‘OAGAIT’: A DECISION SUPPORT SYSTEM FOR GRADING KNEE OSTEOARTHRITIS USING GAIT DATA

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

NİGAR ŞEN KÖKTAŞ

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SCIENCE

JANUARY 2008

Approval of the Graduate School of Informatics

____

Prof. Dr. Nazife BAYKAL
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

____

Assoc. Prof. Dr. Yasemin YARDIMCI
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

____

Prof. Dr. Neşe YALABIK
Supervisor

Examining Committee Members

Prof. Dr. Volkan ATALAY	(METU, CENG)	________________
Prof. Dr. Neşe YALABIK	(METU, CENG)	________________
Assoc. Prof. Dr. Erkan MUMCUOĞLU	(METU, II)	________________
Assoc. Prof. Dr. Sibel TARI	(METU, CENG)	________________
Assoc. Prof. Dr. Güneş YAVUZER	(Ankara Unv., Med. Fac.)	________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Nigar ŞEN KÖKTAŞ

Signature:___________________________

ABSTRACT

‘OAGAIT’: A DECISION SUPPORT SYSTEM FOR GRADING KNEE OSTEOARTHRITIS USING GAIT DATA

Şen Köktaş, Nigar

Ph.D., Department of Information Systems

Supervisor: Prof. Dr. Neşe Yalabık

January 2008, 142 pages

Gait analysis is the process of collecting and analyzing quantitative information about walking patterns of the people. Gait analysis enables the clinicians to differentiate gait deviations objectively. Diagnostic decision making from gait data *only* requires high level of medical expertise of neuromusculoskeletal system trained for the purpose. An automated system is expected to decrease this requirement by a ‘transformed knowledge’ of these experts.

This study presents a clinical decision support system for the detecting and scoring of a knee disorder, namely, Osteoarthritis (OA). Data used for training and recognition is mainly obtained through Computerized Gait Analysis software. Sociodemographic and disease characteristics such as age, body mass index and pain level are also included in decision making. Subjects are allocated into four OA-severity categories, formed in accordance with the Kellgren-Lawrence scale: “Normal”, “Mild”, “Moderate”, and “Severe”.

Different types of classifiers are combined to incorporate the different types of data and to make the best advantages of different classifiers for better accuracy. A decision tree is developed with Multilayer Perceptrons (MLP) at the leaves. This gives an opportunity to use neural networks to extract hidden (i.e., implicit) knowledge in gait measurements and use it back into the explicit form of the decision trees for reasoning.

Individual feature selection is applied using the Mahalanobis Distance measure and most discriminatory features are used for each expert MLP. Significant knowledge about clinical recognition of the OA is derived by feature selection process. The final system is tested with test set and a success rate of about 80% is achieved on the average.

Keywords: Gait analysis, grading knee OA, combining classifiers, clinical decision support systems

ÖZ

‘OAGAIT’: YÜRÜYÜŞ VERİLERİ KULLANARAK DİZ OSTEOARTRİTİ DERECELENDİRMESİ İÇİN BİR KARAR DESTEK SİSTEMİ

Şen Köktaş, Nigar
Doktora, Bilişim Sistemleri
Tez Yöneticisi: Prof. Dr. Neşe Yalabık

Ocak 2008, 142 sayfa

Yürüyüş analizi insanların yürüyüş örüntüleri hakkında niceliksel bilgi toplama ve analiz etme yöntemidir. Yürüyüş analizi klinik uzmanların yürüyüşteki sapmaları objektif olarak ayırt etmelerini sağlar. Sadece yürüyüş verisi ile tanısal karar vermek kas-sinir-iskelet sistemi hakkında ileri medikal uzmanlık gerektirir. Otomatik bir sistemin uzmanların “dönüştürülmüş bilgileri” ile bu ihtiyacı azaltacağı beklenmektedir.

Bu çalışma bir diz rahatsızlığı olan Osteoartrit’in, tespiti ve derecelendirilmesi için tasarlanan bir klinik karar destek sistemini sunmaktadır. Öğrenme ve tanıma için kullanılan veri bir Bilgisayarlı Yürüyüş Analizi yazılımı yolu ile toplanmıştır.

Sosyodemografik ve yaş, vücut kitle endeksi ve ağrı seviyesi gibi hastalık karakteristikleri de karar verme sürecine dahil edilmiştir. Kişiler Kellgren-Lawrence ölçeğine göre dört OA-şiddet derecesine ayrılmıştır: “Normal”, “Hafif”, “Orta” ve “Şiddetli”.

Farklı türlerde verileri kapsamak ve daha iyi doğruluk oranları için farklı sınıflandırıcılar birleştirilmiştir. Yapraklarına Çok Katmanlı Algılayıcılar yerleştirilen bir karar ağacı geliştirilmiştir. Bu yöntem, sinir ağlarını yürüyüş ölçülerinde saklı bilgileri çıkarma ve bunları karar ağacının açık biçimine sebeplendirme için geri bildirme fırsatı verir.

Mahalanobis Uzaklığı ölçüsünü kullanarak bireysel öznitelik seçme uygulanmış ve her bir uzman Çok Katmanlı Algılayıcılar için en ayırt edici öznitelikler kullanılmıştır. Öznitelik seçme sürecinde OA’nın klinik tanınması hakkında önemli bilgiler çıkarılmıştır. Üretilen son sistem test verisi ile test edilmiş ve ortalamada yaklaşık %80 başarı oranı elde edilmiştir.

Anahtar Kelimeler: Yürüyüş analizi, diz OA’sı derecelendirilmesi, birleşik sınıflandırıcılar, klinik karar destek sistemleri

To my beloved husband

Ersoy KÖKTAŞ

ACKNOWLEDGMENTS

I would like to express my deep and sincere gratitude to my supervisor, Prof. Neşe YALABIK. Her understanding, encouraging and personal guidance have provided a good basis for the present thesis. We have not only performed academic researches together, but also enjoyed social activities like cooking, traveling, going out, chatting and more. Her way of thinking, communicating and motivating have been of great value for me during our eight-year togetherness which I hope to be everlasting.

I owe special thanks to Prof. Volkan ATALAY and Assoc. Prof. Güneş YAVUZER who have always welcomed me whenever I had questions throughout the whole progress.

I am also grateful to my committee members Assoc. Prof. Erkan MUMCUOĞLU and Assoc. Prof. Sibel TARI for reading the thesis and for their constructive comments.

I am sincerely grateful to Professor Robert P.W. DUIN, from Delft University of Technology Netherlands, for providing a research opportunity at his department and for providing support and guidance during my stay in Delft.

I also want to thank to my friends, especially to Pınar ONAY DURDU, Habil KALKAN, Fatih NAR, Serkan ŞIK, Burcu AKMAN and Sibel GÜLNAR for their endless support and motivation during our lunches and coffee breaks. Special thanks go to Bülent ÖZTÜRK for his being a gracious friend and for his invaluable technical supports. I also would like to acknowledge my dear friends Semra YILMAZ, Ahmet TAŞTEKİN, Yalçın BOSTAN and Mine IŞIKSAL.

They have created a friendly atmosphere and filled my life with fun whenever I needed.

I express thanks to my friends in Image Processing and Pattern Recognition Laboratory, İsmet YALABIK, Özge ÖZTİMUR, Aykut ERDEM, Erkut ERDEM and Oral DALAY for creating a warm and friendly environment since my first day in the department.

Thanking my family never be enough; I express my special appreciation to my parents and sisters for their encouragements and supports throughout this study and whole my life.

Finally, thanks go to my beloved husband Ersoy KÖKTAŞ. Without his patience, understanding and motivation it was impossible to finish this study.

TABLE OF CONTENTS

ABSTRACT iv
ÖZ vi
DEDICATION viii
ACKNOWLEDGMENTS ix
TABLE OF CONTENTS xi
LIST OF TABLES xv
LIST OF FIGURES xvii
LIST OF ABBREVIATIONS xvii
CHAPTER
1. INTRODUCTION 1
1.1. GAIT ANALYSIS 1
1.2. OSTEOARTHRITIS 3
1.3. LITERATURE SURVEY ABOUT GAIT CLASSIFICATION 5
1.4. OBJECTIVES AND CONTRIBUTIONS OF THE STUDY 8
1.5. SUPPORTING PROJECT 10
1.6. ORGANIZATION OF THE WORK 11
2. OSTEOARTHRITIS AND DATA PROPERTIES 14
2.1. OSTEOARTHRITIS 14
2.1.1. Causes 14

2.1.2. Symptoms 15

2.1.3. Diagnosis 16

2.1.4. Treatment 18

2.2. DATA COLLECTION METHODS 20

2.3. PROPERTIES OF GAIT DATA 23

2.4. DATA STORAGE 31

3. PATTERN RECOGNITION METHODS 33

3.1. FEATURES AND DIMENSIONALITY REDUCTION 33

3.1.1. Feature Extraction 34

3.1.2. Feature selection 37

3.2. CLASSIFIERS 39

3.2.1. Tree Classifiers 39

3.2.2. Neural networks 45

3.2.3. Combining Classifiers 49

3.2.4. Combination Schema 51

4. FEATURE SELECTION AND CLASSIFICATION (GRADING) 54

4.1. INTRODUCTION 54

4.2. STATISTICAL ANALYSIS OF GAIT DATA: COMPARISON OF FEATURE REDUCTION/SELECTION AND CLASSIFICATION ALGORITHMS 55

4.2.1. Feature reduction and selection methods 56

4.2.2. Comparing Classifiers 61

4.2.3. Results of comparisons 63

4.3. COMBINING MLPS FOR GAIT CLASSIFICATION 64

4.3.1. Dataset Properties 65

4.3.2. Classification by Combining MLPs 67

4.4. A DECISION TREE-MLP MULTICLASSIFIER FOR GRADING KNEE OA 71
4.4.1. Preprocessing of data 74
4.4.2. Feature Reduction and Selection 75
4.4.3. Combining classifiers 79
4.4.4. Training and testing 82
5. IMPLEMENTATION AND RESULTS 85
5.1. INTRODUCTION 85
5.2. CREATING THE DECISION TREE 85
5.3. FEATURE REDUCTION AND SELECTION PROCESSES 90
5.4. IMPLEMENTATION OF MLPS 94
5.5. RESULTS 96
5.6. CLASSIFICATION EXAMPLES 101
5.7. OVERALL ANALYSIS OF THE RESULTS 103
6. OAGAIT DECISION SUPPORT SYSTEM 105
6.1. CLINICAL DECISION SUPPORT SYSTEMS 105
6.2. PROPERTIES OF A GOOD CDSS 108
6.3. FEATURES OF OAGAIT 110
6.4. OAGAIT DATABASE 114
6.5. DATA FLOW DIAGRAMS 119
6.6. EVALUATION OF OAGAIT AS A CDSS 122
7. CONCLUSION AND FUTURE DIRECTIONS 125
7.1. PROPERTIES OF OAGAIT SYSTEM 125
7.2. EVALUATION OF THE GRADING ALGORITHM 126
7.3. LIMITATIONS OF THE STUDY 127
7.4. SUGGESTIONS FOR FUTURE WORK 128

REFERENCES 130
APPENDICES 137
A. AN EXAMPLE TO EXCEL FILE OF A SUBJECT 137
VITA 140

LIST OF TABLES

Table 1.1: Stages of the study and results 10
Table 2.1: Limits of the personal features 24
Table 2.2: Properties of the used gait attributes 28
Table 4.1: Mahalanobis Distance values of gait attributes 58
Table 4.2: An example to feature selection by distmaha function of PRTools..... 59
Table 4.3: Datasets after feature reduction and selection processes 59
Table 4.4: Dataset characteristics 66
Table 4.5: Properties of MLPs used in experiment 1 68
Table 4.6: Properties of MLPs used in experiment 2 69
Table 4.7: Success rates (number and percentage) for combining rules 69
Table 4.8: Confusion matrices for used MLPs 70
Table 4.9: Limits of the personal features 74
Table 4.10: Levels of the selected gait attributes for each binary class case 77
Table 4.11: Selection of gait attributes using MD values 78
Table 4.12: Tenfold Crossvalidaton testing 84
Table 5.1: Tree construction algorithm 86
Table 5.2: Pruning algorithm 89
Table 5.3: Selected gait attributes for each two-class case 93
Table 5.4: Classification errors of MLPs 95
Table 5.5: Distribution of samples in decision tree 98
Table 5.6: Confusion matrix for combination (training data) 99
Table 5.7: Confusion matrix for combination (test data) 100
Table 5.8: Confusion matrix for MLP4 100
Table 5.9: Examples of the test set classification 102

Table 6.1: Features of a good CDSS 109
Table 6.2: Features of OAGAIT compared to the ones suggested in 123

LIST OF FIGURES

Figure 1.1: XR image of a normal and OA affected knee joint 4
Figure 1.2: Flowchart for learning and classification phases............................... 12
Figure 1.3: Flowchart for feature reduction and selection processes................... 13
Figure 2.1: OA of the knee 18
Figure 2.2: Replacing knee operation 20
Figure 2.3: Data collection with stereo metric method (visible markers are attached to body of the subject) 22
Figure 2.4: Data collection (a: gait analysis laboratory, b: a subject walking on the platform).. 25
Figure 2.5: 3D analysis of human body 26
Figure 2.6: VICON Software marker adjustment screen 27
Figure 2.7: Calculation of kinetic variables 27
Figure 2.8: Examples to temporal gait attributes (data set C).............................. 30
Figure 2.9: Patient recording form 32
Figure 3.1: Averaging five consecutive time samples of two gait waveforms 36
Figure 3.2: An example to decision tree classifier 40
Figure 3.3: Examples to a) balanced and b) unbalanced trees 42
Figure 3.4: A simple neuron model.. 46
Figure 3.5: An example to multilayer perceptron with one hidden layer. 47
Figure 3.6: Examples to classification regions by one, two and three layer MLPs 48
Figure 3.7: Reasons why an ensemble classifier may be better than an individual one 50
Figure 3.8: Approaches for building ensemble classifiers (differentiation at each level may be considered as an approach).. 52

Figure 4.1: Error rates of the classifiers for the a) averaged datasets b) FFT applied datasets ... 62
Figure 4.2: Learning curves for some classifiers ... 63
Figure 4.3: Graphs of the gait data (*healthy (a) and knee osteoarthritis (b)*) ... 66
Figure 4.4: MLP combination schemas for experiment 1 (a), and experiment 2 (b) H/S: Healthy or sick, FV: Feature vector ... 68
Figure 4.5: Flowchart for classifier design process ... 73
Figure 4.6: Flowchart for feature reduction and selection processes ... 76
Figure 4.7: Flowchart for classifier combining ... 81
Figure 4.8: Proposed example combination in a simplified symbolic representation. ... 84
Figure 5.1: An example to created decision tree (no pruning) ... 88
Figure 5.2: Cost (classification error) versus tree size ... 90
Figure 5.3: Averaging consecutive time samples for feature reduction ... 91
Figure 5.4: Dataset dimension versus misclassification rate (a. averaged datasets, b. FFT applied datasets) ... 92
Figure 5.5: ROC curves for MLPs ... 96
Figure 5.6: Composed decision tree by prune level 3 ... 97
Figure 6.1: Patient recording form ... 111
Figure 6.2: Patient tracking form ... 112
Figure 6.3: WOMAC form ... 113
Figure 6.4: Grading screen ... 114
Figure 6.5: ER diagram for OAGAIT database ... 115
Figure 6.6: Table: gait_data ... 116
Figure 6.7: Table: Womac_results ... 117
Figure 6.8: Table: time_dist ... 118
Figure 6.9: Table: personal_info ... 118
Figure 6.10: Table: measure_data ... 119
Figure 6.11: Level 1 DFD of the OAGAIT for data recording ... 120
Figure 6.12: DFD for patient tracking process ... 121
Figure 6.13: DFD for patient grading process ... 122

LIST OF ABBREVIATIONS

OA	Osteoarthritis
MRI	Magnetic Resonance Imaging
NN	Neural Network
FFT	Fast Fourier Transform
CCD	Charge Coupled Device
VCM	Vicon Clinical Manager
GRF	Ground Reaction Force
WOMAC	Western Ontario and McMaster University Osteoarthritis Index
PCA	Principle Component Analysis
DFT	Discrete Fourier Transform
ANN	Artificial Neural Network
MLP	Multilayer Perceptron
CDSS	Clinical Decision Support System
CDDSS	Clinical Diagnostic Decision Support System
DSS	Decision Support System
BMI	Body Mass Index
DFD	Data Flow Diagram
MD	Mahalanobis Distance
XR	Roentgen Rays

CHAPTER 1

INTRODUCTION

1.1. GAIT ANALYSIS

Gait analysis is the process of collecting and analyzing quantitative information about walking patterns of the people. Gait analysis, when considered as an automated system, is usually used for two major applications: human identification and clinical applications.

Human identification is an important security issue. In most cases it is not so easy to determine the identity of the person but many applications work well for some special cases, such as gender classification **[1],** age classification **[2]** etc. In most of these studies silhouettes are obtained from image sequences and required features are gathered from these **[3-10]**. Since exact human id verification requires more complex systems and huge amount of data, these studies are at their initial phases **[3-10].**

The application of gait analysis in medicine is also a well-studied subject. There are studies that have shown that the number of surgical procedures is reduced after a three-dimensional (3-D) gait analysis **[11-13].** Moreover, gait analysis is important for orthopedists to develop a treatment plan or to track the improvement of persons having gait problems (i.e. Parkinson, cerebral palsy, arthritis **[13-20])**. Examples of the application area of clinical gait analysis include **[11-15]**:

- the assessment of orthopedic diseases' progression to aid in the determination of appropriate surgical or orthotic intervention
- the examination of the progression of neuromuscular disorders such as Parkinson's or muscular dystrophy
- the quantification of the effects of surgery, that is, pre and post-operative patterns
- the evaluation of the effectiveness of prosthetic joint replacement
- the examination of improvements in orthotic design and
- the quantification of changes in prosthetic design

Gait analysis enables the clinicians to differentiate gait deviations objectively. It serves not only as a measure of treatment outcome, but also as a useful tool in planning ongoing care of various neuromuskuloskeletal disorders such as cerebral palsy, stroke, Osteoarthritis (OA), as a support to other approaches such as X-rays, MRI, chemical tests etc. Gait process is realized in a 'gait laboratory' by the use of computer-interfaced video cameras to measure the patient's walking motion by the use of electrodes placed on the skin to follow muscle activity, and by the use of force platforms embedded in a walkway to monitor the forces and torques produced between the patient and the ground. Resultant data (such as knee angle/time) is tabulated in graphic/numerical forms by commercial software. Storing gait data makes the comparison of patients to each other (i.e. normal and patient gait classification) and to themselves (examining improvements of patient by previous data) possible.

In addition to temporal changes of joint angles and ground reaction force data, time-distance parameters of the gait such as velocity, cadence, stride length, step length are recorded. It is not possible to detect the resultant biomechanical musculoskeletal characteristics using other approaches such as the radiographic (X-ray, computerized tomography and/or MRI) evaluations, which makes gait analysis a preferable tool for many cases.

If the physician him/herself interprets the gait data for clinical decision making, then it may be called a "non-automated procedure" and "automated procedure" if it is interpreted partially by any kind of decision support software. Non automated decision making from gait data only requires high level of expertise of neuromusculoskeletal system trained for the purpose. An automated system is expected to decrease this requirement by a 'transformed knowledge' of these experts. This way, clinicians' time is saved, and the probability of human errors is decreased. Automated gait analysis in medicine may also be used as a consultative and educational tool**.**

In this study, 'Knee Osteoarthritis (OA)' is chosen as an application. OA severity levels are established through the Kellgren-Lawrence radiographic grading system **[21].** The decision support system showed here aims to guess the grade of the illness without need of radiography which is a more expensive system and also may have invasive side effects **[22]**.

1.2. OSTEOARTHRITIS

OA is a disorder that affects joint cartilage and surrounding tissue that shows itself by pain, stiffness and loss of function **[22]**

This disease occurs mostly because of cartilage deformations. Bone can overgrow at the edges of the affected joint and bumps can be seen and felt. All the components of the joint deteriorate in some ways and so alter the structure of the joint. OA usually begins with one of a few joints and most often gradually increase. Earliest symptom is the pain which is worsened by weight bearing and relieved by rest. Stiffness is felt after some inactivity and lessens with movement. As OA progresses, joint motion becomes restricted, and tenderness and crepitus may appear.

The muscles surrounding and supporting the joint (such as knee) may stretch. So the joint becomes unstable and stiff, and loses it range of motion. Touching or moving the joint (particularly when standing, climbing stairs, or walking) can be very painful. Figure 1.1 shows X-Ray (XR) images of a normal and an OA affected knee joint. The narrowing of the sick joint can be clearly seen.

Figure 1.1: XR image of a normal and OA affected knee joint [22]

Diagnosis of the disease is made according to symptoms, physical examination and XR images. XR images show evidence of OA especially in weight-bearing joints such as hip and knee. Kellgren-Lawrence is a method used for radiological assessment of OA **[21].** According to this method OA is divided into five grades as follows:

- Grade 0 indicates a definite absence of x-ray changes of OA.
- Grade 1, doubtful narrowing of joint space and possible outgrowth of the bone;
- Grade 2, definite outgrowth of the bone and possible narrowing of joint space;
- Grade 3, moderate multiple outgrowths, definite narrowing of joints space, some sclerosis and possible deformity of bone contour;
- Grade 4, large outgrowths, marked narrowing of joint space, severe sclerosis and definite deformity of bone contour.

Postural exercises for stretching and strengthening are advised for treatment of the OA. Exercises may help maintain healthy cartilage, increase range of motion, and strengthen surrounding muscles. Soft chairs, recliners, mattresses, and car seats may worsen symptoms. Specific exercises may be needed for OA of the spine. Exercises should include muscle strengthening and low impact aerobic exercises (such as walking, swimming, and bicycle riding).

Physical therapy is another effective treatment method. Heat improves muscle function by reducing stiffness and muscle spasm. Massage by trained therapists and deep heat treatment may be useful. Cold may be applied to reduce pain.

Drugs are used to reduce the symptoms and thus allow more appropriate exercises. If a sudden injury occurs the fluid inside the joint may be removed and a form of cortisone may be injected directly to the joint. But this treatment is not a long term relief. Another injection method is done by hyaluronate which is a component of a normal joint fluid. This method may provide significant pain relief for longer time periods.

The replacement of the damaged knee joint with an artificial joint is a surgical operation applied for treatment of OA of the knee. Surgery may help when all other treatments fail to relieve pain. It is usually very successful to improve motion and decrease pain. Since the artificial joint does not last forever, joint replacement should be considered when function becomes limited.

1.3. LITERATURE SURVEY ABOUT GAIT CLASSIFICATION

A large number of studies are conducted for gait classification. Since gait data is high dimensional and complex, to design a complete decision support system may require a combination of all available features. In non automated traditional systems physicians make decisions about the illnesses by interpreting all available data. On the other hand, most of the automated systems ignore history and symptoms of the patient, such as age, pain grade, family history. There are some known facts about the effects of some factors to cause or to develop OA **[20-30],** such as occurring in the same frequency in both sexes before the age of 55 but being more common in women after 55. Obesity places people (particularly women) at increased risk for osteoarthritis because of increased weight on the joints. Injury from different sources can also contribute to osteoarthritis. Repeated minor injuries or a single injury to a joint may change the normal joint structure. A genetic defect may promote breakdown of the protective architecture of cartilage. Actually, in traditional non automated clinical decision making, physicians listen to the patient and use this kind of qualitative information in addition to the lab test. So this kind of non-numeric information should also be included in the decision making process.

The aim of pattern recognition research for clinical gait analysis is to find ways to support doctors in decision making and treating patients using gait data. Standard movement patterns are produced for healthy walkers, and then these patterns served as a baseline for examination of the walking pattern of patients' gait with abnormalities or diseases. The interpretations of quantitative gait data are experimented by pattern recognition techniques before **[31, 32]**. Most popular of these are neural networks (NNs) **[33-40]** and support vector machines (SVMs) **[2, 41].** The use of NNs for experimental gait classification is not new. There are studies in which NNs are trained by force platform data to distinguish 'healthy' from 'pathological' gait **[34, 35, 40]**. In addition to these, people are identified among a few subjects (less than 10) by using joint angles as features **[33, 36, 38, 39]**. These studies produce reasonable results for NNs use in gait classification. A few of these studies will be explained in more detail in following paragraphs.

Kohle et al. **[34]** categorized gait pathology based on the ground reaction forces. They measured two successive ground reaction forces of 131 subjects having various diseases like calcaneus fracture, and limb deficiencies. 94 normal subjects' data was also gathered. FFT coefficients of vertical components of the two ground reaction forces were used as inputs to a standard network with one hidden layer. The accuracy of this network in discriminating healthy from pathological gait was 95%. This study summarized that simple two-category gait classification with large number of input parameters is achievable with neural networks.

A similar study conducted by Barton and Lees **[36]** extended the classification problem to a three class case. A neural network with two hidden layers is used to categorize maximum value of ground reaction forces into one of three categories: healthy feet, pes cavus (a deformity of the foot characterized by an abnormally high arch and hyperextension of the toes) and hallux (big toe) valgus. The pressure patterns of 18 subjects were recorded and scaled to a size and normalized to the interval [0, 1]. The number of inputs of network reached to 1316 which is much more than the previous examples. The accuracy of the network was claimed to vary between 77% and 100% based on the size of the train and test set. Since the conditions classified here can be identified by routine medical exams, the advantage of the proposed system was not clearly understandable. In second stage of the study hip-knee joint angles of the eight healthy subjects were calculated via a set of four reflective markers. They mentioned that hip–knee joint angle diagrams are characteristic of a subject's gait pattern and so could be used for automated identification of gait patterns. Subjects were walked on a walking platform under three different conditions; normal walking, simulated leg length difference and simulated leg weight difference. The angles were normalized in time; Fourier transformed and normalized to the interval [0, 1]. A neural network again with two hidden layers is used and 83.3% average accuracy rate was achieved.

In another study Lafuente et al. **[40]** used a standard feedforward neural network with one hidden layer to classify four-category gait patterns. They collected data from 148 subjects with ankle, knee or hip arthritis and 88 normal subjects without limb pathology. The features consisted of cadence, velocity and five kinetic magnitudes. A three layered network was trained by these inputs to discriminate four classes and an accuracy of 80% was reached.

As a summary these studies have shown the potential for multi-category classification of the gait patterns. However, if the aim of the classification is medical usage, more detailed classification problems (i.e. grading of a disease) with high dimensional and diverse data arise.

Unfortunately, as the dimension of the obtained data increases accuracies may deteriorate. Recently, the focus has been on combining several classifiers and getting a consensus of results for better accuracy **[42, 43]**. Today, combining methods are preferred for many well known pattern recognition problems such as character recognition, speech recognition **[42-55]** etc. On the other hand, decision trees have been widely used for medical decision making processes **[42, 56-62]**. They have not been applied to gait data analysis but have demonstrated potential in analyzing gait data **[31].** These two classifiers neural networks and decision trees have their specific advantages and disadvantages, and most of these characteristics complement each other. Hence, advantages of both approaches might be utilized if combined properly.

Automatic feature selection from many numerical gait parameters is another subject that's not studied well in medical applications. Actually, there are many medical practices, testing the variations in the gait attributes which are caused by the related disease **[20-30]**. Selection of attributes is usually done by using the result of these studies by expert clinicians. However, the judgments may vary in different experts leading to the different interpretations. Obviously, automated selection lessens the dependence and the load on the experts and gives more freedom to the researchers.

1.4. OBJECTIVES AND CONTRIBUTIONS OF THE STUDY

The main objective of this study is to design a decision support system to help physicians by supplying accurate and practical ways to interpret the gait data and further follow the progress of OA. Grading of the diseases is helpful for physicians in treatment and operation plans. The treatment plans for knee OA are made according to the grade of the illness **[20, 21, 26-28]**. This grading is usually done by physicians using radiographic films of the knee. The patient is also walked on the gait laboratory if available and the data is used as assistive to the radiography results. Actually the classification of sick and normal subjects has been studied before **[18, 19, 34-39]**, but the grading of the OA is new. The decision support system developed by this study is expected to be a supportive tool for grading and treatment planning of knee OA. Moreover computerizing the current system is expected to provide some additional benefits for gait laboratory users. Since automated evaluation of the data does not require as much expertise as manual evaluation, the usage of the gait laboratories may increase. In addition the misdiagnosis occurring because of different comments of the different experts are expected to be minimized.

Original contributions of the study are summarized below:

- The system supports a base for patient follow up in time. Gait patterns of the subjects taken in different times can be compared so that more accurate treatment plans are possible.

- Automated feature selection process reveals that gait measurements for different parts of the body such as knee or hip to be more effective for different scores of the OA. These results may be valuable for physicians for effective decision making about treatment of OA and reasoning the conclusion. Also, dependence and the load on the experts are reduced and more freedom to the researchers is given by the automated feature selection process.

- Another original contribution of the study is taking advantage of working with a gait expert at each stage of the study. A physical medicine and rehabilitation expert gives support to the study as both an expert and a target user of the proposed decision support system. So, the results of the data analysis process can be commented for further recognition of the selected illness and helping the treatment.

- Since, all available various structured data is used in a hierarchical way for design of the decision support system, it models the expert's decision making process well. The combined decision tree-MLP approach is also expected to be applicable to similar type of medical decision making processes, where both disease characteristics and clinical measurements and tests are to be combined.

1.5. SUPPORTING PROJECT

This study is implemented as a research project supported by TÜBİTAK. The group is composed of physicians, computer scientists, PhD students and technical personnel. The group communicated well at each stage of the study. Table 1.1 summarizes the milestones of the whole process.

Table 1.1: Stages of the study and results

Stages of the study	Results
Problem definition	• Gait laboratory is seen and problems are listed • Literature survey about previous studies are done • Objectives of the study are determined
Requirements analysis	• Hardware/software requirements of laboratory are determined (i.e. new PCs are bought, a database is created for easing data collection) • A questionnaire is prepared for expert physicians to determine expected features of the system • Face to face interviews are also implemented
Design	• Design of the system is made iteratively • Different feature reduction/selection and classification methods are compared • New approaches are searched for better classification (i.e. combining methods) • Results are discussed with expert, suggested changes are made and iteration started again

Table 1.1 (cont)

Implementation	• Datasets are created by querying the database • Preprocessing of the data (i.e. cleaning empty entries, converting non numeric features to numeric ones) • Feature reduction/ selection methods are applied • Two-class experiments are done by neural network methods • Multi-class classification is done by combining decision trees and neural networks
Testing	• System is tested by unseen data

After determining objectives of the study, requirements of the gait laboratory are determined. The old PC is replaced with the new one, and then required software is uploaded. Then, most important deficiency of the laboratory is determined as need of a database. A complete database is created to keep all data together in one system to ease the query processes. Some software interfaces are created to read data from the current system and user friendly interfaces are supplied for laboratory users.

1.6. ORGANIZATION OF THE WORK

The outline of the report is organized according to the steps of classifier design process. A classifier topology which is similar to Mixture of Experts (MME) approach based on decision trees and a number of Multilayer Perceptrons (MLPs), each expert in separating two adjacent degrees is proposed. A scoring of the OA (0-3), which specifies the degree of the disease, is used as the categories of the classifier. Finally, the system is tested by unseen data and results are presented in forward sections of this report. Figure 1.2 shows the stages in classifier design process.

For better design of the classifier some different feature reduction and selection methods are compared. Averaging method is selected for reduction and the Mahalanobis Distance criterion is selected for feature selection. Further details will be discussed in Chapter 4 of the report. Figure 1.3 summarizes the feature reduction and selection processes.

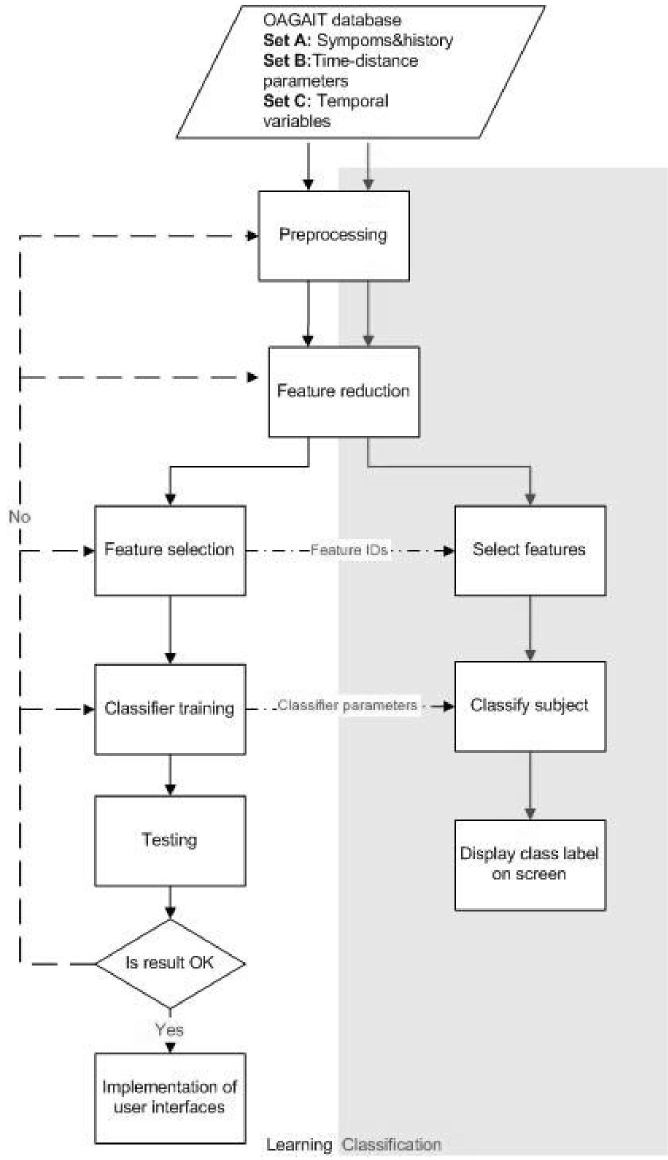


Figure 1.2: Flowchart for learning and classification phases

Since the feature set is quite diverse new classification approaches like combining classifiers are searched in the literature. Following the feature reduction and selection processes a combination algorithm is created by using decision tree and neural network approaches together.

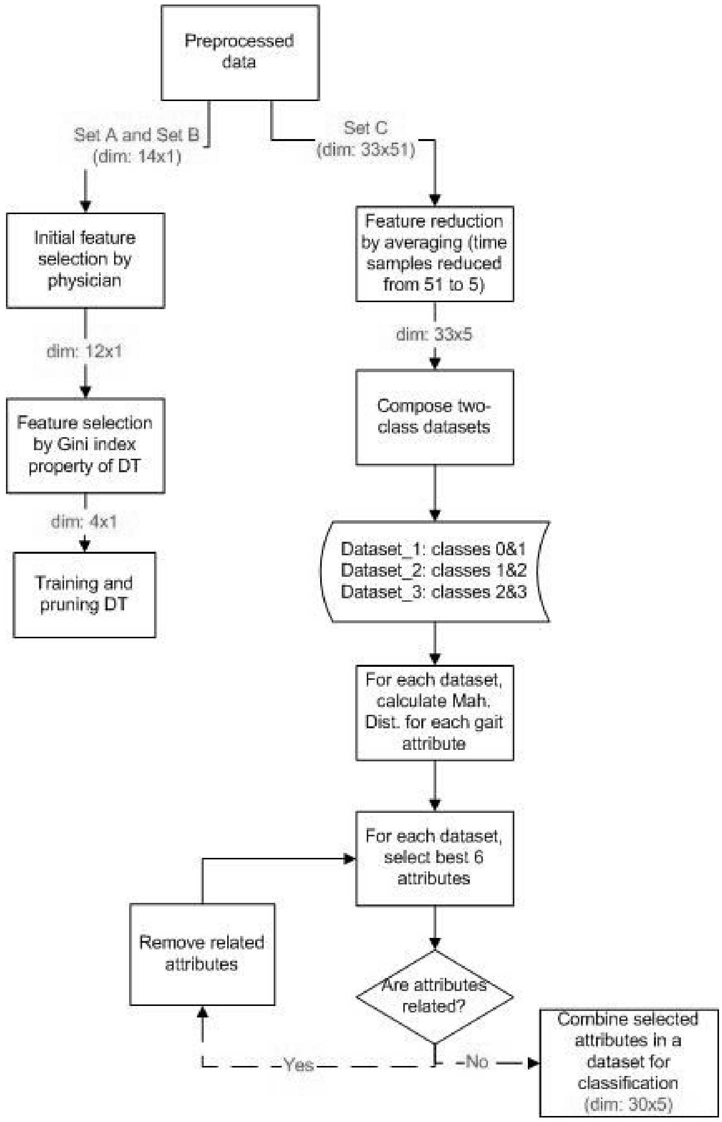


Figure 1.3: Flowchart for feature reduction and selection processes

The remainder of the report is organized as follows. In the next section, data collection and recording methods are discussed. Then in chapter 3 pattern recognition approaches are summarized. In fourth chapter implementation and analysis of the results are given. Detailed information about OAGAIT system is discussed in chapter 5. Finally, conclusion and discussions are presented.

CHAPTER 2

OSTEOARTHRITIS AND DATA PROPERTIES

2.1. OSTEOARTHRITIS

OA is a disorder that affects joint cartilage and surrounding tissue that shows itself by pain, stiffness and loss of function **[20-23].** Although OA is mostly seen in older people, it is not caused by the years of use. But, while the younger people having few symptoms, the older ones develop significant disabilities **[22]**.

2.1.1. Causes

Our joints are normally protected from wearing out by low friction levels, provided by the cartilages between the bones. OA mostly begins with the deformation of the cells that form the cartilage. Then, the cartilage may become soft and cracks on the surface may be seen. Bone can overgrow at the edges of the affected joint and bumps can be seen and felt. All the components of the joint deteriorate in some ways and so alter the structure of the joint **[20-22]**.

OA is classified into two groups; primary and secondary. If the cause of the disease is not known, which is valid for most of the cases, it is called primary OA. If the cause is another disease or condition like infection, deformity, injury, then it is called secondary OA. Some people repetitively stress one joint because of their jobs (i.e. coal miners, bus drivers) and so increase the risk of OA. Obesity may be a major factor in the development of OA, particularly of the knee and especially in women.

2.1.2. Symptoms

Usually, symptoms show themselves in one or a few joints at first. Most commonly affected joints are hip, knee, fingers, neck, lower back and big toes. Pain is the first symptom which usually caused by weight bearing activities. Stiffness is another important symptom which is felt after some inactivity like sleep **[20-22, 25-30]**.

The affected joint may become less movable and it may be more difficult to straighten or bend. The irregular cartilage surfaces cause joints to grind, grate, or crackle when they are moved.

In some joints (such as the knee) the ligaments, which surround and support the joint, may stretch. So the joint becomes unstable and stiff, and loses it range of motion. Touching or moving the joint (particularly when standing, climbing stairs, or walking) can be very painful.

For OA of the spine, the back pain is one of the most common symptoms. Usually, damages of disks or joints in the spine cause only mild pain and stiffness. However, OA in the neck or lower back can cause loss of sense, pain, and weakness in an arm or leg. The overgrowth of bone may presses on the nerves within the spinal canal or before they exit the canal to go to the legs. This leg pain caused by this reason may be confused by the reduced blood supply to the legs.

OA may be stable for many years or may progress very rapidly, but most often it progresses slowly after symptoms are seen. Many people develop some degree of disability. In **[22]** some practical ways to live with OA are advised to the patients.

- Exercise affected joints gently (i.e. in a pool)
- Massage at and around affected joints (trained therapist would do it better)

- Apply a heating pad or a damp and warm towel to affected joints
- Maintain an appropriate weight (extra stress on joints may be dangerous)
- Use special equipment when necessary (for example, walker, neck collar, or elastic knee support to protect joints from overuse)
- Wear well-supported shoes or athletic shoes

2.1.3. Diagnosis

The diagnosis is made according to characteristics of symptoms, physical examination, and the XR images. The XR images of many people aged about 40 show some evidences of OA especially in weight-bearing joints such as the hip and knee. However, XR is not very useful for detecting OA early because it does not show changes in cartilage, which is where the earliest abnormalities occur.

Magnetic resonance imaging (MRI) can reveal early changes in cartilage, but it is rarely used because of expensive cost. There are no blood tests for the diagnosis of OA. But there are some researches about detection of hyaluronic acid within a blood sample **[22]**.

Kellgren-Lawrence is a method used for radiological assessment of OA. Kellgren and Lawrence defined this scoring according to these radiological features **[21]**

- The formation of osteophytes on the joint margins or, in the case of the knee joint, on the tibial spines.
- Periarticular ossicles; these were found chiefly in relation to the distal and proximal interphalangeal joints.
- Narrowing of joint cartilage associated with sclerosis of subehondral bone.
- Small pseudocystic areas with sclerotic walls situated usually in the subchondral bone.
- Altered shape of the bone ends, particularly in the head of femur.

According to these features OA is divided into five grades as follows:

- None
- Doubtful
- Minimal
- Moderate
- Severe

Grade 0 indicates a definite absence of x-ray changes of OA.

Grade 1, doubtful narrowing of joint space and possible osteophytic lipping;

Grade 2, definite osteophytes and possible narrowing of joint space;

Grade 3, moderate multiple osteophytes, definite narrowing of joints space, some sclerosis and possible deformity of bone contour;

Grade 4, large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone contour.

Figure 2.1 shows XR images of knee joints affected by different grades of OA.

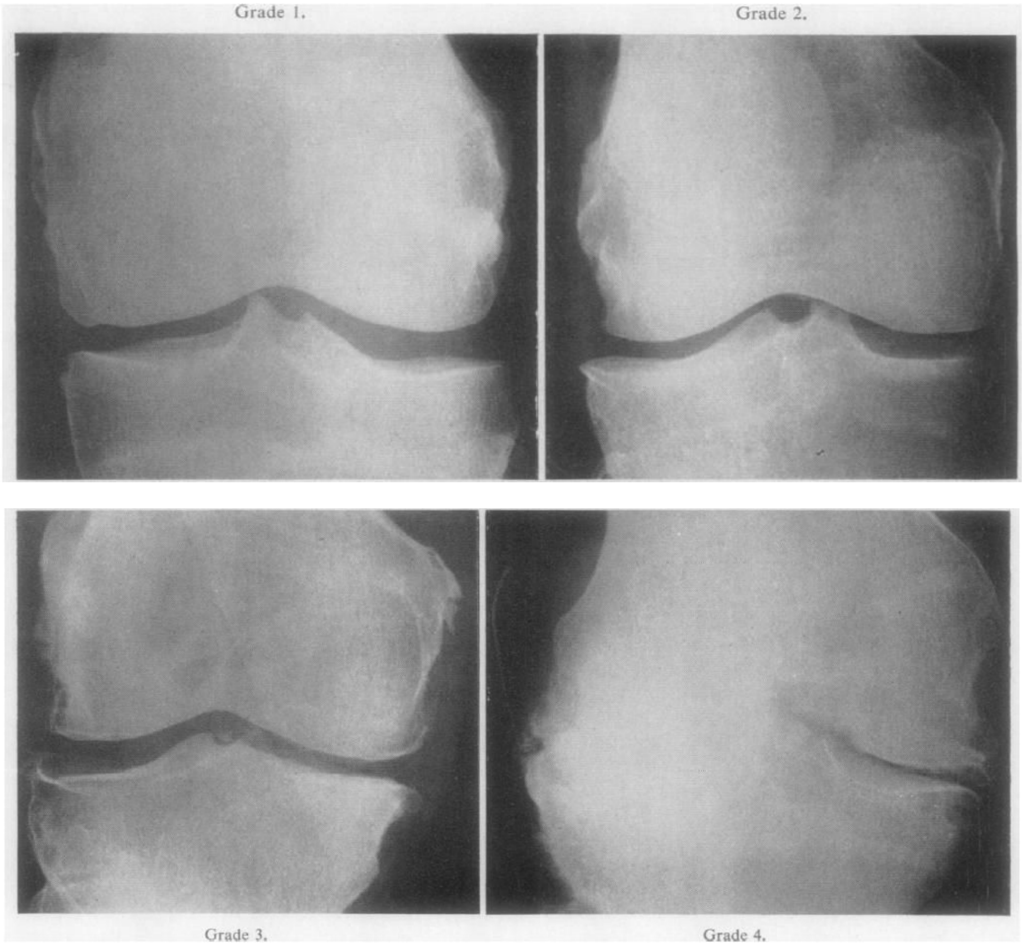


Figure 2.1: OA of the knee [21]

2.1.4. Treatment

Exercises like stretching, strengthening, and postural exercises may help maintain healthy cartilage, increase range of motion, and strengthen surrounding muscles. Exercises must be balanced with rest of painful joints, but immobilizing a joint may make the joint worse. Using excessively soft chairs, recliners, mattresses, and car seats may worsen symptoms; using car seats moved forward, straight-backed chairs with relatively high seats, firm mattresses, and bed boards is often recommended.

For osteoarthritis of the spine, specific exercises may help, and back supports may be needed when pain is severe. Exercises should include muscle strengthening and low impact aerobic exercises (such as walking, swimming, and bicycle riding).

The patient should try to continue his/her normal daily activities such as a hobby or job.

Physical therapy, often with heat therapy can be helpful. Heat improves muscle function by reducing stiffness and muscle spasm. Massage by trained therapists and deep heat treatment may be useful. Cold may be applied to reduce pain. Splints or supports (such as a cane, crutch, and brace) can protect specific joints during painful activities. Shoe inserts (orthotics) may help reduce pain during walking.

Drugs are used to supplement exercise and physical therapy. Drugs do not directly alter the course of osteoarthritis; they are used to reduce symptoms and thus allow more appropriate exercises.

If a joint suddenly becomes inflamed, swollen, and painful, most of the fluid inside the joint may need to be removed and a special form of cortisone may be injected directly into the joint. This treatment may provide only short-term relief, and a joint treated with cortisone should not be used too often or damage may result. A series of injections of hyaluronate (a component of normal joint fluid) into the joint may provide significant pain relief in some people for longer periods of time.

The knee replacement surgery is another treatment method which is applied when the moves become limited. The damaged knee joint may be replaced with an artificial joint. After a general anesthetic is given, ends of the thigh bone (femur) and shinbone (tibia) are smoothed so that the parts of the artificial joint (prothesis) can be attached more easily. One part of the artificial joint is inserted into the thigh bone, and the other part into the shinbone and the parts are cemented in place. Figure 2.2 shows the knee replacement operation **[22]**.

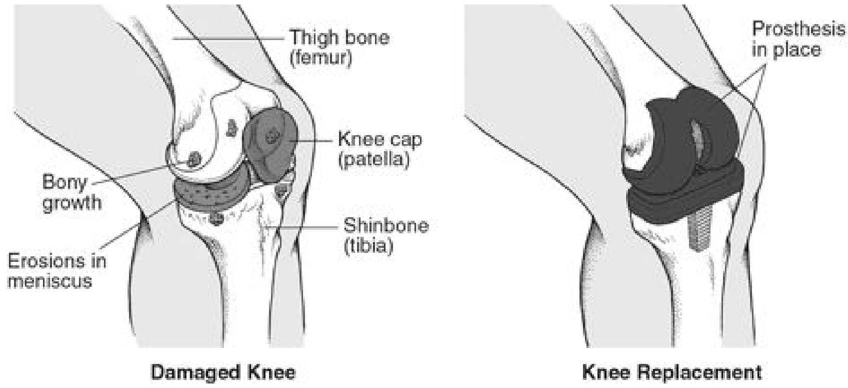


Figure 2.2: Replacing knee operation [22]

Surgery may help when all other treatments fail to relieve pain. Some joints, most commonly the hip and knee, can be replaced with an artificial joint. It is usually very successful to improve motion and decrease pain. Therefore, joint replacement should be considered when function becomes limited. Because the artificial joint does not last forever, such surgery is often delayed as long as possible in young people so the need for repeated replacements can be minimized.

A variety of methods that restore cells inside cartilage have been used in younger people with OA to help cure small defects in cartilage. However, such methods have not yet been proven valuable when cartilage defects are extensive, as commonly occurs in older people.

2.2. DATA COLLECTION METHODS

Kaufman states five existing technologies for data collection for gait analysis **[12]**:

- Electromechanical linkage method
- Stereo metric method
- Roentgen graphic method
- Accelerometer method
- Magnetic coupling method.

An exoskeleton apparatus is employed with the ***electromechanical linkage method*** to measure joint motion. The primary disadvantage for this technique is the cumbersome nature of the instrument and, to a lesser extent, cross coupling of the sensor inputs and joint motion. The requirement for the exoskeleton instrument affects the motion of young subjects making it unusable for clinical measurement.

The ***stereo metric method*** is the most popular one currently used for clinical gait analysis. It employs visible markers attached to the skin on rigid segments of the body structure and tracks their motion using imaging equipment. This technique is implemented using charge coupled device (CCD) cameras and frame-grabber electronics to allow digital images to be captured as the subject moves within the field of view. Digital image analysis allows the physical location of each marker to be computed, using triangulation of the views from an array of camera systems. This technique has minimal impact on the natural motion of the subject and allows data capture without the need to tether the subject to the data acquisition hardware. Figure 2.3 shows a laboratory collecting data with stereo metric method.

A disadvantage of this approach is the increased image analysis complexity resulting from tracking the apparent position of the markers in a two-dimensional (2-D) image on a camera frame-to-frame basis and correlating the position of each marker for the multiple camera positions. Occlusion of markers from the camera field of view and false readings caused by reflection phantoms pose non-trivial, unresolved complications in data capture. In addition, passive markers provide unlabeled trajectory segments that must be manually identified and resolved. This image analysis task requires a significant amount of time for the data gathering process. A second major disadvantage is the reduction in resolution as the camera system is altered to allow a larger field of view. The camera imaging sensors have a fixed number of pixel elements and a compromise must be reached between optical field of view and pixel element resolution size, limiting the clinical measurement volume to approximately a single stride. It is not feasible to measure gait patterns or variability with only one traversal of the instrument walkway. Thus, multiple walking trials need to be collected, which may fatigue the subject.

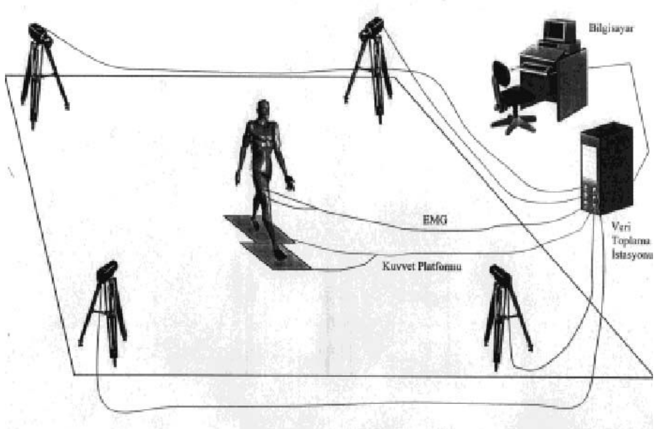


Figure 2.3: Data collection with stereo metric method (visible markers are attached to body of the subject) [12]

The ***biplanar roentgen graphic method*** employs metal markers and x-ray films for the measurement of static positions of a body joint. This approach is not appropriate for the study of dynamic joint motion. Due to the use of ionizing radiation, it also represents a potential health hazard to the subject.

The ***accelerometric approach*** employs sensors attached to the rigid areas of the human subject that measure accelerations in three dimensions. Joint motion is then derived through integration of the accelerometer waveforms given appropriate initial conditions. Integration of the waveforms produces velocities for each of the sensor locations. A second integration step provides the displacement as a function of time. This technique can provide the kinematics motion measurement desired but has been implemented with a tether to the subject for the data acquisition; however, the tether affects the motion of the subject and represents an undesirable feature. In addition, this approach requires an accurate estimate of initial conditions, which is difficult to provide.

The ***magnetic coupling method*** employs a reference magnetic field source that surrounds the subject and an array of magnetic field sensing elements attached to the rigid segments of the subject.

Recent rapid developments in hardware technologies created an attractive environment for image processors. This also created opportunities for gait analysis using video sequences **[3-10]**. Beside above methods, a clinical gait analysis might be limited to a video recording and the measurement of certain gait stride and temporal parameters such as velocity, cadence, stride length, step length and percentage of stance/swing. While the video record is a useful tool in developing and substantiating visual impressions, it is inappropriate to measure joint and segment gait kinematics directly from the videotape or monitor. They do not give an indication of the cause of the gait abnormality and so have limited value in clinical decision-making.

2.3. PROPERTIES OF GAIT DATA

Studies of biomechanical factors in OA are mainly focused on the knee joint. This is primarily for two reasons. First, the knee is the most common joint affected by OA. Second, the anatomy of the knee joint is relatively simpler and more amenable to biomechanical modeling and noninvasive evaluation than other joints **[20]**. Therefore, the data in this study are primarily from measurements on the knee.

In this study, the gait data are collected by the gait experts in Ankara University Faculty of Medicine, Department of Physical Medicine and Rehabilitation Gait Laboratory (shown in Figure 2.4). Before gait analysis, all subjects gave informed consent as advised by the Ethics Committee. The socio-demographic and clinical characteristics of the patients were also collected in the lab before the patients are walked. Electronic format of the form filled by the patients to gather these data is shown in Figure 2.9.

Collected information other than gait is converted to numerical values before they are used as features. For example, while weight and height attributes are not used for classification purpose, body mass index (BMI), which is equal to weight divided by the square of the height, is created as a new feature. Age of the subject is calculated from date of birth, disease periods are converted to months as unit.

Pain and morning stiffness are numeric values between 0 and 10, family history is a binary value indicating whether the same disease exist in family history or not. Sex is another binary valued feature where 0 stands for women and 1 for men. Then, the first subset of the data can be defined as:

A = {age, BMI, pain, stiffness, period, history, sex}

The max and min values of the non binary features of the subjects used in this study are shown in Table 2.1.

Table 2.1: Limits of the personal features

Features	**Normal subjects**			**Patients**		
	Min	max	average	min	max	average
age	19	63	43	41	80	60
BMI	18	46	27	20	49	32
pain	0	0	0	1	10	6,6
stiffness	0	0	0	1	10	5,2
period (year)	0	0	0	0	30	6,6

Subjects underwent gait analysis with the same protocol by one and the same physician. Spatiotemporal and kinematic data were obtained from the Vicon 370 Motion Measurement and Analysis System. This system consisted of 5 video cameras, a computer system for data acquisition, processing and analysis and a data station. The experimental model idealized the lower extremity as a system of rigid links with spherical joints. The joints were assumed to have a fixed axis of rotation. Skeletal movement can be described using surface markers placed in precise anatomical positions.

Figure 2.4: Data collection (a: gait analysis laboratory, b: a subject walking on the platform)

All subjects were instructed to walk at a self selected speed along the walkway and to practice until they could consistently and naturally make contact with both of the force plates. Three acceptable trials were obtained for each foot and averaged to yield representative values. Time-distance parameters of the gait are gathered at the end of one cycle. So the second set of the data is composed of *time distance parameters*:

B = {Cadence, Walking Speed, Stride Time, Step Time, Single Support, Double Support, Stride Length, Step Length}

The kinetic and kinematic features of the gait are gathered by 3D analysis of the human body. Figure 2.5 shows three planes of the human motion. Flexion-extension data is taken in sagittal plane, valgus-varus and abduction-adduction data is taken in frontal plane and rotation data is taken in transvers plane.

Figure 2.5: 3D analysis of human body

External retro-reflective markers, used for computer digitization, were placed on each of the following anatomic locations: anterior superior iliac spine (ASIS), sacrum, lateral thigh, joint line of the knee, lateral shank, calcaneus, lateral malleolus and second metatarsal head. The 3-dimensional position of each reflective marker was sampled 60 times a second. Markers were placed on the bony prominences to minimize artifacts due to skin movement. On the other hand, these locations provided anatomic reference points to locate internally the joint center position of the hip, knee and ankle. The hip joint center was determined using leg length, inter-ASIS distance and ASIS-greater trochanter distance calculated by Vicon Clinical Manager (VCM) **[20].** The knee center was located at one-half the knee width medially along the knee flexion axis. The ankle joint center was located at one-half the ankle medially along the ankle flexion axis. Figure 2.6 shows an example to marker adjustment screen of VICON software.

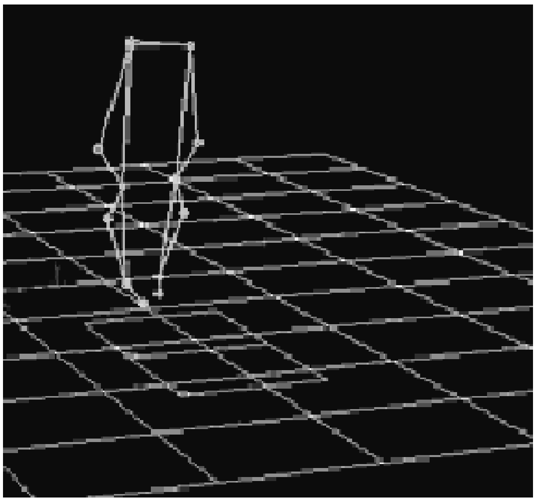


Figure 2.6: VICON Software marker adjustment screen

Besides kinematic variables, ground reaction force (GRF) data is also gathered in one gait cycle. Kinetic variables, which are also important for diagnosis, are calculated by using GRF as shown in Figure 2.7.

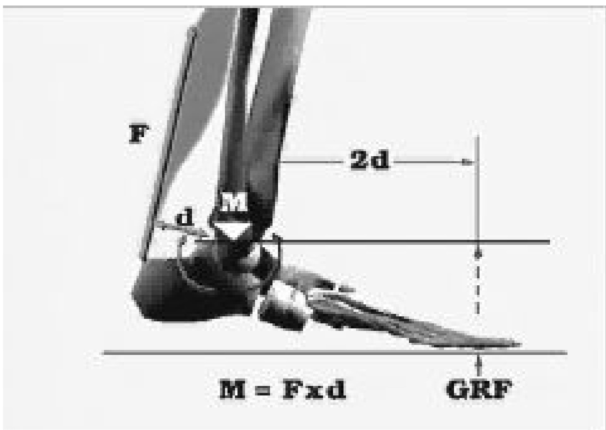


Figure 2.7: Calculation of kinetic variables [19]
Moment = GRF x Distance
Power = Moment x Angular velocity

Ground reaction forces (GRF) were collected using two force plates (Bertec, Columbus, OH). GRF measurements were acquired simultaneously with a measurement of the limb position. Time-to second vertical force peak and values of first and second vertical force peaks were determined. To calculate the moments, each segment of the limb (thigh, shank and foot) was assumed to be a rigid body with a coordinate system chosen to coincide with the anatomic axes. Moments producing flexion-extension, abduction-adduction and internal external rotation at the knee joint were calculated. Angular velocity and acceleration around the longitudinal axis were assumed to be negligible. All moments and ground reaction forces were normalized to body weight and height permitting comparison with other results in the literature. Table 2.2 summarizes the motion planes, the anatomic levels and the types of the 33 kinematic gait attributes used in this study.

Table 2.2: Properties of the used gait attributes ("*x*": exists, "-": not exists, *flex*: flexion, *abd*: abduction, *rot:* rotation

	Joint rotation angles			*Joint net moments*			*Joint net powers*			
Motion plane / *anatomic level*	**flex**	**abd**	**rot**	**flex**	**abd**	**rot**	**total**	**flex**	**abd**	**rot**
Pelvic	x	x	x	-	-	-	-	-	-	-
Hip	x	x	x	x	x	x	x	x	x	x
Knee	x	x	x	x	x	x	x	x	x	x
Ankle	x	x	x	x	x	x	x	x	x	x

Then the final subset of the data can be defined as the temporal changes of the joint angles from four anatomical level and three motion planes (Set C) as below.

C = {PelvicTilt, Pelvic Obliquity Knee Flexion, Knee Varus, … }

Each of these attributes in C above is represented by a graph that contains 51 samples taken in equally spaced intervals for one gait cycle. So the attributes for a given subject can be arranged as a 33-dimensional vector X as below:

$\boldsymbol{X = [X^{(1)}, X^{(2)}, \ldots X^{(33)}]}$ where

$\boldsymbol{X^{(i)} = [X^{(i)}_1, X^{(i)}_2, \ldots X^{(i)}_{51}]}$

$\boldsymbol{X^{(i)}_j}$ is the value of the $\boldsymbol{i^{th}}$ gait attribute at $\boldsymbol{j^{th}}$ time period of the gait cycle. Figure 2.8 shows examples of a graphical representation of the joint angle attributes.

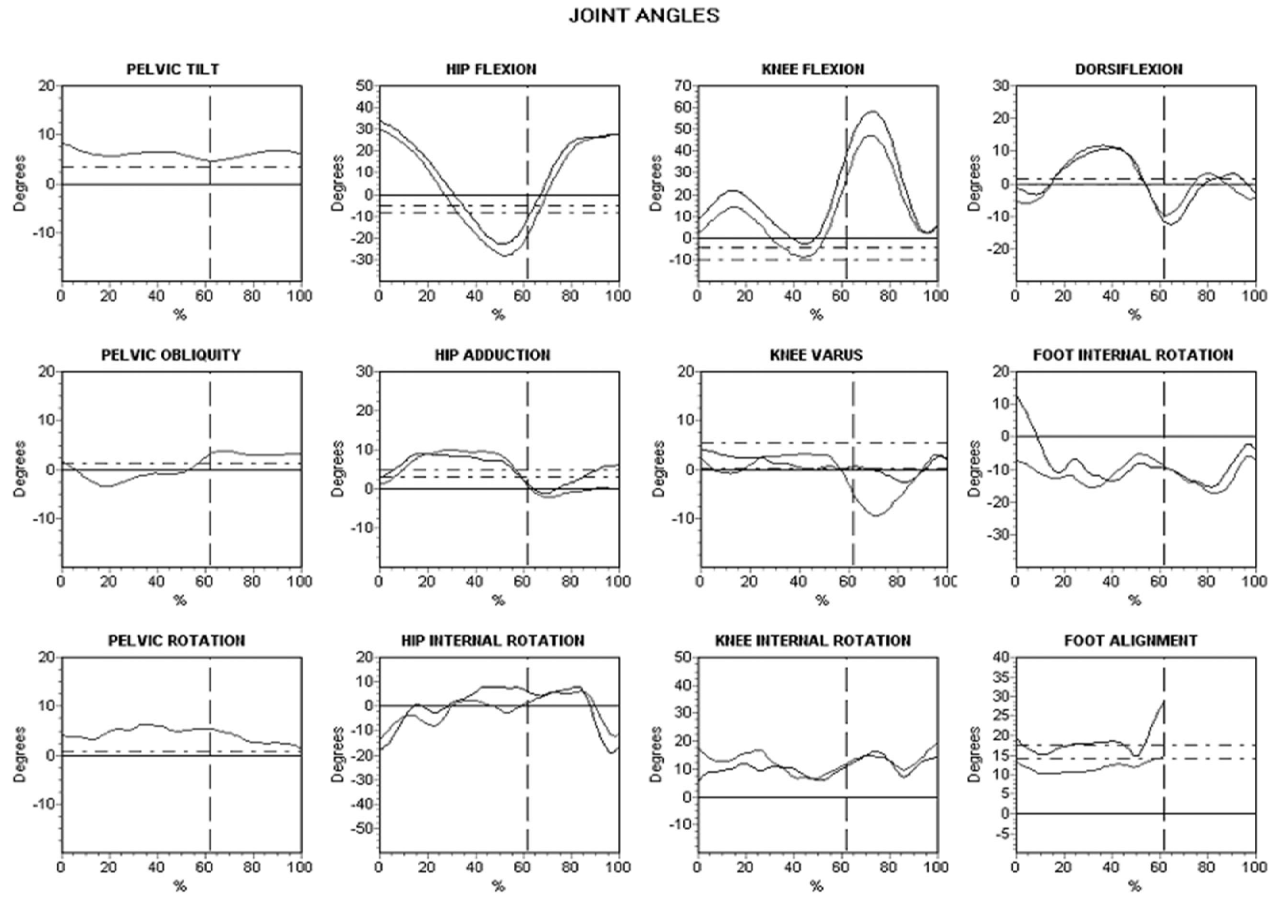


Figure 2.8: Examples to temporal gait attributes (data set C): The data is reported in 2-D charts with the abscissa usually defined as the percentage of the gate cycle and the ordinate displaying the gait parameter.

2.4. DATA STORAGE

As explained above one of the advantages of gait analysis for researchers and medical experts is opportunity of data storage for comparison and other purposes. Gait laboratories need comprehensive, user friendly database systems for efficient data processing. Since each laboratory uses different software, hardware and biomechanics models, it improved its own particular databases and information systems. Standardization of data storage methods is important for opportunity of transferring data between different systems and for creating an international, easily understandable gait terminology. Some institutions have recognized this need and founded some standardization societies. GCMAS (Gait and Clinical Motion Analysis Society) centered in USA and ESMAC (European Society of Movement Analysis for Adults and Children) centered in Europe are most known of these societies**.** But unfortunately a widely used international gait standards could not be created until today. On the other hand, usage of a standard file format (c3d) for data storage is becoming widespread **[14].**

In our gait laboratory collected gait data can be saved as MS Excel file, so the access and the transfer of the data become easier. These files show the time-distance parameters of the gait and temporal changes of the joint angles and graphs of them. An excel file of a patient is shown in Appendix A of the report as an example. Since there was no database keeping all available data in the system earlier, it used to be difficult to combine personal information and gait data of the patients for processing. The patients are asked to fill a form about personal information and to take WOMAC (Western Ontario and McMaster University Osteoarthritis Index) questionnaire before walking in the laboratory. Both of these forms were kept in paper files, so it was mandatory to create an electronic environment to safely save them for further analysis.

A comprehensive database is designed to automate data collection and query processes in the scope of the thesis study. Electronic interfaces are created for entering this information to the database. When the database is opened the first form to be filled by the user is the patient record form, which is shown in Figure

2.9. Detailed information about OAGAIT database system is given in fourth chapter of the report.

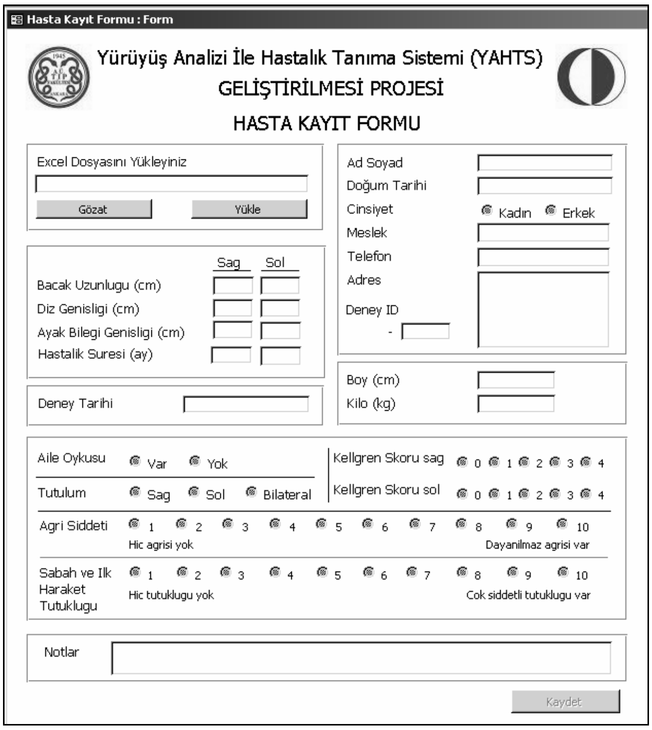


Figure 2.9: Patient recording form

CHAPTER 3

PATTERN RECOGNITION METHODS

3.1. FEATURES AND DIMENSIONALITY REDUCTION

When dealing with high-dimensional data, one faces with the well-known problem of “curse of dimensionality”. The ***curse of dimensionality*** is a term to describe the problem caused by the exponential increase in volume associated with adding extra dimensions to a (mathematical) space **[72]**. In pattern recognition view, the idea of the curse of dimensionality is that high dimensional data is difficult to work with for several reasons **[64, 65]**. Most importantly the need of exponentially increasing number of training samples with dimension. Also, adding more features can increase the noise, and hence the error. There may not be enough samples to get good recognition results. So, dimensionality reduction is a commonly used step before classification in pattern recognition, especially when dealing with very high dimensional feature spaces. The original feature space is mapped onto a new, reduced dimensionality space and the examples to be used by pattern recognition algorithms are represented in that new space. The mapping is usually performed either by selecting a subset of the original features or/and by constructing some new features.

The dimensionality of the feature space may be reduced by the selection of subsets of good features. Several strategies and criteria are possible for searching good subsets. In addition to the improved computational speed, an increase in the accuracy of the classification algorithms is also expected with reduced feature set. **[64].**

Another way to reduce the dimensionality is to map the data on a linear or nonlinear subspace. This is called linear or nonlinear feature extraction. It does not necessarily reduce the number of features to be measured, but the advantage of an increased accuracy may still be gained. Moreover, as lower dimensional representations yield less complex classifiers better generalizations can be obtained **[64].**

There are some significant difficulties in the design of automated medical decision support systems because of the multidimensional and complex structure of the clinical data. So, feature reduction and selection always become an important part of the medical data analysis studies.

3.1.1. Feature Extraction

In most pattern recognition problems the number of samples is smaller than the number of row features due to practical reasons. In that case the actual feature space is mapped to another one having fewer dimensions by minimizing the information lost. There are some widely used mapping algorithms in statistical pattern recognition like FFT, PCA, wavelet etc.

The most commonly used method for the feature extraction in gait classification is based on the estimation of parameters (peak values, ranges) as descriptors of the gait patterns. In that case the classification is done according to the differences between the class averages of the training set and the parameters of the new subjects **[19, 31].** This method is subjective **[31]** and neglects the temporal information of the gait data. There are examples of using statistical feature reduction techniques in gait analysis, such as Fast Fourier Transform (FFT) **[34, 36, 37]**, Principle Component Analysis (PCA) **[19, 24, 31],** wavelet transform **[32]** and averaging **[38, 39].**

FFT:

A **fast Fourier transform** (**FFT**) is an efficient algorithm to compute the discrete Fourier transforms (DFT) and its inverse **[72]**. FFTs are used in great variety of applications like digital signal processing, solving partial differential equation, and quick multiplication of large integers.

Fourier Transform maps a time series into the series of frequencies (their amplitudes and phases) that compose the time series. Applications of Fourier transforms in statistical pattern recognition and image processing include **[72]**:

- Filtering: Since taking Fourier transform of a function means to represent it as the sum of sine functions, eliminating some high/low frequency components and taking inverse Fourier transform produce an image without noises.
- Image Compression: Since a filtered image contains less information than a noisy image, encoding it requires fewer bits to represent than the original image.
- Convolution and Deconvolution: Fourier transforms can be used to efficiently compute convolutions of two sequences.
- Feature reduction for temporal features: FFT coefficients (which is usually less than original number of time samples) are used for classification **[31, 34]**

Let x_0, ..., x_{N-1} be complex numbers. The DFT is defined by the formula

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} n k} \qquad k = 0, \ldots, N-1.$$

(Equation 3.1)

Evaluating these sums directly would take O(N^2) arithmetical operations. An FFT is an algorithm to compute the same result in only O(N log N) operations.

Since the inverse DFT is the same as the DFT, but with the opposite sign in the exponent and a 1/N factor, any FFT algorithm can easily be adapted for it as well.

Averaging:

Averaging methods are similar to mean filtering methods in image processing. Mean filtering is a simple, intuitive and easy to implement method of *smoothing* and scaling images **[71, 72]**. The image is smoothed because the amount of intensity variation between one pixel and the next is reduced. As an example, for the scaling, to halve the size of the image each quad of four pixels is replaced by one pixel with average of the four pixels. This simplest way of scaling images is mostly recommended for downscaling.

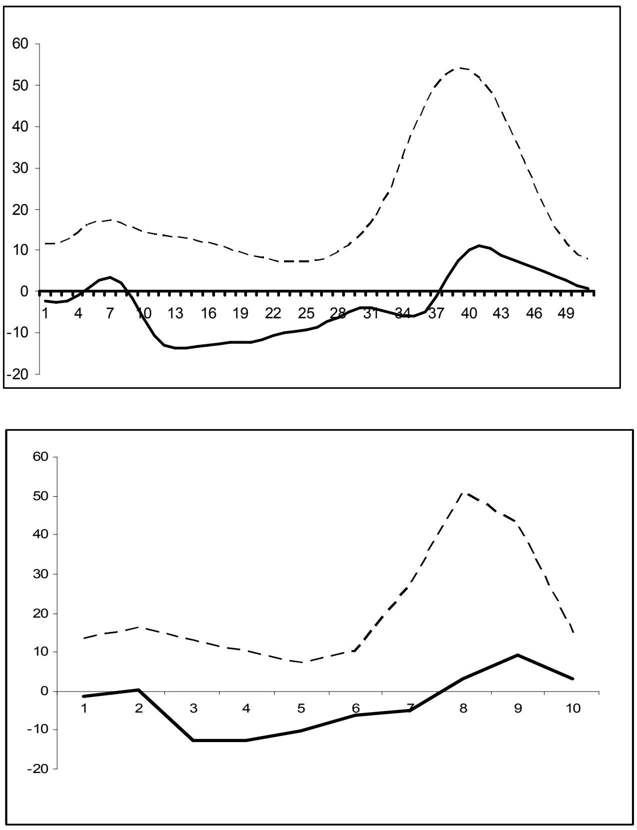


Figure 3.1: Averaging five consecutive time samples of two gait waveforms

This property of the averaging makes it usable for time sample reduction of temporal features. Figure 3.1 shows an example to averaging five consecutive time samples of two gait waveforms. The dimension of the temporal features is reduced from 51 to 10 by averaging method.

3.1.2. Feature selection

Automated selection of the gait attributes is not observed in gait classification literature; medical experts select them or previous studies are taken as references. Actually, there are many medical practices, testing the variations in the gait attributes which are caused by the related illness **[20-30]**. Non-automated selection of the gait attributes are done by using the result of these studies. But it may not always be convenient to work with a medical expert for the feature selection process. Also, the judgments may vary in different experts leading the different interpretations of the classifiers. Obviously, automated selection lessens the dependence and the load on the experts and gives more freedom to the researchers.

Feature selection also helps people to acquire better understanding about their data by telling them that which are the important features and how they are related to each other.

The feature selection process is simple defined as follows: given a set of candidate features, select a subset that performs the best by a given classifier. This procedure can reduce the cost of recognition and in most cases provide better classification accuracy. There are some criterion functions for assessing the goodness of a feature subset. Mahalanobis distance is one of these functions **[64, 68-70].** The selection of the criterion function is very important. If we know which classifier will be used in the problem, then the best criterion is the correct recognition rate of that classifier. However, it is computationally time consuming and very difficult to estimate correct recognition rate of the classifier with a limited number of training samples. This is one of the reasons why Mahalanobis distance, which gives an upper bound of the Bayes error rate with a priori probabilities of classes, is used in many feature selection processes **[69].**

In statistics, **Mahalanobis distance** is defined as distance measure based on correlations between variables by which different patterns can be identified and analyzed. It is a useful way of determining similarity of an unknown sample set to a known one. It differs from Euclidean distance is that; it takes into account the correlations of the data set and is scale-invariant, i.e. not dependent on the scale of measurements **[65, 72]**.

Formally, the *Mahalanobis distance* is defined as

$$D_M(x) = \sqrt{(x-\mu)^T \mathrm{P}^{-1} (x-\mu)}. \quad \textbf{(Equation 3.2)}$$

for a multivariate vector

$$x = (x_1, x_2, x_3, \dots, x_p)^T$$

with mean

$$\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_p)^T$$

and covariance matrix P whose (i, j) entry is the covariance

$$P_{ij} = E\left[(\chi_i - \mu_i)(\chi_j - \mu_j)\right] \quad \textbf{(Equation 3.3)}$$

where $\mu_i = E(x_i)$ is the expected value of the i^{th} entry in the vector X

Mahalanobis distance can also be defined as dissimilarity measure between two random vectors $\vec{x}$ and $\vec{y}$ of the same distribution with the covariance matrix P:

$$d(\vec{x},\vec{y}) = \sqrt{(\vec{x}-\vec{y})^T \mathrm{P}^{-1}(\vec{x}-\vec{y})}. \qquad \textbf{(Equation 3.4)}$$

If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Euclidean distance. If the covariance matrix is diagonal, then the resulting distance measure is called the *normalized Euclidean distance*:

$$d(\vec{x},\vec{y}) = \sqrt{\sum_{i=1}^{p} \frac{(x_i - y_i)^2}{\sigma_i^2}}, \qquad \textbf{(Equation 3.5)}$$

Where, σ_i is the standard deviation of the x_i over the sample set.

3.2. CLASSIFIERS

Here, we will discuss only the ones that are used in this study.

3.2.1. Tree Classifiers

Decision Tree Classifiers are used successfully in many diverse areas such as radar signal classification, character recognition, remote sensing, medical diagnosis, expert systems, and speech recognition. Perhaps, the most important feature of decision trees is their capability to break down a complex decision-making process into a collection of simpler decisions, thus providing a solution which is often easier to interpret **[61].**

As an example, the decision tree in Figure 3.2 is constructed to decide whether the weather is convenient for playing tennis or not. The weather attributes are outlook, temperature, humidity, and wind speed and the target classification is “yes” or “no”.

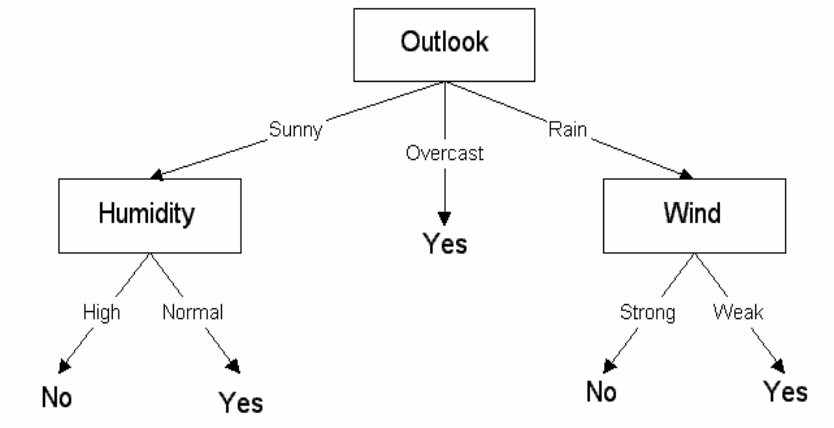


Figure 3.2: An example to decision tree classifier

Comparing to other classification methods the most advantageous differences of decision trees are **[62]:**

- They produce understandable tree-structures which clarify the reasoning of the method (many other techniques lack this and are harder to interpret)
- They can show the problem in as a disjunction of the hypotheses.
- They can be faster in the average than many other approaches.

A classification tree or a decision tree is an example of a multistage decision process. Instead of using the complete set of features, subsets are used at different levels of the tree. Three important properties of the decision trees are **[42]:**

- Decision trees are instable classifiers. Means they are capable of memorizing the training data so that small changes in data might create a different structure tree. Instability can be an advantage when ensembles of classifiers considered.
- Since decision process can be traced as a sequence of simple decisions, tree classifiers can be defined as intuitive. Tree can capture a knowledge

base in a hierarchical way; most popular examples are botany, zoology and medical diagnosis.

- Both quantitative and qualitative features are suitable for building decision tree classifiers. Binary features and features with a small number of categories are useful because the decision can be easily branched out. Since decision trees are not based on the distances in the feature space they are regarded as nonmetric methods for classification.

A decision tree construction starts with the root and continues separating the parts of the data to child nodes which is called *splitting the tree*. Splitting into small parts continue until a termination criterion is met. A termination criterion may be that all objects be labeled correctly. In this case the tree has to be pruned to prevent overtraining.

One can reach from root of the tree to the final class label by asking small number of questions at nodes (root is the top node of the tree) of the tree. Depending on the answer a branch is selected and the related child node is visited. Another decision is made at next node and the process continues until reaching to a leaf (a terminal node). This leaf shows a class label which can be repeated at other nodes. If the same number of branches are visited to reach to every leaf of tree then tree is called balanced. Otherwise it is called imbalanced.

Figure 3.3 shows examples to the balanced and unbalanced trees. Imbalanced trees indicate that objects near the classification boundaries may need longer decision chains than the others **[42]**.

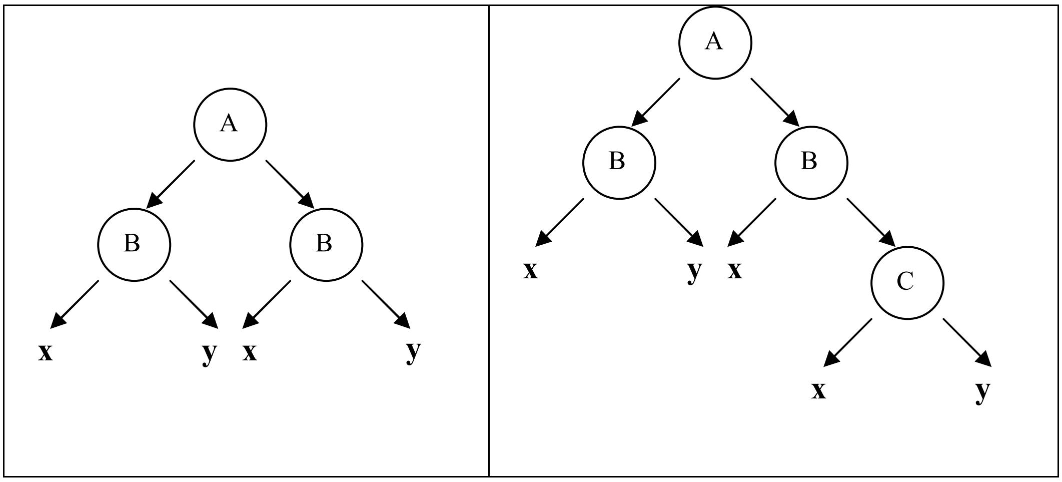


Figure 3.3: Examples to a) balanced and b) unbalanced trees

Splitting Criteria

Consider a c-class problem with $\mathbf{\Omega} = \{w_1, w_2, ..., w_c\}$. Let $\mathbf{P_j}$ be the probability for class $\mathbf{w_j}$ at a node $\mathbf{t}$. These probabilities can be calculated by the proportion of the points from related class within the whole dataset at that node. The impurity of the distribution of the class labels at $\mathbf{t}$ can be measured in different ways.

Entropy based Impurity:

$$i(t) = -\sum_{j=1}^{c} P_j \log P_j$$ **(Equation 3.6)**

According to this formula impurity takes its minimum value when only one class label exists at the node (0 log 0 = 0). The most impure situation occurs when the classes have uniform distribution. In that case *i(t)*=log c

Gini Impurity:

$$i(t) = 1 - \sum_{j=1}^{c} P_j^{\,2}$$ **(Equation 3.7)**

For the most pure case again *i(t)* = 0. The highest impurity in the case of uniform distribution is *i(t)* = (c-1)/c. The Gini index can be defined as the expected classification error when a random class label is chosen from the distribution of the labels at **t**.

Misclassification Impurity:

$$i(t) = 1 - \max_{j=1}^{c}\{P_j\} \qquad \textbf{(Equation 3.8)}$$

Misclassification impurity gives the expected error if the node was replaced by a leaf and the chosen label was the corresponding label of the largest P_j .

Gain:

Assume that the tree is split into child nodes based on the feature X. Then the gain in splitting the t is defined as:

$$\Delta i(t, X) = i(t) - \sum_{v \in X} \frac{t_v}{t} i(t_v) \qquad \textbf{(Equation 3.9)}$$

If the features are binary then splitting is easier; try each one in turn and choose the feature with highest gain. However is the features are multiple categories or continues valued, then an optimal threshold to split node is have to be found.

Among the above methods Gini index is the mostly used one **[65]**. The choice of the impurity index is not seem to be very important for the success of the tree classifier **[65]**. The more important issues are stopping criteria and the pruning methods.

Stopping Criterion

The tree construction can be continued until there are no impure nodes. But this time overtraining may be a problem for the test dataset. So the training should be stopped before reaching pure nodes. But if the splitting is stopped too early the tree may be under-trained. In **[65]** some options are listed to avoid this problem:

- Use a validation set
- Set a small impurity-reduction threshold. When the greatest possible reduction of impurity is less than or equal to this threshold stop splitting. But the problem here is to determine this threshold value.
- Set a threshold values for the number of point at a node
- Use hypothesis testing to see whether a one more split is beneficial or not

Pruning methods

Sometimes early stopping can prevent further beneficial splits. This phenomenon is called horizon effect **[65]**. To avoid this one can construct full tree and then prune it to a smaller size. The aim of pruning is to optimize training error and the size of the tree.

Reduced error pruning is the simplest pruning method. An additional training set (pruning set) is used for a simple error check at all non-leaf node. A node is replaced with a leaf and labeled to the majority class. The error of the tree on the pruning set is calculated and compared to the error of the first tree. If the new error is smaller than the previous one, the node is replaced with the leaf. Otherwise the sub-tree is kept **[63].**

In pessimistic error pruning method the same dataset is used for both constructing and pruning the tree. If the number of errors at node is smaller than the number of errors with a complexity correction at sub-tree of that node, then the node is replaced by a leaf **[63]**.

In critical value pruning a critical value is set as a threshold. The tree is checked in a bottom-up fashion. If a node has a gain in error rate smaller than the critical value is replaced by leaf **[63]**.

In **[63]** two more pruning methods are defined: Cost-complexity pruning and error-based pruning. According to this study critical value pruning and error based pruning have tendency towards over pruning where as reduced error pruning has opposite trend. Also it is concluded that, using an aside pruning set does not always work. Methods used the whole trading set to construct and prune the tree are found to be more successful.

3.2.2. Neural networks

Artificial neural network (ANN), which is often called Neural Network (NN), is a mathematical model created by inspiration of biological neural networks. A NN is created by artificial neurons by connecting them to each other in different fashions. The information flows through these connections and update the structure of the networks. So NNs are defined as adaptive systems.

In pattern recognition view, NNs are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data.

Neurons

The basic schema of a neuron is shown in Figure 3.4.

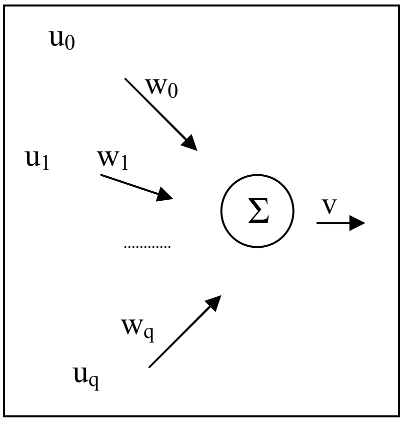


Figure 3.4: A simple neuron model

Let $u = [u_0, u_1, \ldots u_q] \in R^{q+1}$ be the input vector, $v \in R$ be its output.

The vector $w = [w_1, w_2, \ldots w_q]^T \in R^{q+1}$ is synaptic weights. Where

$v = \varphi(\varepsilon)$ and $\varepsilon = \sum_{i=0}^{q} w_i u_i$ where φ is the activation function and ε is the net sum.

The activation function may be hard-limit, linear or sigmoid function. The sigmoid function is the most used one, because;

- It can model both linear and hard limit (threshold) functions. It is almost linear near the origin and hard limited for large weights.
- It is differentiable, which is important for the training algorithms.

Perceptron

The simplest kind of neural network is a *perceptron* network, which consists of a single layer of output nodes. The inputs are given directly to the outputs via related weights. The sum of the products of the inputs and weights are calculated and the output is produced according to threshold activation function.

This one-neuron linear classifier can separate two classes. The weights are initialized randomly and modified as each sample is subsequently presented to the inputs of the perceptron. The modification occurs only if the current ample is misclassified. Perceptron training has following properties **[42]**:

- If the classes are linearly separable the algorithm always converges in a finite number of steps.
- If the classes are not linearly separable the algorithm will enter to a loop and never converges.

Multilayer Perceptron

By connecting two or more perceptron one can construct a Multilayer Perceptron (MLP). MLP has a feedforward structure, means all units in input layer and hidden layers are submitted to the only higher layer. A generic example of a MLP is shown in Figure 3.5.

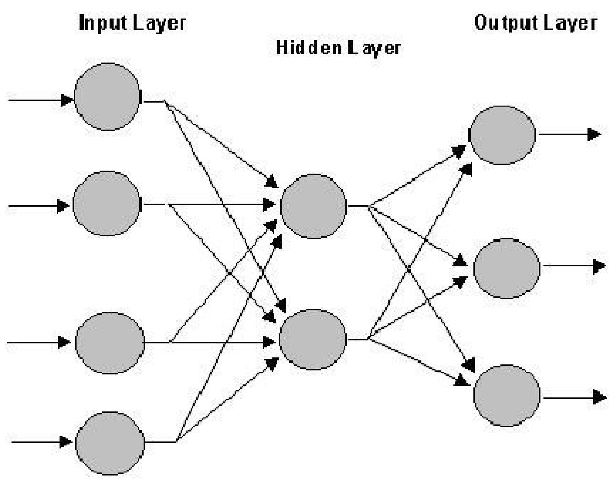


Figure 3.5: An example to multilayer perceptron with one hidden layer.

The number of hidden layers and number of nodes is not limited, but there are lots of studies to find best numbers. In late 80s it was shown that an MLP with two hidden layers with threshold nodes can approximate any classification problem **[42, 65].** In F

Figure 3.6 classification regions that could be constructed by one, two and three layers are shown **[42]**. Later, it is proven that even an MLP with single hidden layer can approximate any function **[42]**.

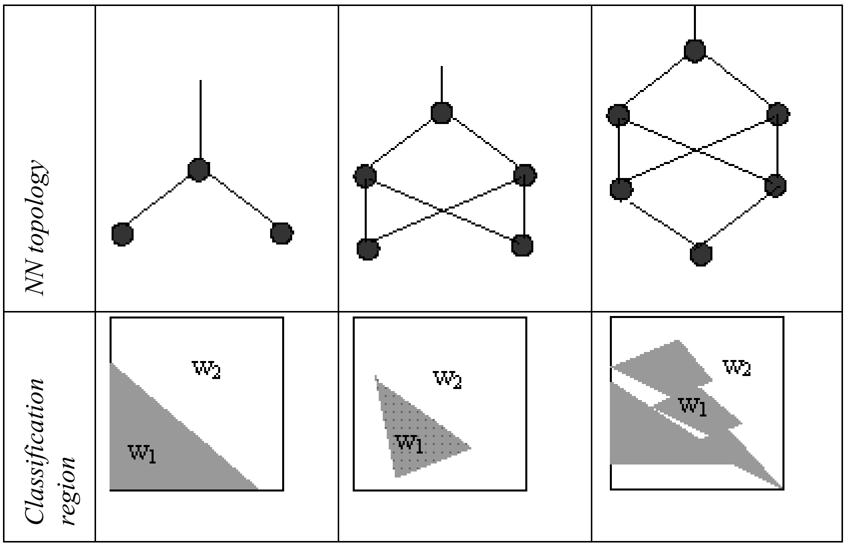


Figure 3.6: Examples to classification regions by one, two and three layer MLPs [42]

Most common properties of MLPs are **[42, 65]**:

- The activation function at input layer is the identity (linear) function
- There are no connections between the nodes at same layer (feedforward)
- There no connection between the nodes at nonadjacent layers
- All the nodes at all hidden layers have the same activation function

Multi-layer networks use a variety of learning techniques; the most popular of these is backpropagation algorithm. For training of the neural networks, the output values are compared with the correct answer to compute the error function. Then this error is fed back to the network by various methods. Algorithm adjusts the weights of each connection in order to reduce the error some small amount. To adjust weights a general method for non-linear optimization that is called gradient descent is used. For this, the derivative of the error function with respect to weights is calculated and the weights are updated to decrease the error. So the activation function of the network applying backpropagation should be differentiable. Repeating this process for a sufficiently large number of training cycles the network will usually converge to some state where the error is small. The network converged this final state is said that it has learned the target function.

If a network is trained by very limited number of training samples it can overfit the data. So the network can not perform well on test data set. Some special methods should be applied to avoid overfitting. Other problems of network training are speed of the convergence and stopping convergence in a local minimum. This causes networks to take a non-optimum state. Decreasing or increasing the number of hidden layers and nodes may prevent the local minima problem. Rerunning the algorithm may also work because the weights will be reinitialized to a different numbers.

3.2.3. Combining Classifiers

The concept of combining classifiers is proposed as a new direction for the improvement of the performance of individual classifiers. These classifiers can be based on a variety of classification methodologies, and could achieve better rates than individual classifiers. The goal of classification result integration algorithms is to generate more certain, precise and accurate system results. Dietterich **[55]** provides an accessible and informal reasoning as shown in Figure 3.7 from statistical, computational and representational viewpoints, of why ensembles can improve results.

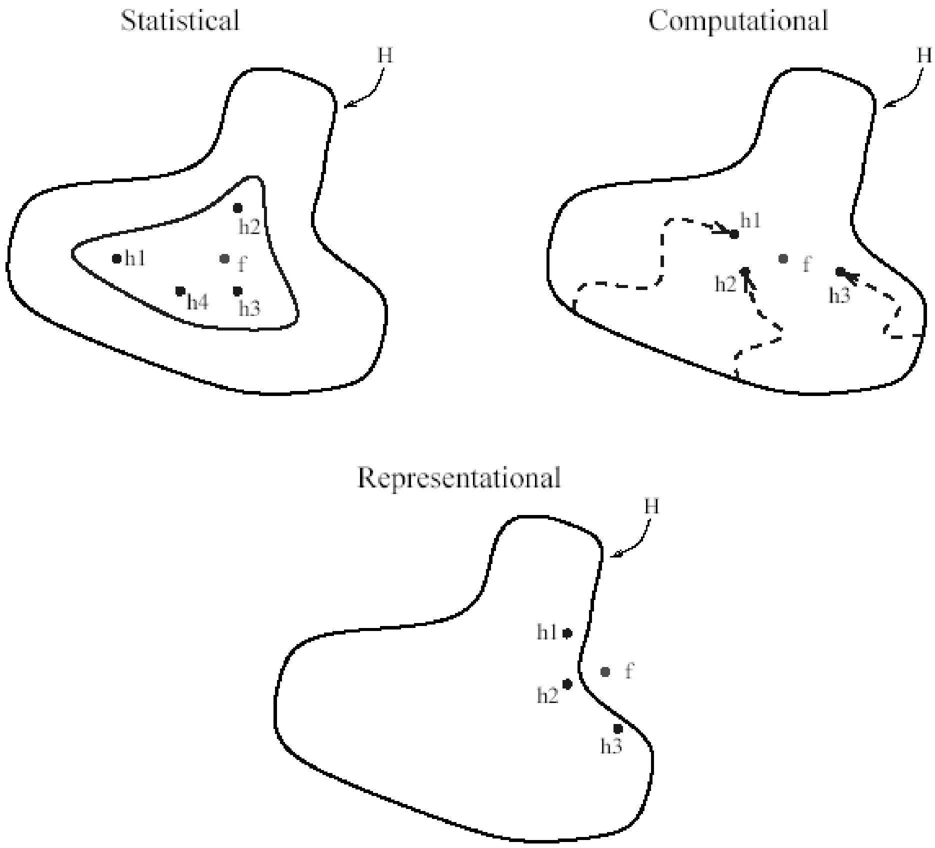


Figure 3.7: Reasons why an ensemble classifier may be better than an individual one [55]

- Statistically: Instead of selecting a single classifier one option may be use them all and average their outputs. The new classifiers may not be better than the single best classifier but the risk of selecting an inadequate classifier is eliminated.

- Computational: Assuming the training process of each classifier start somewhere in the space and end closer to the best one (f), combining them may cause to better approximation than a single classifier.

- Representational: Training an ensemble of simple classifiers to achieve a high accuracy is more straightforward than training a single more complex classifier.

There are several ways of creating multiple classifier system. In **[73]** three broad categories are defined.

Different feature spaces: This describes the combination of a set of classifiers, each designed to use different feature spaces. For example in a person verification application, several classifiers may be used for different sensor data like retina scan, facial image etc.

Common feature spaces: This describes the combination of different classifiers trained on the same feature space. The classifiers can differ from each other in some ways.

- The classifiers may be of different type, for example nearest neighbor, neural network, decision tree etc.
- They may be similar types but use different part of training set.
- They may be similar type but use different initialization parameters, for example weight initialization of neural networks.

Repeated measurements: This category of combination is about different classification of an object through repeated measurements.

3.2.4. Combination Schema

Combination schemas may be classified according to some characteristics including, level of combination, structure, form of classifiers and training styles.

Level of combination

Combination may be done at different levels as suggested by Kuncheva **[42]** in Figure 3.8.

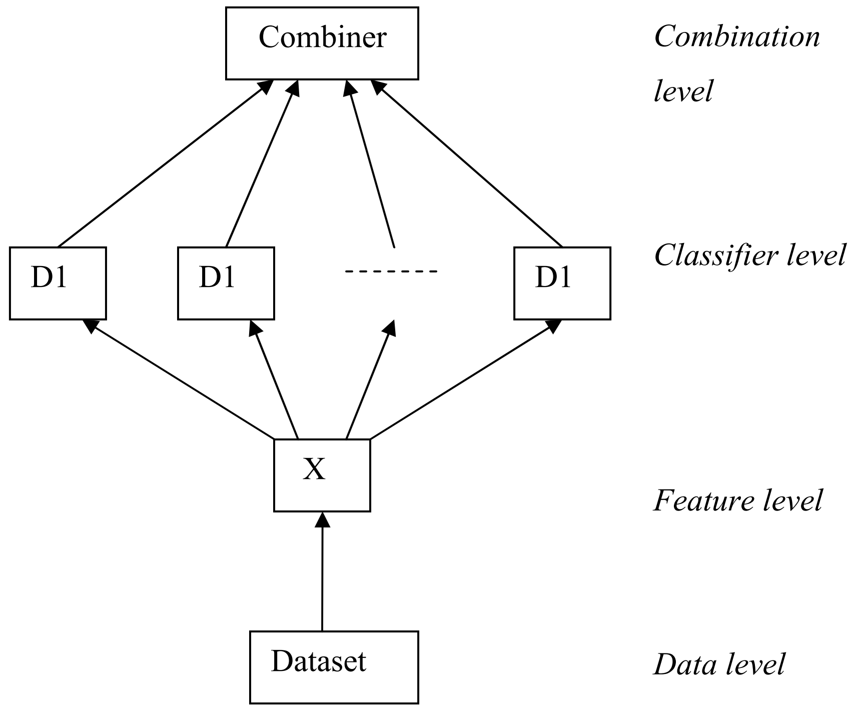


Figure 3.8: Approaches for building ensemble classifiers (differentiation at each level may be considered as an approach) [42]

Data level: Raw measurements are given to the combiner that produces posterior probabilities of class membership. This requires defining a classifier on all sensor variables. Different datasets may be created by also different preprocessing methods.

Feature level: Each of features may have its own techniques for reducing dimension. Different classifiers may perform some local preprocessing on feature subsets.

Classifier level: Using different base classifiers is a common approach in combination schemas. Different base classifiers may be preferred due to various structure of the feature set. Another approach may be differentiating the same classifier by changing some parameters of it, for example changing initializing or training parameters of a neural network.

Combination level: Most important issue in combining classifiers is the way they are combined. According to **[43, 71]** there are two types of combining rules, trained and fixed rules. Trained combiners are different than fixed combiners in methods of producing final decision. After gathering outputs of the base classifiers, they are used as an input vector of the combining classifier. The training set is used for both training base classifiers and combining classifier.

The fixed combining rules make use of the fact that the outputs of the base classifiers are not just numbers, but that they have a clear interpretation: class labels, distances, or confidences. The confidence is sometimes interpreted or generated by fuzzy class membership functions sometimes by class posterior probabilities. Majority vote, product rule, sum rule, maximum rule are examples of fixed rules.

For example if only labels are available a majority vote is used **[44]**. Sometimes label ranking may be preferred. If continuous outputs like posterior probabilities are gathered, a linear combination like sum or average may be used. Moreover, it is possible to train a classifier with the output of another as new features **[43]**.

The structure of a multiple classifier system may be discussed by three style **[71]**. In parallel combining, results from the base classifiers are passed to the combiner together. In serial combining the base classifiers are invoked sequentially. In hierarchical combining, the classifiers are combined in a hierarchy, with the outputs of one base classifier are given to another one as inputs, in a similar manner to decision trees.

CHAPTER 4

FEATURE SELECTION AND CLASSIFICATION (GRADING)

4.1. INTRODUCTION

In this study the grading algorithm is designed as a combination of different types of classifiers, which allows including all available data in decision making process. Before finalizing this form of the grading algorithm many experiments have been implemented by different classifiers, different feature reduction/selection methods. The results of these experiments were helpful for design of the final algorithm.

- In first trials, the gaits of 111 patients with 110 age-matched normal subjects are compared. Two different feature reduction techniques, FFT and averaging are compared by performances of some well known pattern classifiers. The MD measure is used as an individual feature filtering criterion and most discriminatory features are determined. Then a set of linear and non-linear classifiers is tested by a ten-fold cross validation approach. The details of this trial are given in Section 5.2.
- In second experiment, two popular methods of combining neural networks are implemented for discrimination of normal and sick patterns. The

results of classifiers are compared with different output combining rules. The details of this trial are given in Section 5.3.

- Finally, a decision tree MLP combination is implemented for grading of the knee OA. Automated feature selection is used for composing datasets to train MLPs responsible for discriminating neighbor classes. Last section of this chapter is about design and implementation stages of this algorithm.

4.2. STATISTICAL ANALYSIS OF GAIT DATA: COMPARISON OF FEATURE REDUCTION/SELECTION AND CLASSIFICATION ALGORITHMS

The objective of this experiment is to compare the convenient methods for preprocessing (feature reduction/selection), classification and further analysis (such as learning curves of the classifiers) of the gait data. For this purpose two feature reduction techniques (averaging and FFT) are compared by performances of some well known pattern recognition classifiers. The MD method is used as an individual feature selection criterion and the most discriminatory features are determined automatically and the selected attributes are compared with the ones suggested by previous OA classification studies to discuss the differences of automated and non automated selection procedures. Next, a set of linear and non linear classifiers is tested on datasets with different dimensionalities by a crossvalidation approach. Finally, learning curves of some classifiers are compared to discuss data size issues (the number of subjects in the training set) for further studies.

The classification and feature selection algorithms are used from PRTools which is a Matlab based toolbox for pattern recognition. PRTools supplies about 200 user routines for traditional statistical pattern recognition tasks. **[67]**.

4.2.1. Feature reduction and selection methods

In this experiment, the data that were formerly collected in Ankara University Faculty of Medicine gait laboratory from 110 normal and 111 OA patients are used. All joint angles features from both kinetic and kinematic domains are included in feature reduction and selection processes.

Data collection process produces a dataset for each subject, including 33 gait attributes, each having 51 sample points in time, as explained before. Combining these files into a complete dataset, we got *33 (attributes) x 51(time samples)* dimensional arrays for each subject. The final dataset is thereby composed of 221 subjects presented by 1653 points in feature space. Since the total number of the features is too large relative to the number of subjects, most of the commonly used classifiers will suffer from the curse of dimensionality **[64, 65]**. So, a reduction in the number of features is needed before the classification process. Two different reduction techniques are applied to the same dataset for comparison. Six datasets of the different dimensionalities are composed by averaging consecutive time sample points for reducing the size of the feature vectors. Also, FFT is applied to each waveform and each attribute is represented by one, five, ten or 25 FFT coefficients rather than 51 time samples. At the end of the reduction process, ten datasets of the different dimensionality are created.

Most of the datasets still have a too high dimension, which forces elimination of the redundant features. Since the term “features” here represent the time samples of the gait attributes, they have no meaning by themselves. So the selection is done by the gait attributes, not by the features. The MD measure is used as a selection criterion (detailed information about MD based feature selection is given in Section 3.1.2) Individual performances of the each gait attributes to discriminate two classes are compared. Instead of the individual selection, a forward or a backward selection can also be tried in order to avoid the selection of the similar attributes. However, since the data is from different motion planes and different anatomic levels of the body, to have similar attributes is not very probable.

Table 4.1 summarizes how these datasets are created. The values on this table represent the MD values produced by each gait attribute with shown number of time samples. The marked values show the selected attributes for creating corresponding dataset. All of one dimensional attributes are used for dataset creation.

The number of selected attributes is limited to reach the best ratio of the number of subjects and the number of features, which is suggested as one over five in **[64].** Except for the all-mean datasets, the dimensions of the datasets are fixed to 50, since it is reachable by integer number of the attributes for all datasets and about the ideal ratio (which is 44,2 here). For feature selection *distmaha* property of the PRTools to calculate MDs of classes in a dataset is used, as an example code segment is shown in Table 4.2. Table 4.3 shows the properties of the created new datasets by combining the selected best features and the MD values produced by these new datasets.

Table 4.1: Mahalanobis Distance values of gait attributes

Gait	Averaged datasets						FFT applied datasets			
Attributes	Best51d	Best25d	Best10d	Best5d	Best2d	Best1d	BestFFT25	BestFFT10	BestFFT5	BestFFT1
PTilt	6,09	4,93	3,60	**3,20**	**2,85**	2,85	1,76	0,67	0,32	2,20
PObliq	5,28	3,76	2,48	1,72	**0,45**	0,01	2,35	1,73	1,09	0,23
Prot	4,05	2,71	1,99	**1,89**	0,05	0,00	2,07	1,37	1,23	0,00
HFlex	10,26	7,32	**5,70**	**2,00**	**1,14**	0,92	1,84	1,32	0,83	1,01
HAbd	10,63	7,63	**5,92**	**3,10**	**1,27**	1,19	2,81	1,67	1,14	1,40
HRot	6,79	4,40	2,99	1,86	**0,90**	0,64	3,37	2,54	**1,83**	0,02
KFlex	**14,57**	**9,95**	**8,36**	**3,49**	**3,21**	0,83	5,26	**4,33**	**4,25**	0,83
KVal	6,68	3,83	2,97	**2,37**	**1,89**	0,37	2,81	2,02	**1,65**	0,13
KRot	8,79	6,35	5,24	**2,98**	**2,14**	0,02	4,52	**3,66**	**2,56**	0,10
FDor	6,80	4,59	3,24	**2,32**	**0,83**	0,62	1,84	1,22	0,86	0,49
FRot	6,30	4,32	3,33	1,44	**1,34**	0,95	2,47	1,74	1,28	0,66
FPro	5,77	2,93	1,77	1,34	**0,74**	0,71	0,69	0,25	0,07	0,70
HMFlex	5,64	4,08	1,96	0,82	0,11	0,04	4,50	2,06	**1,75**	0,13
HMAbd	10,40	**8,26**	**6,38**	**5,92**	**3,49**	2,91	**7,39**	**6,34**	**5,56**	2,92
HMRot	5,57	3,02	1,72	0,98	**0,31**	0,16	3,96	1,63	1,41	0,80
KMFlex	10,24	7,98	**5,72**	0,77	**0,44**	0,40	**5,37**	**3,36**	**2,78**	0,19
KMVal	9,08	6,37	3,57	**2,92**	**1,05**	0,46	4,86	**3,47**	**2,64**	0,61
KMRot	6,89	4,15	2,33	1,08	**0,51**	0,17	2,16	0,43	0,39	0,23
FMDor	8,68	5,70	3,13	1,20	**0,32**	0,22	2,72	1,77	1,49	0,22
FMAbd	4,12	2,67	1,58	1,21	0,29	0,25	2,43	1,72	0,89	0,21
FMRot	5,06	2,97	2,10	1,22	**0,93**	0,66	1,78	1,15	1,06	0,66
HPTot	3,21	2,08	1,27	0,35	0,15	0,02	2,50	1,24	0,70	0,01
HPFlex	3,36	1,66	1,01	0,47	**0,55**	0,04	2,77	1,59	0,81	0,16
HPAbd	6,25	3,74	1,49	0,91	**0,38**	0,16	3,30	1,93	1,31	0,08
HPRot	3,98	2,85	1,26	1,01	**0,56**	0,22	2,23	1,43	0,91	0,38
KPTot	8,52	6,12	5,05	1,23	**1,05**	0,56	4,68	3,31	**3,19**	0,89
KPFlex	9,89	6,80	5,50	1,76	**1,36**	0,85	4,21	2,88	**2,65**	1,13
KPVal	5,78	4,25	2,40	0,61	0,09	0,09	1,67	0,85	0,39	0,00
KPRot	3,42	2,11	1,39	1,06	0,28	0,25	2,28	0,70	0,53	0,02
APTot	5,92	3,73	2,38	0,90	0,16	0,15	3,36	2,19	1,51	0,01
APDor	6,22	4,68	2,73	0,64	**0,38**	0,29	3,38	1,49	0,81	0,00
APAbd	1,98	1,11	0,84	0,70	**0,41**	0,05	1,15	0,74	0,66	0,00
APRot	2,18	1,38	0,89	0,38	0,27	0,01	1,42	0,53	0,33	0,00

Table 4.2: An example to feature selection by distmaha function of PRTools

`for i = 1 to #gait_attributes`
`1. x = featfft5 (:,:,i);`
`2. y = dataset (x, labs);`
`3. Dm = distmaha (y);`
`4. Dmfft5(i) = Dm (1,2);`
`End`
	1. x is an array composed by first 5 fft coefficients of the ith attribute
	2. Convert x to dataset y by presenting class labels array labs
	3. Dm is a 2x2 symmetric matrix where Dm(i,j) represent the Mahalanobis Dist. of classes are written to a one-dimensional array
	4. Mahalanobis Distance of classes i and j of dataset y.

Here the naming procedure for datasets are based on the number of time samples and feature reduction criteria, for example "BesFFT10" represent the dataset that the time sample reduction is done by FFT algorithm and each attribute is represented by 10 time samples.

Table 4.3: Datasets after feature reduction and selection processes

Composed dataset	# time samples to represent gait attributes	# selected gait attributes	Dimension of the dataset	*Mahalanobis Distance*
Best51d	51	1	51	14.568
Best25d	25	2	50	18.273
Best10d	10	5	50	23.545
Best5d	5	10	50	19.804
Best2d	2	25	50	20.387
Best1d	1	33	33	14.577
BestFFT25	25	2	50	13.244
BestFFT10	10	5	50	15.589
BestFFT5	5	10	50	18.029
BestFFT1	1	33	33	10.625

As Table 4.3 shows, datasets with about the same dimensionality are composed of different numbers of gait attributes represented by different numbers of time samples. So at the end of the classification process it will be possible to discuss whether the number of gait attributes is more important than the number of time samples for classification accuracy.

Comparing two reduction techniques based on the MD criterion, averaged datasets perform better than the corresponding FFT based dataset. While composing these new datasets, the MD values of the attributes are ordered and the required number of the best of them is added to the new dataset. So, while some of the gait attributes may appear in many of the datasets, some may not. As the table shows, two of ten datasets are created by including all attributes and eight of them are created by selecting the best ones. The number of appearance times of attributes in eight datasets are used to compare the discriminatory ability of them for the classification of the gait patterns of OA patients (only 4 and above are showed). Appearance times of the gait attributes in eight datasets are:

- KFlex (Knee Flexion): 7
- HMAbd (Hip Abduction Moment): 7
- KMFlex (Knee Flexion Moment): 5
- KRot (Knee Rotation): 4
- KMVal (Knee Valgus Moment): 4

In other trial of the study, the gait attributes were selected by an expert physician who has suggested four knee-related attributes **[38]**. All of these four attributes are appeared in most of the current datasets, too, and moreover three of them are included in the above table as the most apparent attributes (KFlex, KMFlex, KMVal). It can be concluded that our selection criterion approximates the expert knowledge and so contributes to the validation of the approach.

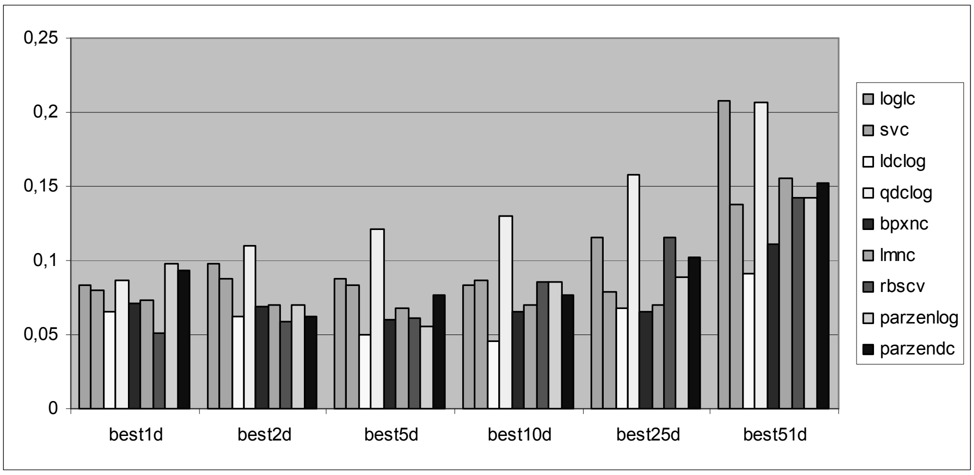
4.2.2. Comparing Classifiers

As an initial study for classifier selection, s set of linear and nonlinear classifiers is tested by a ten-fold crossvalidation method. PRTools **[67]** is used for classifier construction. Total of nine classifiers are used by some adjustments, for more information see also **[64, 65, 67, 71]**.

1. Logistic Linear Classifier (loglc)
2. Support vector classifier (svc)
3. Linear Bayes Normal Classifier (ldc): *log* function is used for adjustment
4. Quadratic Bayes Normal Classifier (qdc)
5. Back-propagation trained feed-forward neural net classifier (bpxnc): 1 hidden layer with 5 nodes
6. Levenberg-Marquardt trained feed-forward neural net classifier (lmnc): 1 hidden layer with 5 nodes
7. Automaric radial basis SVM (rbsvc)
8. Parzen Classifier (parzenc): Datasets are scaled and *log* functions are used
9. Parzen density based classifier (parzendc): Datasets are scaled

Since density based classifiers (ldc, qdc, parzenc) suffer from a low numeric accuracy in the tails of the distributions 'log' function is used to compute log-densities of them. This is needed for overtrained or high dimensional classifiers. Almost zero-density estimates may otherwise arise for many test samples, resulting in a bad performance due to numerical problems. Loglc, and SVC, are other linear classifiers added to set. Loglc is a linear classifier that maximizes the likelihood criterion using the logistic (sigmoid) function. SVC is a linear support vector classifier maximizing the distance between support vectors of two classes.

For neural network classifiers (lmnc, bpxnc), defaults for the numbers of hidden layers (one) and hidden nodes (five) are used, and the optimization of weights is done by the Matlab Neural Network Toolbox. Other nonlinear classifiers added to set are rbsvc and parzendc. rbsvc is a support vector classifier having a radial basis kernel. Parzendc is a density based classifier using different kernels for the density estimated for each of the classes. Figure 4.1 shows the error rates of these classifiers for the averaged and FFT based datasets, respectively.



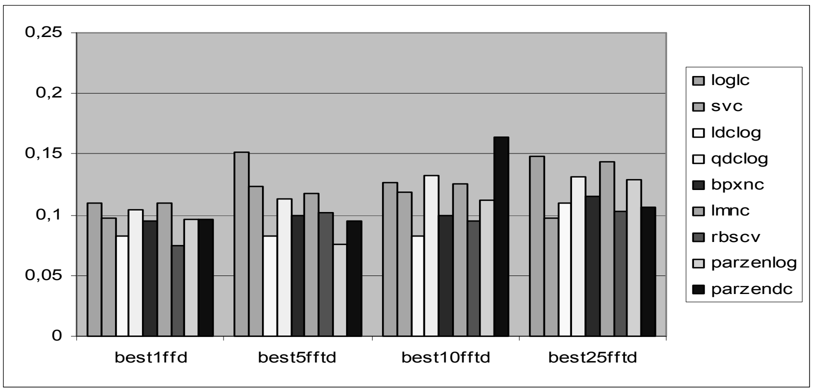


Figure 4.1: Error rates of the classifiers for the a) averaged datasets b) FFT applied datasets

As can be seen in the figure the averaged datasets perform better than the ones composed of FFT coefficients. One of the best datasets (best5d) is selected for further analysis of the gait data.

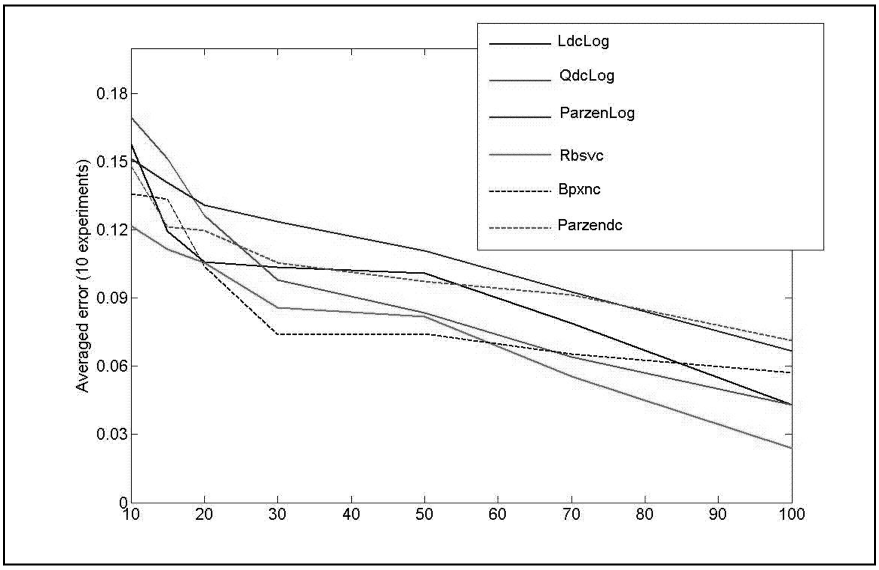


Figure 4.2: Learning curves for some classifiers

4.2.3. Results of comparisons

The results of the experiment performed in this study are important for setting a base for further study of OA grading. Starting from the beginning of the study two different feature reduction techniques are compared first by the MD criterion and then by performances of the classifiers. It can be observed that datasets created by FFT techniques produced worse results in MD calculations and also in the classification process. Even the datasets composed by averaging all time samples produced reasonable classification rates, which makes clear how important activities in the measured angles are for diagnosis. Temporal information of the waveforms is not so significant for the classifier performances. Since FFT coefficients represent the changes of the data in time, the higher error rates for these datasets support the derived result. Also, the severe difference between the performances of the datasets, best1d (all gait attributes with one time sample), and best51d (one gait attribute with all time samples) shows that including more gait attributes is more informative than including more time samples.

We found a high match between currently selected features on the basis of the MD and the ones suggested by gait analysis expert. In the current study, besides the knee features a hip related feature (HMAbd) appeared to perform as well as the best knee related feature (KFlex). The high discriminatory ability of this feature shows that the knee OA causes high variation in hip abduction moments as much as in knee flexion moments of the patients. Moreover, the dataset selected as the best one (best5d), includes data about the pelvic besides the hip and knee related ones (PTilt, PRot, HFlex, HAbd, KFlex, KVal, KRot, FDor, HMAbd, KMVal). To be able to find variation in all levels and motion planes of the subjects automatic feature selection may be preferred.

Comparing the performances of the classifiers on the basis of the current number of subjects, it may be concluded that nonlinear classifiers performed quite well and better than the linear ones. We have compared the learning curves of the classifiers to investigate whether more data might be helpful. Figure 4.2 shows, Backpropagation Neural Network (bpxnc) and Parzen density based (parzendc) classifiers converge faster than the others. Therefore, more data may increase the performances of the linear classifiers more than the nonlinear ones. We have also observed that high regularization prevents linear classifiers learning from more data. Considering the training costs of the algorithms, linear classifiers with a convenient regularization rate may be included in the further studies with more data.

These experiments showed us that statistical pattern recognition algorithms produce promising results for automated analysis of the gait data.

4.3. COMBINING MLPS FOR GAIT CLASSIFICATION

The objective of this part of study is to design a classification algorithm for discrimination of normal and sick gait patterns. The accuracy of the proposed system will be safeguarded by using all gait features. To be able to combine all features in one classification system, combination methods are expected to be most suitable. As our previous studies and similar studies proved MLP usage for gait classification produces reasonable results.

As the dimension of the features and the size of the data increase same accuracies may not be guaranteed. In similar pattern recognition studies this problem is tried to be solved by combining classifiers. Combination of NNs **[45-51]** are widely used today especially in speech recognition and character recognition studies and they have showed an increase in the performance of the classifiers. In **[48]** Sharkey made a comprehensive experiment to compare two different NNs combining methods; modular and ensemble ones. She concluded that using an entire set for training produces more accurate results than decomposing it. In this study comparison of these two approaches are done in the context of gait classification. There are also different approaches on combining outputs of classifiers. In **[43]** the authors have comparative studies on efficiency of output combination rules such as majority voting, sum, product, max., and min. rules. In **[43]**, they concluded that sum rule is superior to others in most of the cases.

A group of MLPs are used to classify the subjects as healthy or sick, using temporal changes of knee joint angle and time-distance parameters as features. Two different NNs combination methods are tried. In the first experiment data set is decomposed into five different sets and five MLPs are trained and tested by these sets. Then test set results are combined by sum, majority vote and max rules to produce final class label. In the second experiment, entire data set is used to train three different architectural MLPs and again outputs are combined by three different rules and accuracy rates on test set are compared.

4.3.1. Dataset Properties

In this study, decision of which gait attributes to use is done by medical expert, and four knee related attributes are selected; knee flexion, knee flexion moment, knee valgus moment and total knee power graphs of which are shown in Figure

4.3. In addition, walking velocity, single support and step length are selected as the time-distance parameters of the gait.

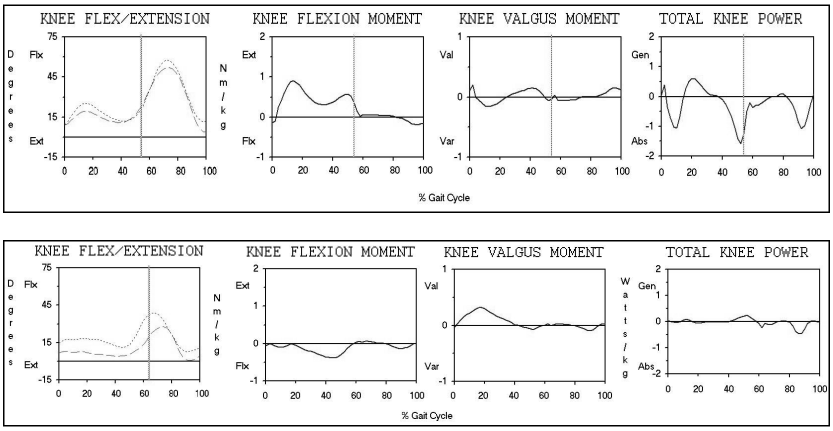


Figure 4.3: **Graphs of the gait data (*healthy (a) and knee osteoarthritis (b)*)**

Each of joint angle related features are represented by a graph that contains 51 samples taken in equally spaced intervals in the time for gait cycle, which is the time spent for one step. These points composed feature vectors which are used as inputs of the related MLP. On the other hand, time-distance parameters are static numerical values which are also used to train another MLP.

Table 4.4: Dataset characteristics

<table>
<tr><th rowspan="2">FEATURE VECTOR
(FV)</th><th rowspan="2">DATASET</th><th rowspan="2">#SMP.</th><th colspan="2">#TRAIN</th><th colspan="2">#TEST</th></tr>
<tr><th>H</th><th>S</th><th>H</th><th>S</th></tr>
<tr><td>FV1</td><td>KFlex: Knee flexion/extension</td><td>51</td><td rowspan="6">61</td><td rowspan="6">77</td><td rowspan="6">30</td><td rowspan="6">33</td></tr>
<tr><td>FV2</td><td>KMFlex: Knee flexion/ extension moment</td><td>51</td></tr>
<tr><td>FV3</td><td>KMVal: Knee Valgus Moment</td><td>51</td></tr>
<tr><td>FV4</td><td>KPTot: Total Knee Power</td><td>51</td></tr>
<tr><td>FV5</td><td>Time-dist: Velocity, single support, step length</td><td>3</td></tr>
<tr><td>FV6</td><td>Entire set (all of above)</td><td>207</td></tr>
</table>

Before passing to classification phase data is cleaned by eliminating rows having missing values. Finally, 91 healthy and 110 sick subjects' data is prepared for classification purpose and shared for training and testing purposes as shown in Table 4.4 (H: healthy, S: Sick, SMP: Samples).

4.3.2. Classification by Combining MLPs

The basic classifier structure, used in this study is MLPs combination. Weaknesses of each classifier are diminished by combining classifiers, and more accurate results are expected. In **[48],** two methods are described for combining multiple networks. The first one is the modular approach, in which the task is first decomposed into several subtasks and a specialist network is then trained using the inputs pertaining to the corresponding subtask. The second approach is the ensemble one, in which each network is trained using the same inputs and provides a different solution to the same task. Outputs from these networks are combined to reach an integrated result. Complexity is an important issue to be considered in this case. Differentiation among classifiers may be done by using initial random weights, different topologies, and varying the input data.

As stated previously the final data that is used here has five feature vectors; four for temporal changes of knee joint angle (KFlex, KMFlex, KMVal, KPTot) and one for time-distance parameters. Before training all data sets are scaled to interval [-1, 1]. Totally eight MLPs are trained using MATLAB neural network toolbox. These MLPs are combined in different schemas for experiment 1 and experiment 2 as shown in Figure 4.4. Table 4.5 and Table 4.6 show the topology of each network and their individual success rates on test data. For the first five networks number of hidden nodes and hidden layers are determined experimentally.

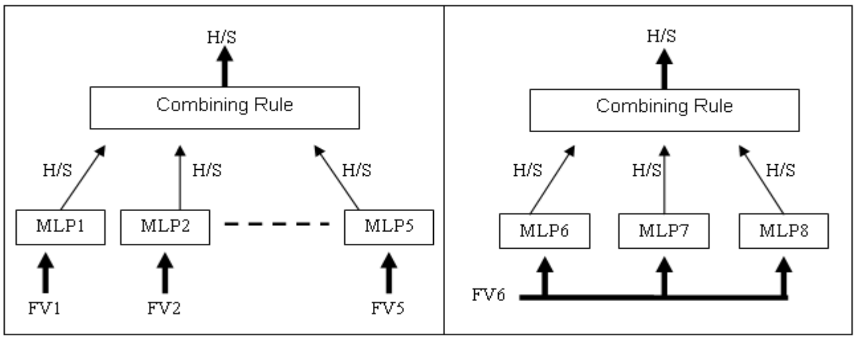


Figure 4.4: MLP combination schemas for experiment 1 (a), and experiment 2 (b) H/S: Healthy or sick, FV: Feature vector

Experiment 1: Input data is decomposed in five sets composed of different feature vectors. Five MLPs are trained by these input sets and then outputs of test set are combined by three different combining rules to reach a final result. So, accuracy of different combining rules is compared.

Table 4.5: Properties of MLPs used in experiment 1

NETWORK	#NODE			#MIS-CLASSIFIED	SUCCESS RATE (%)
	input	*hidden1*	*hidden2*		
MLP1	51	35	10	10	84
MLP2	51	35	10	8	87
MLP3	51	35	10	15	76
MLP4	51	35	10	18	71
MLP5	3	2	-	13	79

Experiment 2: Three different MLPs are trained by using the same composite input set without any decomposition. Here, differentiation of each network is done by different number of hidden layers and hidden nodes.

In both experiments different combination approaches are used, but in both cases combining outputs of classifiers became and important issue. In this study three of these rules, sum, majority vote and max rules, are experimented and results are compared by success rates on test data set.

Table 4.6: Properties of MLPs used in experiment 2

NETWORK	#NODE			#MIS-CLASSIFIED	SUCCESS RATE (%)
	input	*hidden1*	*hidden2*		
MLP6	207	50	-	6	90
MLP7	207	150	40	6	90
MLP8	207	207	50	7	89

After training each network with corresponding input set, test data are presented and the outputs are normalized to use them as posterior probabilities. Since *tansig* function is used as the activation function in all layers of networks, outputs are in interval [-1, 1]. To normalize an output, its absolute value is taken as posterior probability, and its sign is taken as class label (i.e minus sign is for normal and plus sign is for sick subject). Then, its 1-complement is recorded as posterior probability of the other class. Thus, sum and max rules for combining outputs can be applied.

For sum rule, created posterior probabilities are added up for two classes and higher value determined the class label. In max rule, the network, producing the maximum of posterior probabilities determined the class label and the others are ignored. To find the majority vote, each networks' output is converted to class labels by applying a threshold and three agreeing classifiers determine the class label of the test datum. Table 4.7 shows the obtained success rates on test set by applying these combining rules.

Table 4.7: Success rates (number and percentage) for combining rules

Combining rule	Combined networks			
	MLP1-MLP5		MLP6-MLP8	
	#misclassified	*success rate (%)*	*#misclassified*	*success rate (%)*
Sum	4	94	6	90
majority vote	5	92	6	90
Max	5	92	6	90

As seen in Table 4.6 MLPs used in second experiment produced better performances which are expected. However, when the dimension of the dataset increased, and so there are more parameters (like weights) to be tuned, a local extreme of the error function is likely to be found. An ensemble of simple classifiers might be better option for such problems **[42]**. Combining simple classifiers require condition of diversity. Obviously combining identical classifiers does not contribute to accuracy. To check whether our five MLPs have identical classification results or not, we have tested whole dataset by crossvalidation approach and seen that only % 1.5 of the subjects has been misclassified by all classifiers, which proves the disagreement of the classifiers. According to the test set results confusion matrices of the MLPs are created to see where the misclassifications have occurred. Table 4.8 shows these confusion matrices where positive (P) means sick subjects.

Table 4.8: Confusion matrices for used MLPs

		Predicted	
MLP	***Actual***	**Negative (N)**	**Positive (P)**
MLP 1	**N**	21	9
	P	1	32
MLP 2	**N**	28	2
	P	6	27
MLP 3	**N**	26	4
	P	11	22
MLP 4	**N**	29	1
	P	17	16
MLP 5	**N**	26	4
	P	9	24

These matrices proved that while some MLPs are successful at malfunction detection, others are good at separating normal subjects. These properties of simple classifiers support the idea that combining them increases the classification accuracy. To test this idea, these MLPs are combined by different combining rules. In this study three of these rules, sum, majority vote and max rules, are experimented and results are compared by success rates on test data set.

According to these results, it can be concluded that the best individual performance is produced by MLP6 and MLP7 in which entire data set is used for training and testing purpose. However, as the dimension of the data and relatively network size increase, complexity becomes an important drawback. Since it is difficult to process a large set of data training time increases. However, smaller MLPs which use only one feature vector produce less accurate results and combining their outputs increase the accuracy reasonably.

In addition, combining outputs do not increase the accuracy in experiment 2 as much as in the first one. Increasing the number of networks does not cause any improvement after an optimum number, which is "three" in our experiment.

The combining rules show equal performance in experiment 2, but in experiment 1 sum rule is superior to others. Then, as complexities are considered combining many small networks may be preferred when dealing with large dimensional data. The confusion matrices suggest that further study is needed for the effectiveness of the selected features.

4.4. A DECISION TREE-MLP MULTICLASSIFIER FOR GRADING KNEE OA

This part of the study presents the ultimate algorithm behind OAGAIT clinical decision support system for the detecting and grading of a knee OA. The objective of this study is to design a classification algorithm to help physicians by interpreting and further following the progress of OA. The accuracy of the proposed system is expected to be improved compared to our previous work as discussed above by using symptoms and history in addition to numeric gait data, with a multi-classifier approach. In the previous studies, selected knee joint angle features are used to train a single neural network, namely a 3 layer perceptron (MP) and an 89% success rate was achieved for classification of healthy and sick patterns **[38]**. In the next stage of the study, MLP's with identical topology are trained by different feature sets and the outputs are combined by fixed combining rules. This time the success rate of the classifier reached to 94%, for again binary classification **[39],** which suggested that the use of combination classifiers may be used in further work.

Sociodemographic and disease characteristics such as age, body mass index and pain level are also included in decision making. A grade of the OA (0-3, including normal with grade zero) is sought. The grade of the disease is already determined by radiographic methods.

Different types of classifiers are combined to incorporate the different types of data and to make the best advantages of different classifiers for better accuracy. A decision tree is developed with Multilayer Perceptrons (MLP) at the leaves. This gives an opportunity to use neural networks to extract hidden (i.e., implicit) knowledge in gait measurements and use it back into the explicit form of the decision trees for reasoning. The approach is similar to the Mixture of Experts method since different expert MLP's are used for discriminating different grades (our categories) of the disease. Individual feature selection criterion is used with MD measure for feature selection and most discriminatory features are used for each expert MLP.

Automatic feature selection from many numerical gait parameters is another subject that's not studied well before. In this experiment automatic feature selection process produced some valuable results for further analysis of the progress of OA.

The main stages of this experiment are shown by a flowchart in Figure 4.5. The processes followed within these stages are given in detail in following sections.

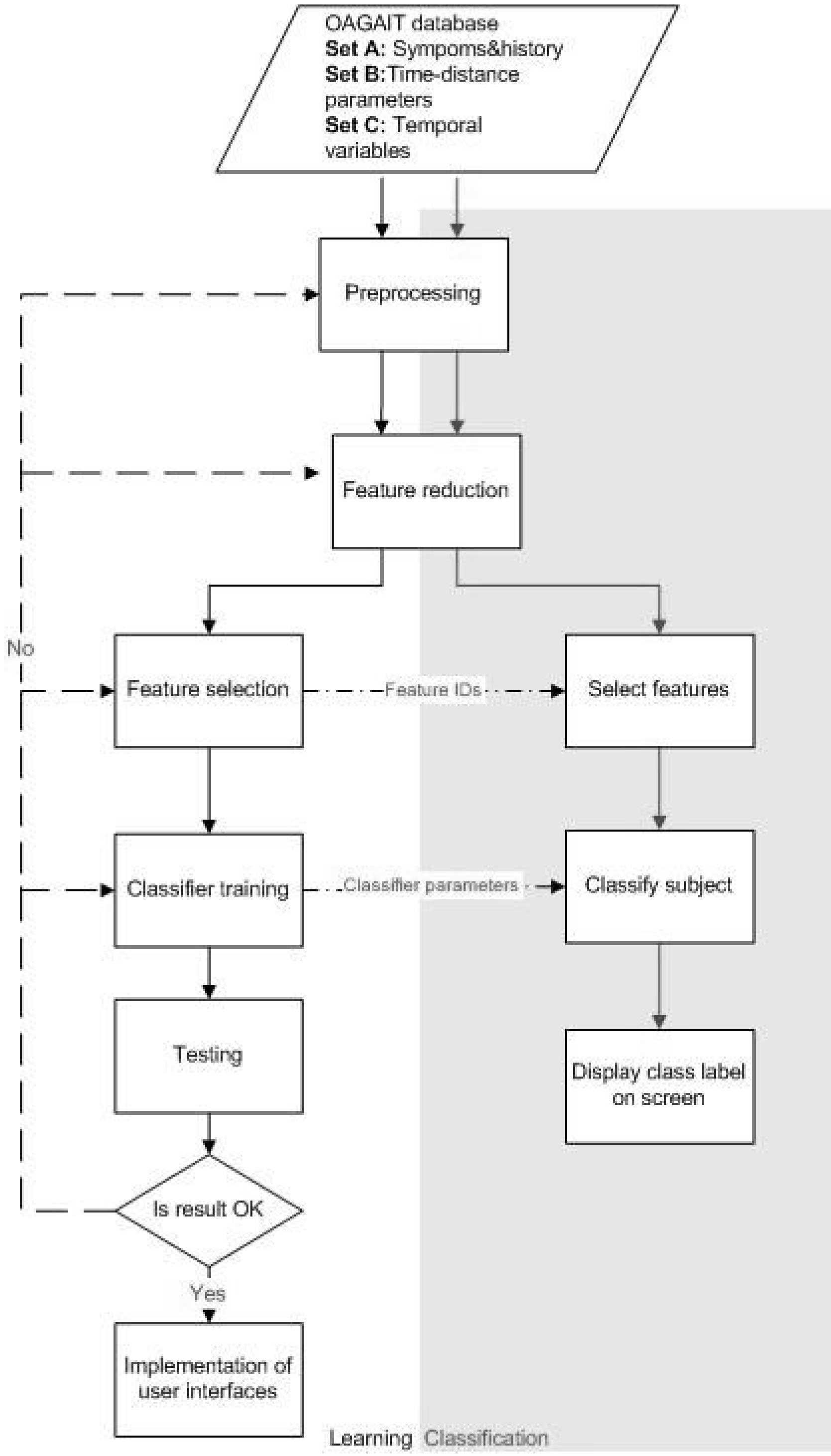


Figure 4.5: Flowchart for classifier design process

4.4.1. Preprocessing of data

As explained in detail in chapter 2 of the report, the gait data is mainly composed of three different sets of data. The first subset (Set A) of the data is about symptoms and history of the subjects and defined as:

A = {age, BMI, pain, stiffness, period, history, sex}

In preprocessing steps, some of these variables are converted to numerical values before they are used as features. For example, while weight and height attributes are not used for classification purpose, body mass index (BMI), which is equal to weight divided by the square of the height, is created as a new feature. Age of the subject is calculated from date of birth, disease periods are converted to months as unit. Pain and morning stiffness are numeric values between 0 and 10, family history is a binary value indicating whether the same disease exist in family history or not. Sex is another binary valued feature where 0 stands for women and 1 for men. The distribution of the subjects for these dataset are shown in Table 2.1 by summarizing min, max and average values of the features.

Table 4.9: Limits of the personal features

Features	**Normal subjects**			**Patients**		
	min	max	average	min	max	average
Age	19	63	43	41	80	60
BMI	18	46	27	20	49	32
Pain	0	0	0	1	10	6,6
Stiffness	0	0	0	1	10	5,2
period (year)	0	0	0	0	30	6,6

The second set (Set B) of the data is composed of *time distance parameters* which are gathered in one cycle of gait.

B = {Cadence, Walking Speed, Stride Time, Step Time, Single Support, Double Support, Stride Length, Step Length}

The final subset of the data can be defined as the temporal changes of the joint angles from four anatomical level and three motion planes (Set C) as below.

C = {PelvicTilt, Pelvic Obliquity Knee Flexion, Knee Varus, … }

These sets of data are grouped according to usage purpose of the data. First two sets are combined for decision tree training and the third one is for MLPs training. Before going to further steps the data sets are cleared by deleting the samples having missing entries from the database. If possible, some missing values are completed by using previous information of the subject.

4.4.2. Feature Reduction and Selection

The first and second sets of data (A and B) is combined and used for constructing decision tree. Since the nature of the decision tree algorithms is based on the selecting best feature and best split point in a top down fashion, an additional feature selection is not applied to this set.

However at the leaves of the tree attributes from set C is used. Combining all these attributes into a complete dataset, 33 (attributes) x 51(time samples) dimensional arrays for each subject are obtained. Figure 4.6 shows the flowchart for feature reduction and selection processes used in this experiment. The changes in dimension of the datasets are shown on the lines by “dim” term to clarify the process.

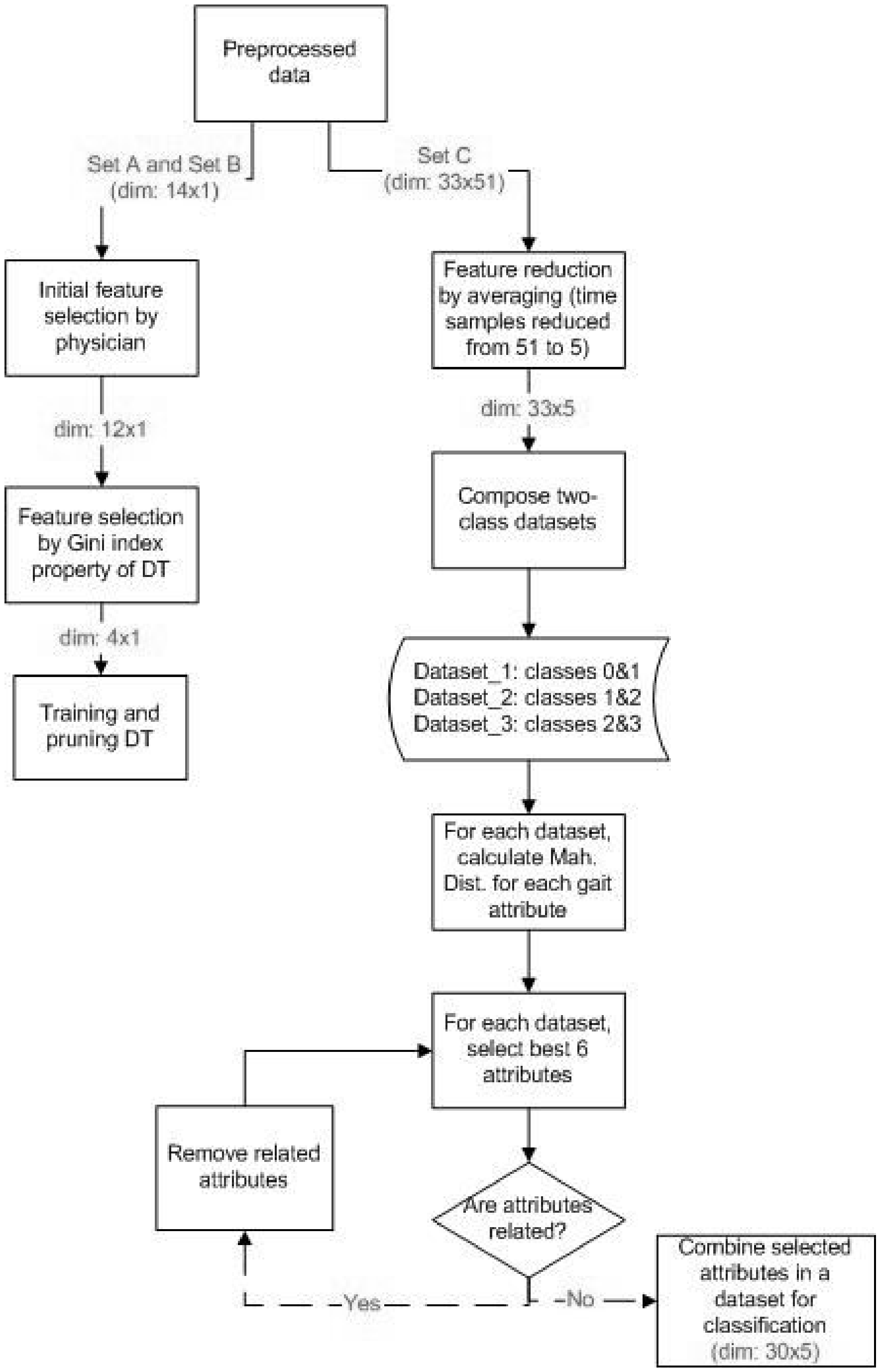


Figure 4.6: Flowchart for feature reduction and selection processes

A basic reduction technique is applied before selection as in previous trials. The dimensions of all attributes are reduced from 51 to 5, by taking means of 10 consecutive time samples. The MD is used as a selection criterion.

Binary class datasets are created for selection of most discriminative gait attributes. The reason for binary case is due to the expert MLP's able to discriminate successive categories, as will be explained later. Since the number of samples in each case is about 130, the dimensions of the datasets are fixed to 30 by including 6 of attributes. Table 4.10 displays the classes that datasets include and the levels of selected gait attributes for those datasets.

Table 4.10: Levels of the selected gait attributes for each binary class case (P: Pelvic, F: Foot, H: Hip, K: Knee)

Classes	**Attribute number**					
	1	*2*	*3*	*4*	*5*	*6*
0-1	F	H	K	K	P	F
1-2	**K**	**K**	**K**	H	**K**	**K**
2-3	**H**	**H**	F	**H**	**H**	K
1-3	**K**	**K**	**K**	P	**K**	**K**
0-2	F	K	H	F	K	P

In the next stage, these gait attributes are used for creating input vector of the related expert MLP. Multidimensional input vectors are created by combining six best features to train MLPs. The selected feature list is revised by the expert physician and the similar features are removed to prevent including highly correlated features in the datasets. Then 6 of 33 features are selected for composing datasets for each expert MLP. Stages of selection process are summarized in Table 4.11.

Table 4.11: Selection of gait attributes using MD values

```
Calculate Mahalanobis distance matrix between classes in
dataset: the distance matrix between the class means,
sphered using the average covariance matrix of the per-class
centered data.
```

```
For i=1 to #gait_attributes

  • Select ith gait attribute as feature set
  • Create a two class dataset
  • Calculate covariance matrix P
  • Using the below formula find the Mahalanobis distances
    of class means (x, y)
```

$$d(\vec{x},\vec{y}) = \sqrt{(\vec{x}-\vec{y})^T \mathrm{P}^{-1} (\vec{x}-\vec{y})}.$$

```
  • Write distance value to array Dm

End

  • Sort Dm
  • According to number of time samples of gait attribute
    select first 1,2,5 or 10 best attributes
  • Create a new dataset with selected gait attributes
```

Automatic selection of these attributes gives some valuable information about progress of the OA. For example while knee related attributes seen more frequently in dataset composed of classes 1 and 2, hip related ones seem more discriminative for classes 2 and 3. This shows that as the grade of the illness increase hip angles are affected more. This kind of information is valuable for clinical decision making and training physicians.

4.4.3. Combining classifiers

It is difficult to combine these features due to their diversified units, e.g., continuous variables, binary values, and discrete labels. Therefore, the combination of multiple classifiers is a good solution for a problem involving a variety of features. It is also important for an M.D. to understand how and/or why the classifier makes its decisions rather than black-box solutions. Hence, a classifier that's able to do that should be aimed.

In this study a Mixture of Decision tree classifiers and Multilayer Perceptrons that are experts for different regions of the feature space are used for classifying four levels (0 to3) of the knee OA. In that aspect, the algorithm is similar to the 'mixture of experts (ME)' approach in the literature **[45].** ME algorithm is based on the principle of' 'divide and conquer' in which a large, hard to solve problem is divided into many smaller, easier to solve ones **[42].** ME is a tree-structured architecture for supervised learning and further for classification with the participation of the experts in the final decision making. The ME architecture has been proposed for neural networks **[42]**. The experts are neural networks, which are responsible for a part of the feature space. The selector uses the output of another neural network, namely 'the gating network'. If the input to the gating network is called **X**, then the output can be defined as a set of coefficients $\mathbf{p_1(x)}$**, ….,**$\mathbf{p_L(x)}$ where $\mathbf{p_i(x)}$ is interpreted as the probability that expert $\mathbf{D_i}$ is the most competent expert to label the particular input **x [17].**

In this study it is aimed to train a decision tree with another subset **y** of the feature space instead of training a gating network with the same input **x.** The output of the decision tree is again a set of probabilities which are served to expert neural network at that leaf as prior probabilities of the classes. The neural networks at each leaf are the experts for classifying the subjects of type falling in that leaf. Different experts are created by running the feature selection algorithm at each leaf and designing different structured neural networks according to these selected features. So one of the networks is responsible for categorization of, for example, first and second classes, the other is responsible for third and fourth classes.

Decision tree classifiers are widely used for building classifier ensembles **[42, 56, 58]**. Binary features and features with small number of discrete values are especially useful for the purpose since the decision can be easily branched out. Since distance is not easy to formulate when the objects are described by categorical or mixed-type features, the decision trees are regarded as nonnumeric methods for classification **[42, 59]**. Using decision trees for clinical medical decision making problems [59**]** is a popular approach since they are cost effective, easy to implement and they have descriptive quality, which makes them advantageous over black-box approaches such as ANN. It is difficult to incorporate a neural network model into a computer system in an explicit form since inner formulation and calculations are not shown to the user. In contrast, once a decision tree model has been built, it can be converted to *if…then…else* statements that can be implemented easily in most computer languages without requiring an additional effort **[59, 60]**.

The four grades of OA are defined by Kellgren grade which is based on the radiographic assessment of the joint space narrowing. The accuracy of the proposed system will be safeguarded by using all feature sets A, B, C above with Kellgren grade-labeled subjects.

The learning and classification processes consist of two stages. In the first stage a decision tree is trained by using data set A and B. In the second stage, the samples falling at each leaf is analyzed for feature selection and an expert MLP is trained by composed datasets using attributes from set C to classify the data into one of the two categories 0-1, 1-2 etc. Figure 4.7 shows the steps of combining and training of classifiers with a flowchart.

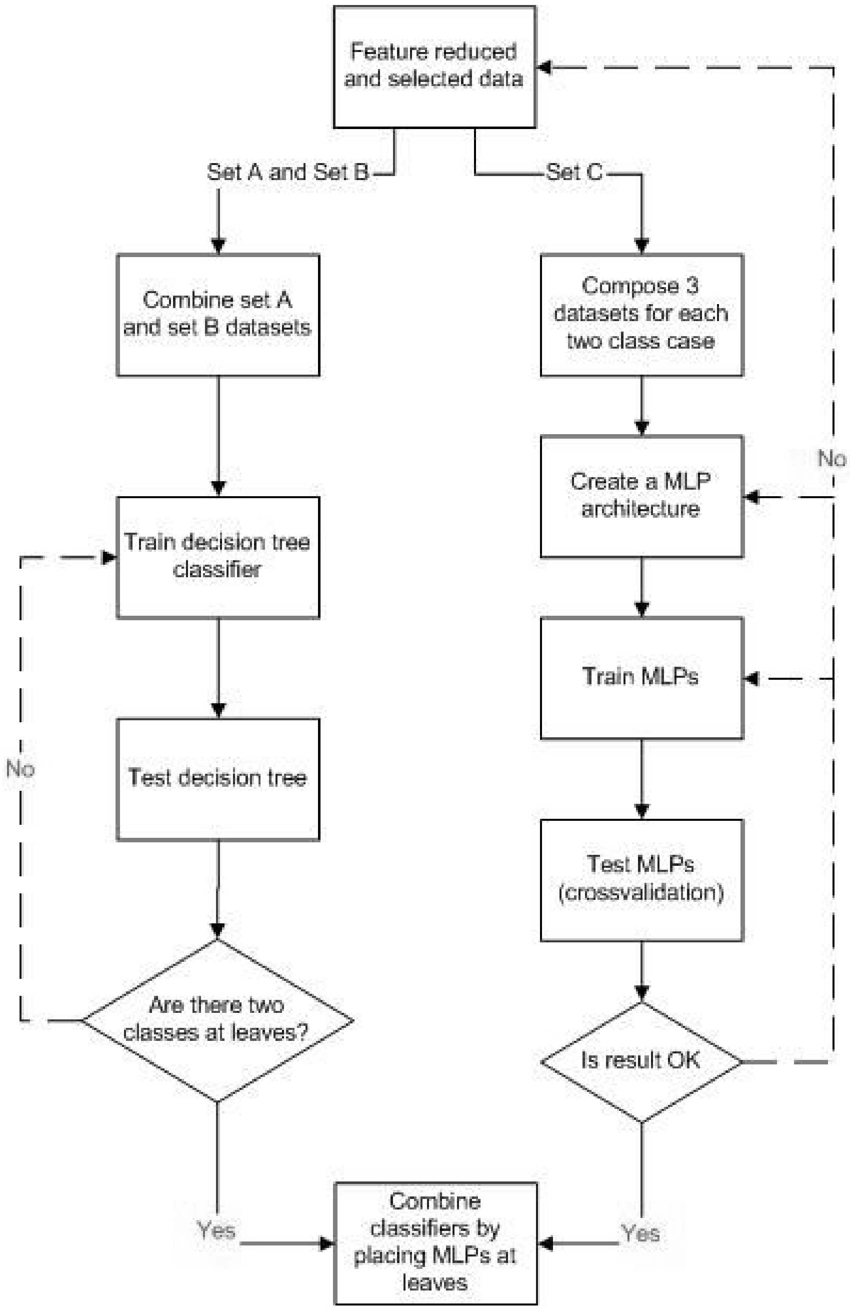


Figure 4.7: Flowchart for classifier combining

4.4.4. Training and testing

The training of our combined algorithm has two divisions, decision tree training and training of MLPs, as shown in Figure 4.7. Decision tree training is different than training of many other classifiers for which a topology is created first and then training data is presented. Decision tree construction and training processes can not be considered separately. Tree is constructed according to the training data set by using some predefined criteria. These criteria like *splitting* and *stopping* criteria are important for constructing best tree representing the training set and avoiding overfitting.

The basic idea of tree learning is to choose a split among all the possible splits at each node so that the resulting child nodes are pure enough. In our algorithm, only univariate splits are considered. That is, each split depends on the value of only one feature variable. All possible splits consist of possible splits of each feature. If X is a *nominal categorical* feature of I categories, there are $\mathbf{2^{I-1} - 1}$ possible splits for it. If X is an *ordinal categorical* or *continuous* feature with K different values, there are $\boldsymbol{K - 1}$ different splits on X. A classification tree is grown starting from the root node by repeatedly using the following steps on each node.

1. *Find each feature's best split:* For each feature, sort its values from the smallest to the largest. For the sorted feature, go through each value from top to examine each candidate split point (call it v, if x <= v, the case goes to the left child node, otherwise, goes to the right.) to determine the best. The best split point is the one that maximize the splitting criterion when the node is split according to it.

2. *Find the node's best split:* Among the best splits found in step 1, choose the one that maximizes the splitting criterion.

3. *Split the node:* Split the node by using its best split found in step 2 if the stopping rules are not satisfied.

At node, the best split is chosen to maximize a splitting criterion. This splitting criterion may be Gini impurity, misclassification impurity etc. For classification trees the impurity is defined with the Gini index of diversity **[42]**.

Stopping and pruning criteria are also important for tree construction. Stopping rules control if the tree growing process should be stopped or not. In this study the number of samples in a node is restricted to be at least 10 so the node is not split any more.

Pruning step of the tree construction is done by considering structure of our combination. Means, we applied a method to prune the tree that each leaf has samples from two classes. In next stage of the combination, MLPs responsible for discriminating these two classes are replaced to the related leaf. This approach changed the dimension of the problem from a recognizer for four-categories to three recognizers with 2 categories, using expert neural networks for discriminating neighbor classes.

The basic classifiers structure used in the leaves of the decision tree are MLPs. Three- layered (one hidden layer) MLPs are trained by different input vectors. These input vectors are trained by automatically selected gait attributes, different for each leaf as mentioned above. So, they are assumed to be experts in the region of the binary decision of the category.

The trained MLPs are placed at the leaves of the trained decision tree by class matching. Figure 4.8 summarizes the proposed combination in a simplified form. Where;

- $\mathbf{Y = \{y_1, y_2, ... y_n\}}$ is the union of set A and B above
- $\mathbf{T = \{t_1, t_2, ... t_n\}}$ is the set of corresponding threshold values for above, used for composing tree.
- $\mathbf{X = \{x_1, x_2, \dots x_m\}}$ is the set of datasets composed by selected attributes of set C and presented to the expert networks

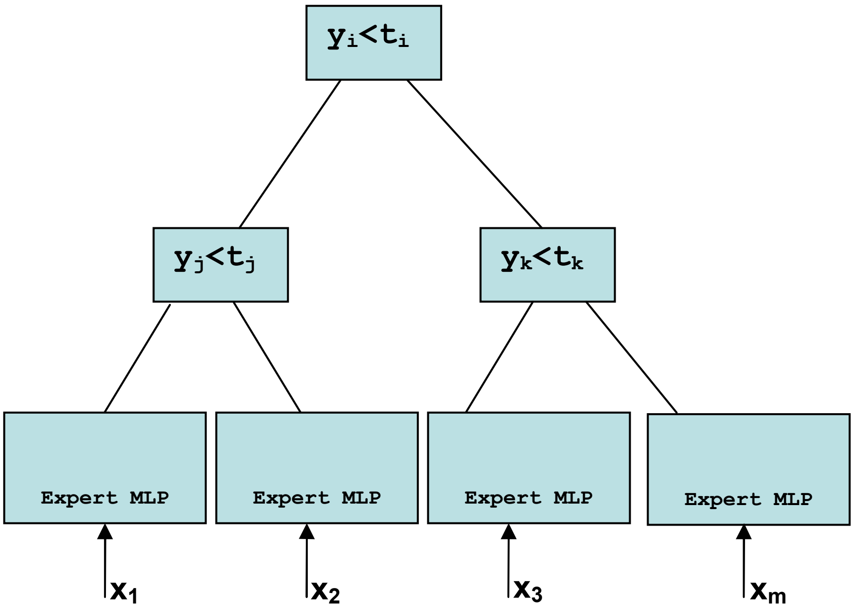


Figure 4.8: Proposed example combination in a simplified symbolic representation.

Implementation of the algorithm and analysis of the results are explained in detail in next chapter of the report.

Testing of the MLPs is done by crossvalidation method. By this method, the set is randomly permutated and divided in N (almost) equally sized parts. The classifier is trained on N-1 parts and the remaining part is used for testing. This is rotated over all parts. Average error rate and standard deviation is returned as the result. In this experiment tenfold crossvaldiation is used a short pseudo code is given in Table 4.12.

Table 4.12: Tenfold Crossvalidaton testing

```
For i=1 to iteration count

   •  Separate 1/10 of the samples for testing randomly
   •  Train the related MLP with rest of the samples
   •  Test the trained MLP with separated test samples
   •  Record the error to an array call Err
End

   •  Produce the mean of values of Err and standard
      deviations as the result of the testing process
```

CHAPTER 5

IMPLEMENTATION AND RESULTS

5.1. INTRODUCTION

This chapter is about the implementation of the classification algorithms and results. As explained in previous chapter the classification algorithm is created by combining a decision tree with a number of MLPs. First a decision tree is created by using symptoms and history of the patients and time-distance parameters together. Then MLPs are trained with different feature sets of gait data to make them experts for discriminating different grades of the illness. Finally, trained MLPs are replaced at the leaves of the tree and the algorithm is tested by unseen data.

5.2. CREATING THE DECISION TREE

A data set having about 40 subjects from each category (0, 1, 2, 3) is composed for training the decision tree. A tree is fitted to the composed dataset by using ***treefit*** property of the MATLAB Statistical Toolbox. Decision tree is constructed in top-down fashion with binary splits, where each node checks a numerical value of a single feature. A termination criterion at a node could be that all objects be labeled as belonging to the same category. Unfortunately, this is an ideal situation where the number of samples and features should represent the problem perfectly, which is not valid for the data at hand. Here it is aimed to continue until samples from 2 categories are left. The steps of our tree construction algorithm are shown in Table 5.1:

Table 5.1: Tree construction algorithm

```
1. Assign all objects to root node.

2. Split each feature at all its possible split points.

3. For each split point, split the parent node into two
   child nodes by separating the samples with values
   lower and higher than the split point for the
   considered feature.

4. Select the feature and split point with the highest
   reduction of impurity.

5. Perform the split of the parent node into the two
   child nodes according to the selected split point.

6. Repeat steps 2-5, using each node as a new parent
   node, until the tree has maximum size.

7. Prune the tree back using cross-validation to select
   the optimal sized tree.
```

Since we are constructing a classification tree, Gini index impurity is used at step 4 of the algorithm. Consider a c-class problem with $\mathbf{\Omega} = \{w_1, w_2, ..., w_c\}$. Let $\mathbf{P_j}$ be the probability for class $\mathbf{w_j}$ at a node $\mathbf{t}$. Gini impurity is defined as

$$i(t) = 1 - \sum_{j=1}^{c} P_j^{\,2} \qquad \textbf{(Equation 5.1)}$$

For the most pure case $i(t) = 0$. The highest impurity in the case of uniform distribution is $i(t) = (c-1)/c$.

A set of possible stopping criteria are explained in Section 3.2 of the report. In this implementation the method of "setting a threshold value for the number of samples at a node" is used. The number of samples in each node is limited to 10. An example tree before pruning is shown in Figure 5.1.

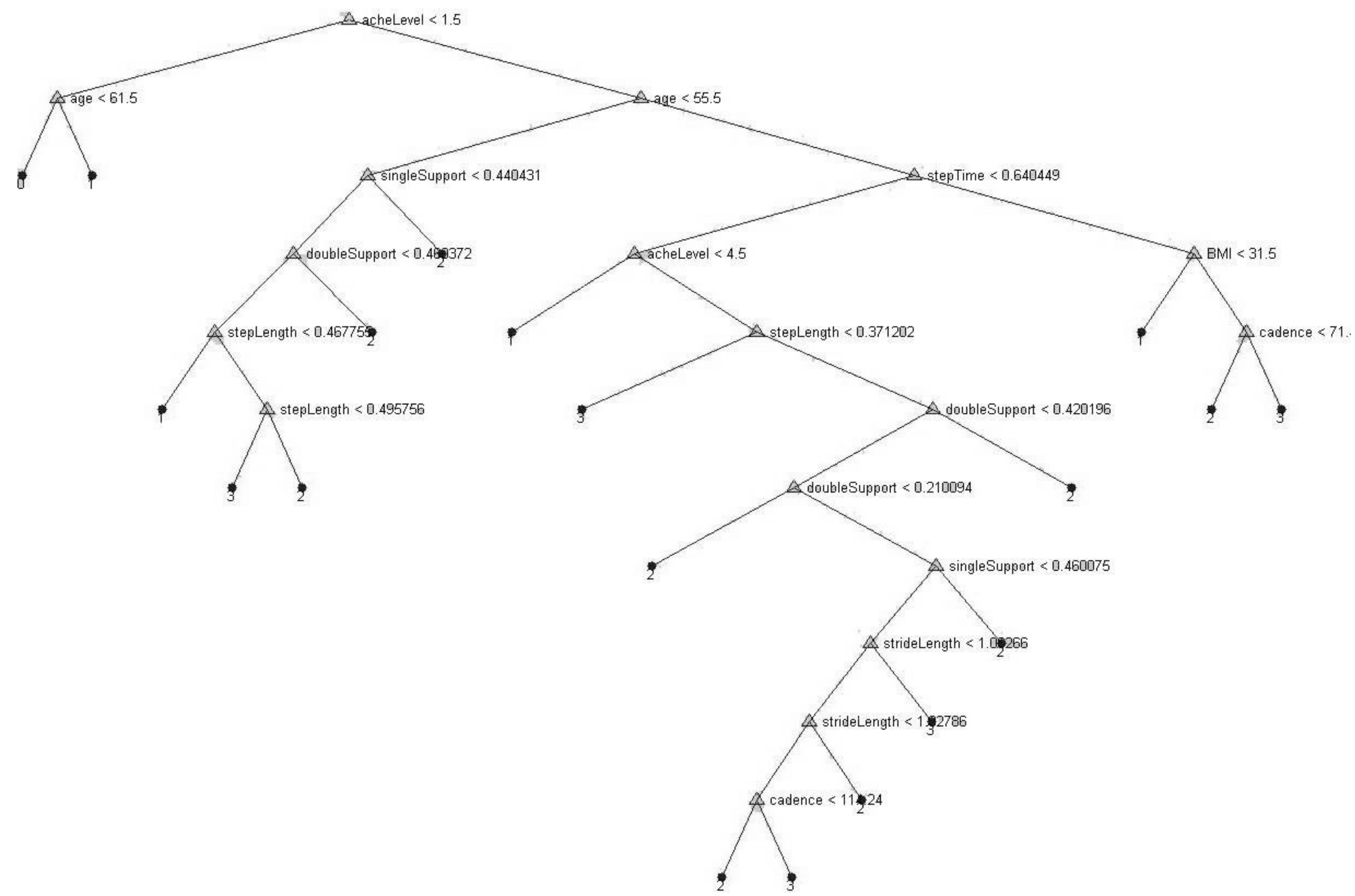


Figure 5.1: An example to created decision tree (no pruning)

Pruning is also applied when composing the decision tree. As explained in section 3.2, pruning is a significant step for decision tree construction to avoid overfitting of the tree to training data. The tree is pruned based on an optimal pruning scheme that first prunes branches giving less improvement in error cost. To determine the best size of the tree, it is tested by crossvalidation approach as shown in Table 5.2:

Table 5.2: Pruning algorithm

```
1. Partition training data in "training" and "validation"
   sets.
2. Build a complete tree from the "training" data.
3. Until accuracy on validation set decreases do:
     a. For each non-leaf node, N, in the tree do:
     b. Temporarily prune the subtree below N and
        replace it with a leaf labeled with the current
        majority class at that node.
     c. Measure and record the accuracy of the pruned
        tree on the validation set.
4. Permanently prune the node that results in the
   greatest increase in accuracy on the validation set.
```

This algorithm pools the information from all subsamples to compute the cost for the whole samples. Applying this method to our tree shown in Figure 5.1 we got a graph showing the cost versus number of final nodes. The cost value in this graph represents the misclassification rate for classification trees. The pruning of the tree is done considering the optimum number of terminal nodes shown in Figure 5.2.

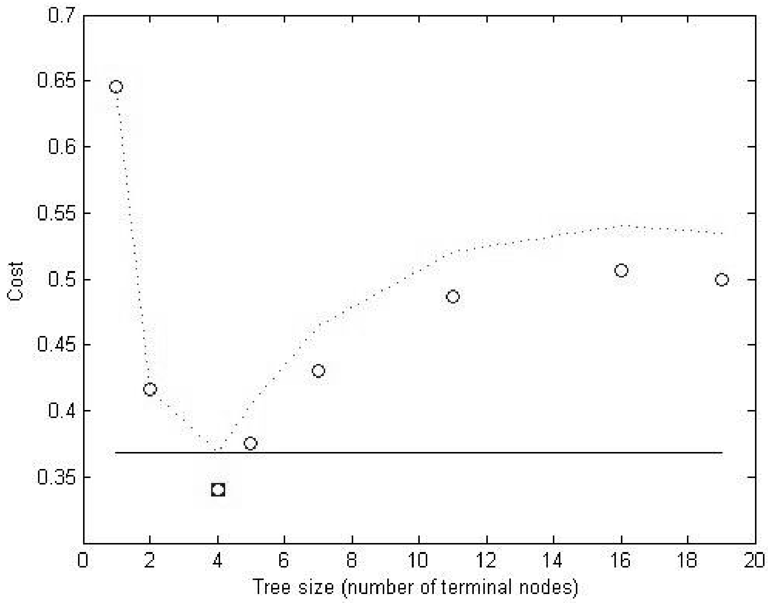


Figure 5.2: Cost (classification error) versus tree size

Cost graph hints that the number of terminal nodes (leaves) should be chosen around 2-6 for minimizing the classification error. The tree shown in Figure 5.1 is pruned 3 times by cost reduction algorithm, and once manually. The aim of manual pruning is to leave samples from two categories at each leaf. Unfortunately the desired situation of leaving only two categories at each leaf is not satisfied fully. Those small numbers of samples at a leaf after pruning will be ignored by the MLPs.

5.3. FEATURE REDUCTION AND SELECTION PROCESSES

As previously explained gait attributes are represented by 51 time samples. Consecutive time samples are averaged before training, so each is represented by five time samples as shown in Figure 5.3. In a previous trial two different feature reduction techniques are compared first by the MD criterion and then by performances of the classifiers (section 5.1). It is observed that datasets created by FFT techniques produced worse results than averaged datasets in MD calculations and also in the classification process. This study showed that number of time samples and number of features should be optimized for better accuracies. In Figure 5.4 graphs derived from this study are shown; dimension of the datasets are represented by *(#of_time_samples) x (#of_gait_attributes)*. Also, the optimum ratio of the number of samples and feature size was tried to reach by limiting the number of selected gait attributes to six.

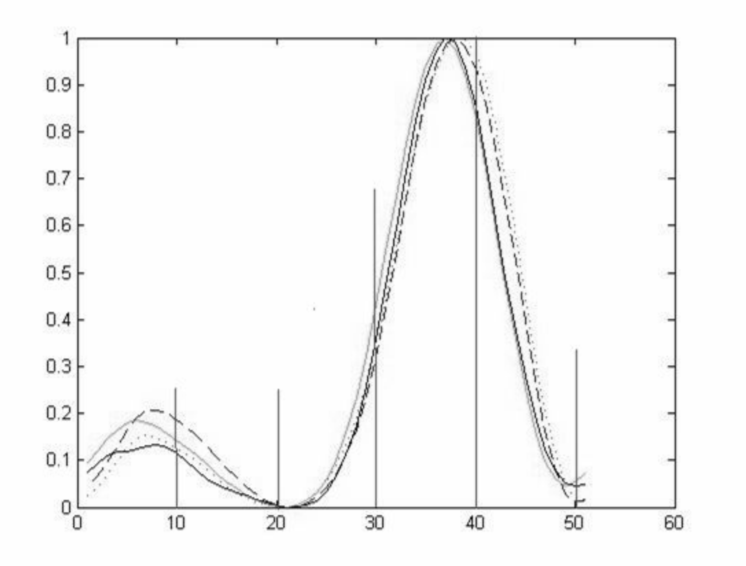
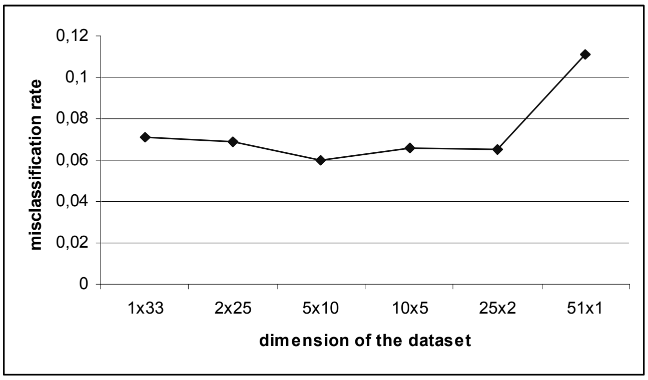


Figure 5.3: Averaging consecutive time samples for feature reduction

MD based feature selection was implemented after reduction in time samples. The automated feature selection method is compared with the manual feature selection done by expert physician in a previous trial (section 5.3). A high match is recognized between the automated selected features on the basis of the MD criterion and the ones suggested by the gait analysis expert. This result encouraged us for using MD criterion in further stages of the study.



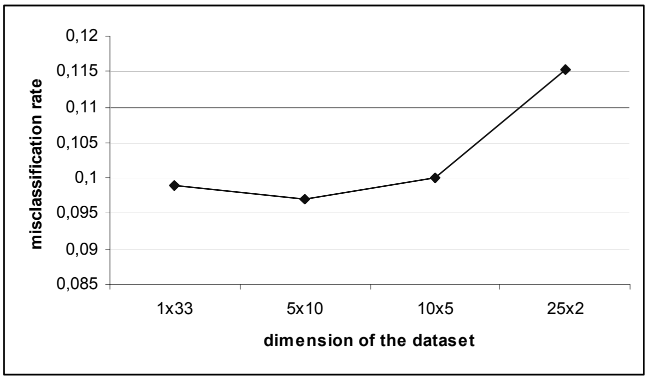


Figure 5.4: Dataset dimension versus misclassification rate (a. averaged datasets, b. FFT applied datasets)

As a result of selection process the features shown in Table 5.3 are obtained. Different attributes are effective at different stages of OA, as seen in the table. The parts of the body that the gait attributes are related are important for deriving information about progression of the knee OA.

Grades 0-1: Most discriminative gait attributes for classification of these grades are from different parts of the body. This can be commented as in early grades of the disease a severe deformation in knee joint does not exist.

Grades 1-2: Five of the selected gait attributes for discrimination of these classes are related to knee joint. This proves that if the grade of the disease progress from 1 to 2, the knee joint is affected seriously. One hip related feature may be the indicator of the early deformation of this joint.

Grades 2-3: The results of the selection process for these grades are really remarkable. Since the number of hip related gait attributes is higher than the knee related ones for these grade levels, it can be concluded that at advanced levels of the disease different parts of the body other than knee are also affected. Medically, it would be incorrect to say that the same disease is also seen in other joints of the body in high grades but it started to affect the other joints. It is known that most patients living with the same disease along time may change their postures to compensate some bad effects of the disease like pain **[20, 26, 29, 30].** That may be why patients with high grade of OA have abnormal hip joint patterns.

Grades 0-2 and grades 1-3: Selected attributes for these grades are not used for classification purpose but added here to comment about the progress of the disease. It is seen that most of the attributes for grades 0-2 are common with the ones for grades 0-1, which shows equal deformation in different parts of the body. Similarly the attribute for grades 1-3 are common with the ones for grades 1-3 which shows deformation of the knee joint more than the others.

Table 5.3: Selected gait attributes for each two-class case (P: Pelvic, F: Foot, H: Hip, K: Knee, Flex: Flexion, M: Moment, Tot: Total, Dor: Dosrflexion, Rot: Rotation, Val: Valgus, Obliq: Obliquity, Adb: Abduction)

Classes	Selected Gait Attributes					
0-1	**F**.M.Dor.	**H.**Flex	**K**.M.Flex	**K**.Flex	**P**.Tilt	**F**.Rot
1-2	**K**.Flex	**K**.M.Flex	**K**.P.Flex	**H**.P.Tot	**K**.P.Tot	**K**.Val
2-3	**H**.P.Abd	**H**.Flex	**A**.P.Dor	**H**.Rot	**H.**P.Flex	**K.**Val
1-3	K.P.Flex	K.P.Tot	K.Flex	P.Obliq	K.M.Rot	K.Val
0-2	F.M.Dor	K.Flex	H.Flex	F.Dor	K.Rot	P.Tilt

5.4. IMPLEMENTATION OF MLPS

A set of classifiers are compared by same dimensional datasets. It is concluded that nonlinear classifiers performed quite well and better than the linear ones. Backpropagation Neural Network (bpxnc) and Radial Basis Support Vector Machine (rbsvc) classifiers produced best generalization accuracy by almost all datasets. Comparing learning rate of the classifiers it was concluded that more data is needed to increase the performances of the linear classifiers. These results formed the direction of the study towards using MLPs in further stages.

In second trial of the study combining classifiers approaches are investigated. Different combination schemas for combining a group of MLPs are experimented. MLPs are used to classify the subjects as healthy or sick, using temporal changes of knee joint angle and time-distance parameters as features. Two different combination methods are tried. In the first experiment five MLPs are trained by different subsets of the feature space, and in second one three MLPs are trained by entire feature set. Then the outputs are combined by sum, majority vote and max rules to produce final class labels. These two experiments show that using entire data set produces more accurate results than using decomposed data sets, but complexity becomes an important drawback. However, when a proper combining rule is applied to decomposed sets, results are more accurate than entire set. So, for design of the final classification algorithm expert MLPs are trained by different subsets of the feature set.

Almost 60 subjects from each category are used for composing datasets for MLP training. Three datasets are created by the selected gait attributes for training three MLPs. These MLPs are responsible for discrimination of classes 0-1, 1-2 and 2-3 respectively. These MLPs are trained by *bpxnc* property of the PRTools **[27]**. This function creates a feedforward neural network and uses Backpropagation algorithm for training. MLPs have three-layer structures with binary outputs. They are tested by crossvalidation approach and an average error rate for each is gathered as shown in Table 5.4.

Table 5.4: Classification errors of MLPs

MLP	Classes (grades)	Classification error
MLP1	0-1	% 11
MLP2	1-2	% 16
MLP3	2-3	% 21

These MLPs are also tested by receiver operating characteristic (ROC) curve, which are shown in Figure 4 where x and y axis represent the error of the first and second classes respectively. ROC curve is a graphical plot of the sensitivity vs. (1 - specificity) for a binary classifier system as its discrimination threshold is varied, where;

$$sensitivity = \frac{number\_of\_true\_positives}{number\_of\_true\_positives + number\_of\_false\_negatives}$$

$$specificity = \frac{number\_of\_true\_negatives}{number\_of\_true\_negatives + number\_of\_false\_positives}$$

The ROC can also be represented by plotting the fraction of false negatives vs. the fraction of false positives as in Figure 5.5. For MLPs setting a threshold values for posterior probability values determine a point on the ROC curve. Plotting these points for each possible threshold value creates a curve.

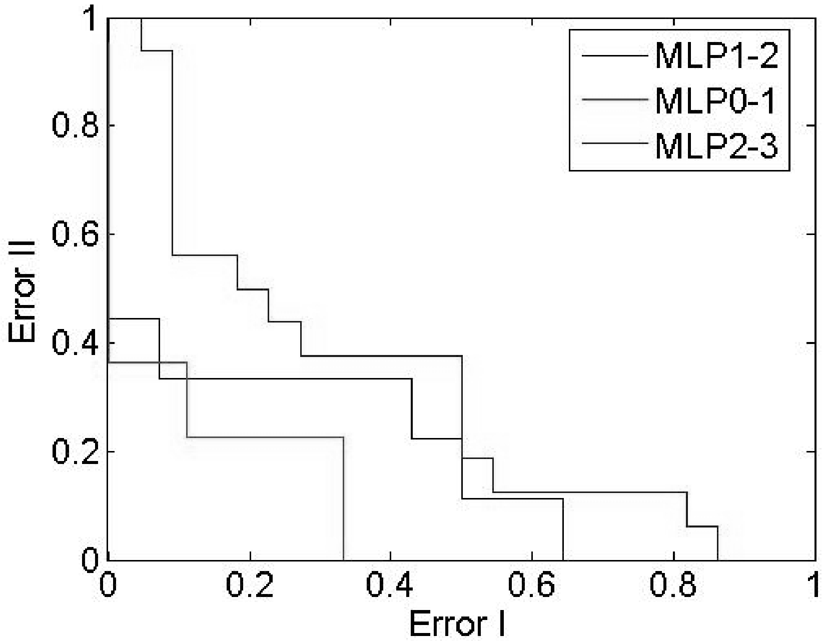


Figure 5.5: ROC curves for MLPs

Error rates and ROC curves of MLPs show that discrimination of some classes are difficult than the others. In ROC curves, the smaller area under curve (AUC) shows better classifier. In graph we see that the AUC of MLP discriminating classes 0-1 is the smallest and the one discriminating classes 2-3 is the greatest. Then it can be commented that, as the grade of the illness increase the discrimination power of the gait patterns decrease. Therefore, for discrimination of these high grades of the disease some more details about these grades may be added to the classification algorithm. For instance, ignored gait attributes may also be contributed to the classification process if the number of samples increases.

5.5. RESULTS

In the second stage of the classification process, expert MLPs are placed at the leaves of the tree. Figure 5.6 shows an example of the composed tree pruned to level 3 end corresponding MLPs at leaves of it.

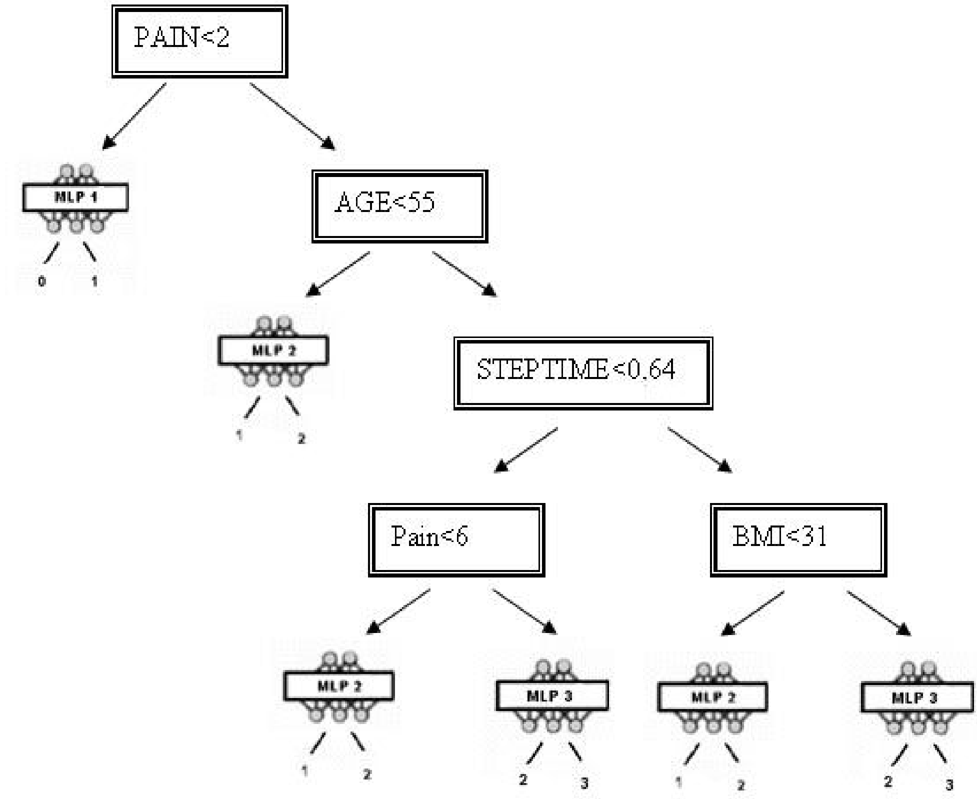


Figure 5.6: Composed decision tree by prune level 3

Table 5.5 shows the number of samples at each leaf of the tree. The signed ones represent the samples at unexpected leaf; means the MLP at that leaf is not responsible for detection of that class. For example 2 subjects from class 3 placed at leaf2, but MLP2 placed at that leaf is responsible for detection of classes 1-2. Summing up the samples at unexpected leaves we got a %95 training error for construction of decision tree.

Table 5.5: Distribution of samples in decision tree

Leaves:	L1	L2	L3	L4	L5	L6	Total Error
Used MLP:	MLP1 (0-1)	MLP2 (1-2)	MLP2 (1-2)	MLP3 (2-3)	MLP2 (1-2)	MLP3 (2-3)	
Class-0	40	0	0	0	0	0	0
Class-1	7	11	12	<u>2</u>	7	<u>1</u>	3
Class-2	0	14	8	11	2	5	0
Class-3	0	<u>2</u>	<u>2</u>	13	<u>1</u>	22	5
Total	47	27	22	26	10	28	8

To calculate the overall error rate of the combination, MLP errors and DT error are used in a formulation. Following code segments summarizes the steps of calculating error rates for detection of each level of the OA.

```
For level = 0 to 3
      For leaf = 1 to #leaves
```

$$Error[level] = \frac{\sum samples\_from\_level\_in\_leaf}{\sum samples\_from\_level} \times error\_rate\_of\_MLP\_in\_leaf$$

```
      End
End
```

The error rate of the decision tree and MLPs are combined in a different fashion by this formulation. The tree is constructed by equal number of samples (which is

40) from each level of the illness to objectively calculate the error rates. According to above formulation, if a sample is sent to wrong leaf of the tree, means there is no MLP to classify it, then the error rate is taken as 1 in that leaf. Confusion matrices created by training and test data as shown in Table 5.6 and Table 5.7 give more detailed information about success rate of the combination.

Table 5.6: Confusion matrix for combination (training data)

		Estimated Classes					
		0	**1**	**2**	**3**	**total**	**error rate**
Actual classes	**0**	39	1	0	0	40	0,025
	1	1	34	2	3	40	0,15
	2	0	1	37	2	40	0,075
	3	0	5	2	33	40	0,175
	total	40	41	41	38	160	**0,10625**

For creation of Table 5.6 data used for training of the combination is presented to it and the numbers of misclassified samples are detected and almost %90 classification rate is achieved. Since the MLPs were trained with the same data the contribution of MLP errors on this total error is very small. The reason for most of these misclassified subjects is that they have been assigned to wrong expert MLP in decision tree. This wrong assignment is most probable because of pain level which is a subjective feature determined by subject himself. For example, a subject from grade 1 can determine his pain level as 10 (max value for pain level) while the other from grade 3 can say 3. These kind of subjective features are not preferred in classification processes but experts and medical studies in literature show that pain level is one of the important indicators of the selected illness, so added to datasets here.

Table 5.7: Confusion matrix for combination (test data)

		Estimated Classes					
		0	**1**	**2**	**3**	**total**	**error rate**
Actual Classes	**0**	18	2	0	0	20	0,1
	1	1	15	2	2	20	0,25
	2	0	1	17	2	20	0,15
	3	0	2	4	14	20	0,3
	total	19	20	23	18	80	**0,2**

Then the algorithm is tested by an unseen dataset composed of 20 samples and classification rate of %80 is achieved. Since the system is tested by unseen data the generalization rate reduced from %90 to %80. Because this time the errors from MLPs are also effective in error calculation.

Finally, for comparing the success of our combination schema with a single classifier a four-class neural network is trained. A three layered MLP, call it MLP4, is created having 50 units in input, 10 units in hidden layer and 2 outputs to discriminate four classes. The same feature reduction and selection processes are applied to this new dataset, but the number of inputs is increased to 50. Crossvalidation testing is applied during training. The estimated labels are gathered as an output of crossvalidation algorithm. Table 5.8 shows the confusion matrix for MLP4.

Table 5.8: Confusion matrix for MLP4

		Estimated classes					
		0	**1**	**2**	**3**	**total**	**error rate**
Actual classes	**0**	27	7	5	1	40	0,325
	1	3	24	7	6	40	0,4
	2	3	8	21	8	40	0,475
	3	3	5	12	20	40	0,5
	Total	36	44	45	35	160	**0,425**

The classification rate of the MLP4 which is about %58, proved us that using different expert for different part of the feature space and then combining them produced reasonable better result than using a single multi-class classifier.

5.6. CLASSIFICATION EXAMPLES

In medical applications, physicians want to see some reasons for class assignments rather than just learning the results that software produced. The proposed combination is designed to produce class labels of the new subjects and the reasons for this assignment. Table 5.9 shows examples of reasoning procedure which is created by tracking the nodes of the tree in Figure 5.6 and checking related gait attributes. The physician is able to see which controls are done to subject for detecting grade of his disease by using this table. The affected gait attribute field decreases the number of graphs from 33 to 6 that physician may need to analyze. These are important for treatment planning and supplying immediate feedback to the subject.

The misclassified examples are also added to the table to be able to discuss reasons for wrong classification. The reason for most of these misclassified subjects is that they have been assigned to wrong expert MLP in decision tree. This wrong assignment is most probable because of pain level which is a subjective feature determined by subject himself. For example, a subject from grade 1 can determine his pain level as 10 (max value for pain level) while the other from grade 3 can say 3. These kind of subjective features are not preferred in classification processes but experts and medical studies in literature show that pain level is one of the important indicators of the selected illness, so added to datasets here.

Table 5.9: Examples of the test set classification

Subject ID	Pain<2	Age<55	Step time <0,64	Pain<6	BMI<31	Affected gait attributes	Actual Kellgren	Assigned Kellgren
S1	yes	-	-	-	-	F.M.Dor., H.Flex, K.M.Flex, K.Flex, P.Tilt, F.Rot	0	0
S2	no	no	no	-	no	H.P.Abd, H.Flex, A.P.Dor, H.Rot, H.P.Flex, KVal	3	3
S3	no	no	no	-	no	H.P.Abd, H.Flex, A.P.Dor, H.Rot, H.P.Flex, KVal	3	3
S4	no	no	yes	yes	-	H.P.Abd, H.Flex, A.P.Dor, H.Rot, H.P.Flex, KVal	2	2
S5	no	no	yes	no	-	H.P.Abd, H.Flex, A.P.Dor, H.Rot, H.P.Flex, KVal	1	2
S6	no	yes	-	-	-	K.Flex, K.M.Flex, K.P.Flex, H.P.Tot, K.P.Tot, K.Val	3	2
S7	no	no	yes	no	-	H.P.Abd, H.Flex, A.P.Dor, H.Rot, H.P.Flex, KVal	2	2
S8	no	no	no	-	no	H.P.Abd, H.Flex, A.P.Dor, H.Rot, H.P.Flex, KVal	1	2
S9	no	yes	-	-	-	K.Flex, K.M.Flex, K.P.Flex, H.P.Tot, K.P.Tot, K.Val	1	1

5.7. OVERALL ANALYSIS OF THE RESULTS

- Comparing accuracy rate of our algorithm with the manual grading done by expert physicians it can be summarized that; the physicians can detect the knee OA by evaluating gait data only**.** Physicians use radiographic imaging techniques to grade severity of knee OA and use Kellgren-Lawrence score as gold standard. There is no study about using gait data for grading of the OA. This study represents a new approach for estimating Kellgren-Lawrence grades of the subjects by using only gait data with an accuracy rate of 80% which is an acceptable rate to use it as a clinical test.

- The causes and compensatory effects of the knee OA has been searched by some studies before **[20, 26, 29]**. They concluded that patients with OA of the knee joint often adapt a gait for alleviating pain, so the motion of the other joints may be affected **[26].** But it is not known if gait adaptation is mainly related to the severity of the disease **[26].** Our data analysis process indicated relations of severity of the OA and the joints affected by gait adaptations. We proved the hypothesis of **[26]** that reduced motion of the knee joint would be compensated by an increased motion of the hip joint.

- The classification success of the implemented combining classifier is proved by comparing the generalization accuracy of it with a single uncombined MLP. It can be concluded that our algorithm performed significantly better than it and supplied some additional advantages like reasoning etc. But, there are still some drawbacks about detection of high grades of the disease. Actually, as the grade of the disease increase, discriminating it from nearest grade becomes more difficult. Most probable reason for this is patients' progressive ability to compensate their gaits by changing gait pattern of some joints. That is why hip joint waveforms selected as more discriminative features than knee joint ones.

- If we discuss the misclassified examples in all levels of the disease, it is seen that subjective features like pain contributes to misclassification rate most. Since the high correlation between the severity of the OA of the knee and pain level is detected by many studies **[20, 26],** it is included in classification process.

CHAPTER 6

OAGAIT DECISION SUPPORT SYSTEM

6.1. CLINICAL DECISION SUPPORT SYSTEMS

Clinical (or **diagnostic**) **decision support systems (CDSS)** can be defined as interactive computer programs assisting physicians and other health professionals with decision making tasks **[72].**

The basic components of a CDSS include a medical knowledge and logical rules derived from experts. There are many computer applications designed to be a CDSS. Programs that perform database search or check drug interactions support decisions, but usually they are not called CDSS. In **[73]** a CDSS is defined as a program that supports a reasoning task, implemented behind the user interfaces and based on the clinical data. For example, a program that takes the laboratory results as inputs and generates a list of possible diseases is recognized as a clinical diagnostic decision support system (CDDSS). General purpose programs accepting clinical findings and generating diagnostic results are typical CDDSSs. These programs use numerical, logical or artificial intelligence techniques to convert clinical data to the information that a physician might use for diagnostic reasoning.

The use of artificial intelligence in medicine started in the early 1970's and produced a number of experimental systems **[73].** INTERNIST I was one of the first CDSSs, designed to support diagnosis. It was a rule-based expert system designed at the University of Pittsburgh in 1974 for the diagnosis of complex problems in general internal medicine. It uses a tree-structured database that links diseases with symptoms. Most valuable product of the system was its medical knowledge base which was used as a basis for successor systems. **[74]**

MYCIN was another rule-based expert system designed to diagnose and recommend treatment for certain blood infections. Clinical knowledge in it is represented as a set of IF-THEN rules. It was a goal-directed system, using a basic backward chaining reasoning strategy. It was developed in Stanford University. The EMYCIN (Essential MYCIN) expert system shell, employing MYCIN's control structures was developed at Stanford in 1980. This domain-independent framework was used to build diagnostic rule-based expert systems **[73]**.

PIP, the Present Illness Program, was a system built by MIT and Tufts-New England Medical Center in the 1970s. They gathered data and generated hypotheses about disease processes in patients with renal disease **[73]**.

The review studies to evaluate the effects of computer based CDSSs on physician performance and patient outcomes have concluded that CDSSs developed in 70s as summarized above can improve clinical performance for drug dosing, preventing care and other aspects of medical care, but not too convincingly for diagnosis. On the other hand, legal issues such as who would be responsible as a result of misdiagnosis also prevented these systems to be commercially accepted. In **[73]** factors that affect the acceptance and use of CDSSs in clinical practice are defined as follows.

- Cost
- Degree of user acceptance prior to and after installation
- Ease of use
- Interoperability: Integration with existing systems (hardware, other devices) and existing software programs (integration with patient record and/or any relevant clinical terminologies)

- Ease of integration within organizational context and routine
- Legal and ethical issues
- User interface: design, structure, number of forms
- Style, manner of presentation of advice/ recommendations/ results to user
- Provision of evidence justifying advice and/or recommendations
- Involvement of local users during development phase

Today, positive aspects of medical experts about computer usage in clinical applications, need of rapid access to recent information and need for time saving increases the number of commercialized CDSSs. Other potential benefits of using electronic CDSSs in clinical practice are grouped in three broad categories **[76]:**

1) Improved patient safety
   a) Reducing medication errors
   b) Improving medication and test ordering
2) Improved quality of care
   a) Increasing clinicians time for patient care
   b) Providing immediate feedback to the patient
   c) Reducing variations in quality of care
   d) Increasing application of clinical pathways and guidelines
   e) Facilitating the use of up-to-date clinical evidence
   f) Improving the clinical documentation and patient satisfaction
3) Improved efficiency in health care delivery
   a) Reducing the cost by faster processing after initial capital cost
   b) Reducing the test duplications

DXplain, QMR, ERA and ATHENA are good examples to commercialized successful systems originating after 80s **[73, 75, 77].** DXplain uses a set of clinical findings (signs, symptoms, and laboratory data) to produce a ranked list of diagnoses which might explain the clinical manifestations. It provides justification for why each of these diseases might be considered, suggests what further clinical information would be useful to collect for each disease, and lists what clinical manifestations, if any, would be unusual or atypical for each of the specific diseases **[77].** DXplain includes 2,200 diseases and 5,000 symptoms in its knowledge base. It is developed by Laboratory of Computer Science, Massachusetts General Hospital, and Harvard Medical School.

QMR has a knowledge base composed of diseases, diagnoses, findings, disease associations and lab information. It includes information about almost 700 diseases and more than 5,000 symptoms, signs, and labs. It was designed for 3 types of use: as an electronic textbook, as an intermediate level spreadsheet for the combination and exploration of simple diagnostic concepts and as an expert consultant program [**75**]. It is developed by the University of Pittsburgh and First Databank in California in 1980.

The ATHENA DSS implements guidelines for hypertension, encourages blood pressure control and recommends guideline-concordant choice of drug therapy. It is designed to allow clinical experts to customize the knowledge base to incorporate new evidence or to reflect local interpretations. It has a independent database so can be integrated into a variety of electronic medical record systems.

6.2. PROPERTIES OF A GOOD CDSS

The implementation of effective CDSS is a challenging task that should involve interactions between technologies and organizations. There are no obvious solutions to guarantee success or to avoid failure in this complex process. There are many factors to reduce errors or to improve health processes, so measuring the effectiveness of decision support systems is difficult. So, evaluation studies of CDSSs have typically aimed to measure the impact of a system on a limited part of the process **[73]**. Evaluated systems are mostly designed for providing support for diagnosis, disease management, drug management or preventive interventions. Other evaluation topics have included the impact of a system on the quality of decision making, impact on clinical actions, usability, integration with workflow, the quality of the clinical advice offered. The cost effectiveness of CDSSs and their ability to help improve clinical outcomes have been infrequently evaluated.

In a review of computer based systems, most (66%) significantly improved clinical practice, but 34% did **not [78].** There is little scientific evidence to explain why systems succeed or fail. Some researchers have tried to identify the system features most important for improving clinical practice by relying on opinion of a limited number of experts. In **[79]** the authors systematically reviewed the literature published up to 2003 to identify features of CDSSs critical for improving clinical practice. Table 6.1 shows the 15 features of CDSSs derived from this study.

Table 6.1: Features of a good CDSS [79]

General system features
Integration with charting or order entry system to support workflow integration
Use of a computer to generate the decision support
Clinician-system interaction features
Automatic provision of decision support as part of clinician workflow
No need for additional clinician data entry
Request documentation of the reason for not following CDSS recommendations
Provision of decision support at time and location of decision making
Recommendations executed by noting agreement

Table 6.1 (cont.)

Communication content features
Provision of a recommendation, not just an assessment
Promotion of action rather than inaction
Justification of decision support via provision of reasoning
Justification of decision support via provision of research evidence
Auxiliary features
Local user involvement in development process
Provision of decision support results to patients as well as providers
CDSS accompanied by periodic performance feedback
CDSS accompanied by conventional education

6.3. FEATURES OF OAGAIT

OAGAIT is designed as a CDSS for grading of the knee OA. On the other hand it has significant differences from commercialized CDSSs explained in examples. It has a small knowledge base and database about knee OA but not for all gait disorders. Its implementation is done by a small amount of money as a part of academic research project.

A grading method is implemented by a combined pattern recognition system and embedded into the OAGAIT system. User friendly interfaces are designed for different modules. The main objective of the OAGAIT is to help physicians for grading of an already diagnosed disease to ease the treatment planning. Moreover, some other functions are designed to guide gait analysis process. As discussed in previous chapters, various structured data such as personal information, time-distance parameters are collected in the gait laboratory for each subject. Some of these are stored in paper files. So it used to be difficult to access and combine data for processing. A complete database is integrated to OAGAIT system to keep it together and to access easily when needed.

In our gait laboratory, the commercial software (VICON) used allows gait data to be saved as MS Excel file. These files show the time-distance parameters of the gait and temporal changes of the joint angles and their graphs. An excel file of a patient report is shown in Appendix A as an example. Electronic interfaces are created for entering this information to the database. When the database is opened the first form to be filled by the user is the patient record form, which is shown in Figure 6.1.

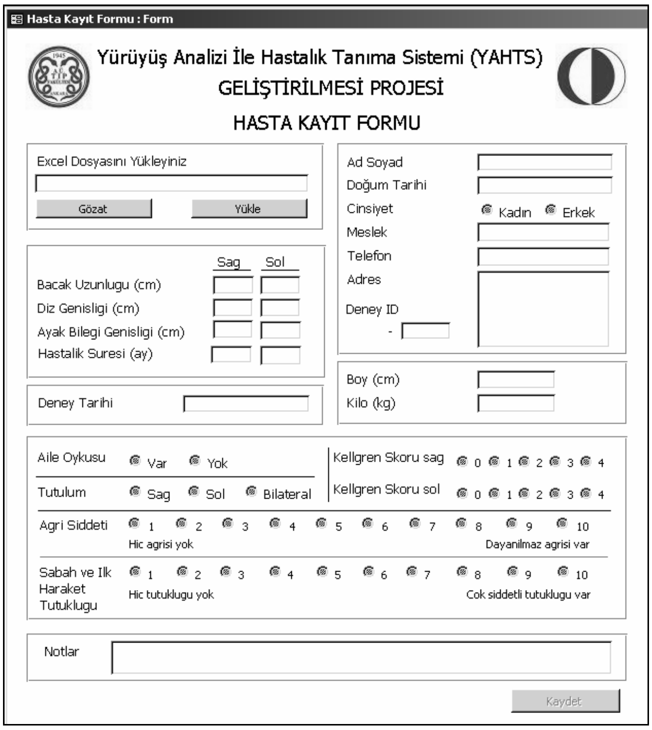


Figure 6.1: Patient recording form

Besides these, patient tracking forms are created to automate some often used queries. For example Figure 6.2 shows the query results for a patient’s time distance parameters and personal information. These types of query forms are important for database user to learn the number and the date of the experiments of the selected subject. Moreover, the physician can compare the changes in the gait parameters by selecting the different experiments from the list.

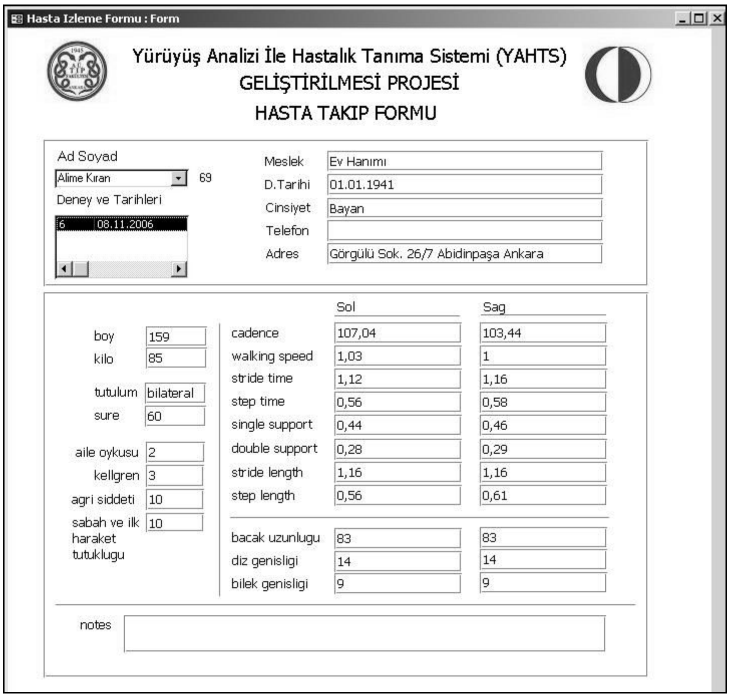


Figure 6.2: Patient tracking form

Another form supplied by the database system is WOMAC entry form. This is a validated test designed specifically for the assessment of lower extremity pain and function in OA of the knee or hip **[27, 29]**. Figure 6.3 shows the WOMAC form of the OAGAIT system.

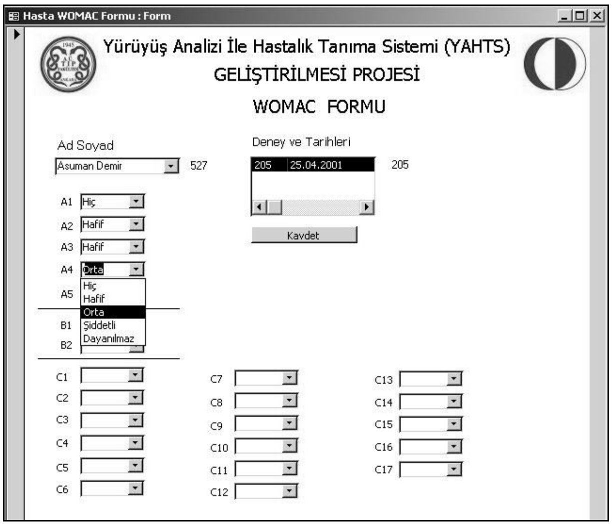


Figure 6.3: WOMAC form

These forms and database tables are used for the main purpose of the system; grading of the OA. An explanatory and easy to use grading screen is designed for grading results of the system (Figure 6.4). The grading algorithm will be explained in more detail in further sections of the report.

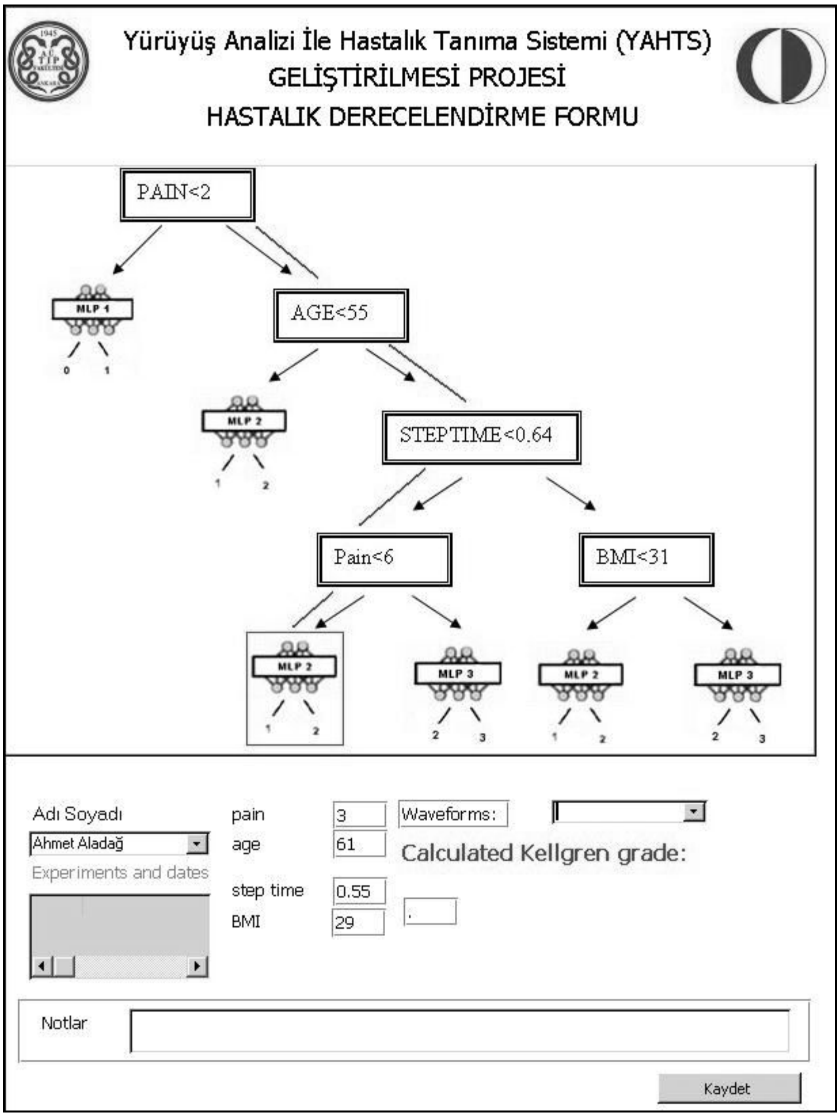


Figure 6.4: Grading screen

6.4. OAGAIT DATABASE

A database with five main tables is used for OAGAIT system. Figure 6.5 shows these and their relations. The relations of the table are constructed by the used database system automatically. These relations are important for consistency of the data, and show the database's structure of how this data is arranged.

The ID entries of the tables represent the unique numbers assigned for the subjects and construct the relations of all tables. For the tables containing information about gait experiment the relation is also provided by experiment ID (expID) entry.

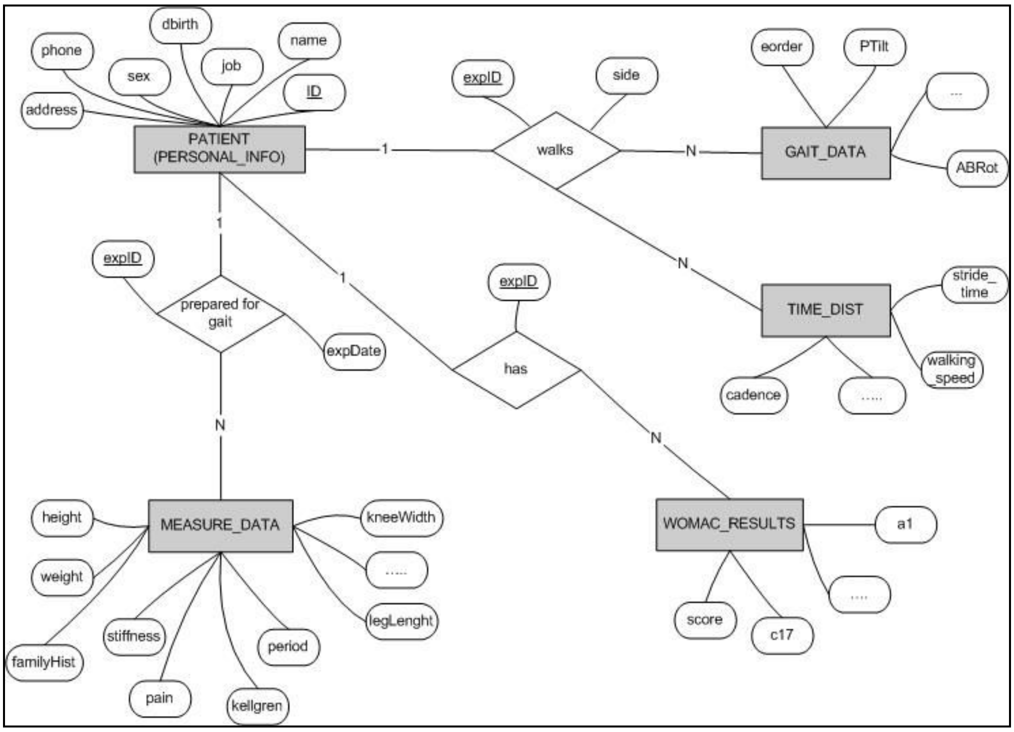


Figure 6.5: ER diagram for OAGAIT database

Some of these tables are created for provision of integration of VICON system and OAGAIT system, and some are designed for storing paper based data on a computer based system. Detailed explanation of the tables and entries are presented in this section.

Table "gait_data": This table is the main table storing temporal variables of the gait and provides integration of VICON system and OAGAIT system. It is accessed by both patient tracking and grading functions. Patient record function writes data to the table.

gait_data : Table

Field Name	Data Type
ID	Number
expID	Number
eorder	Number
side	Text
expDate	Date/Time
PTilt	Number
PObliq	Number
PRot	Number
HFlex	Number
HAbd	Number
HRot	Number
KFlex	Number
KVal	Number
KRot	Number
FDor	Number
FRot	Number
FPro	Number
HMFlex	Number
HMAbd	Number
HMRot	Number
KMFlex	Number
KMVal	Number
KMRot	Number
FMDor	Number
FMAbd	Number
FMRot	Number
HPTot	Number
HPFlex	Number
HPAbd	Number
HPRot	Number
KPTot	Number
KPFlex	Number
KPVal	Number
KPRot	Number
APTot	Number
APDor	Number
APAbd	Number
APRot	Number

Figure 6.6: Table: gait_data

Entries: It has 38 entries as shown in Figure 6.6, 33 of those are related to temporal gait variables and the rest is about experiment that the subject is walked. The ID entry of this table is read from VICON system which assigns a numeric ID to each subject during first visit. Experiment data (expDate) keeps the date of the gait and used for deriving "age" feature of the subjects with date of birth entry of another table. Experiment ID (expID) is another system assigned number for each gait trial of the subject. So the experiments done in different times are stored by a unique number and allow the patient tracking property of OAGAIT. Eorder is an automatically generated number to show order of the time sample points for temporal variables of gait. Side entry is a single character value (L stands for left and R stands for right) shows which knee of the patient is affected by OA.

Table "womac_scores": This table is responsible for keeping answers of the subjects to the WOMAC questionnaire. The WOMAC form of the database allows subject ID selection and writes his/her answers to table.

womac_results : Table

Field Name	Data Type
ID	Number
expID	Number
a1	Number
a2	Number
a3	Number
a4	Number
a5	Number
b1	Number
b2	Number
c1	Number
c2	Number
c3	Number
c4	Number
c5	Number
c6	Number
c7	Number
c8	Number
c9	Number
c10	Number
c11	Number
c12	Number
c13	Number
c14	Number
c15	Number
c16	Number
c17	Number
score	Number

Figure 6.7: Table: Womac_results

Entries: This table has 20 entries as shown in Figure 6.7, 17 of which are related to WOMAC questionnaire answers of the subjects. ID and expID entries are primary key of the table. The entries from a1 to c17 are numbers taking values 0, 1, 2, 3, 4 or 5 according to answers. The system calculates the WOMAC scores of the subjects by summing up these values and writes the results to "score" of this table.

Table "time_dist": Data of this table are entered by automatic file reading function of the OAGAIT. The parameters are read from the VICON system and written to the table by patient record function. It is accessed by patient tracking and grading functions of the system.

time_dist : Table

Field Name	Data Type
ID	Number
expID	Number
side	Text
cadence	Number
walkingSpeed	Number
strideTime	Number
stepTime	Number
singleSupport	Number
doubleSupport	Number
strideLength	Number
stepLength	Number

Figure 6.8: Table: time_dist

Entries: This table stores the time distance parameters of the gait as shown in Figure 6.8. It has 11 entries 8 of which are time distance parameters and first three are about subject and experiment information as in other tables.

Table "personal_info": This table is created for storing personal information of the subjects which were saved in paper files before. Figure 6.9 shows the entries and data types of the table.

personal_info : Table

Field Name	Data Type
ID	Number
name	Text
job	Text
dbirth	Date/Time
sex	Number
phone	Text
address	Text

Figure 6.9: Table: personal_info

Entries: This table has seven entries most of which are not used for grading or tracking functions. Only the date of birth (dbirth) entry is used for deriving "age" features of the subjects.

Table "measure data": This table stores the measurements of the subjects that are taken just before the gait. These measurements are used for calculation of time-distance parameters and temporal variables of the gait by VICON system. Left and right side of the subjects are measured separately and the initial characters (“r” or “l”) of the entries represent these sides as shown in Figure 6.10. The measurements are done by a laboratory expert who writes these numbers to a paper form. Then this information is saved to the OAGAIT database by patient record function.

measure_data : Table

Field Name	Data Type
ID	Number
expID	Number
expDate	Date/Time
direction	Text
rlegLength	Number
llegLength	Number
rkneeWidth	Number
lkneeWidth	Number
rankleWidth	Number
lankleWidth	Number
height	Number
weight	Number
familyHist	Number
lperiod	Text
rperiod	Text
lkellgren	Number
rkellgren	Number
acheLevel	Number
firstMovAche	Number
notes	Memo

Figure 6.10: Table: measure_data

Entries: This table has 20 entries, 17 of which are about measurements of the subjects. The height and weight entries are used for calculation of BMI features of the subjects. acheLevel and firstMovAche entries are used as pain and stiffness features, respectively. These features together with BMI feature are used for grading function of the OAGAIT system.

6.5. DATA FLOW DIAGRAMS

A **data flow diagram** (**DFD**) is a graphical representation of the "flow" of data through an information system. A DFD is mostly used for the visualization of data processing. A first level DFD is called context-level DFD which shows the interaction between the system and the outside entities. This context-level DFD is then "exploded" to show more detail of the system.

Figure 6.11 shows level 1 DFD of the designed decision support system. Integration of VICON Clinical Manager and OAGAIT is also shown here.

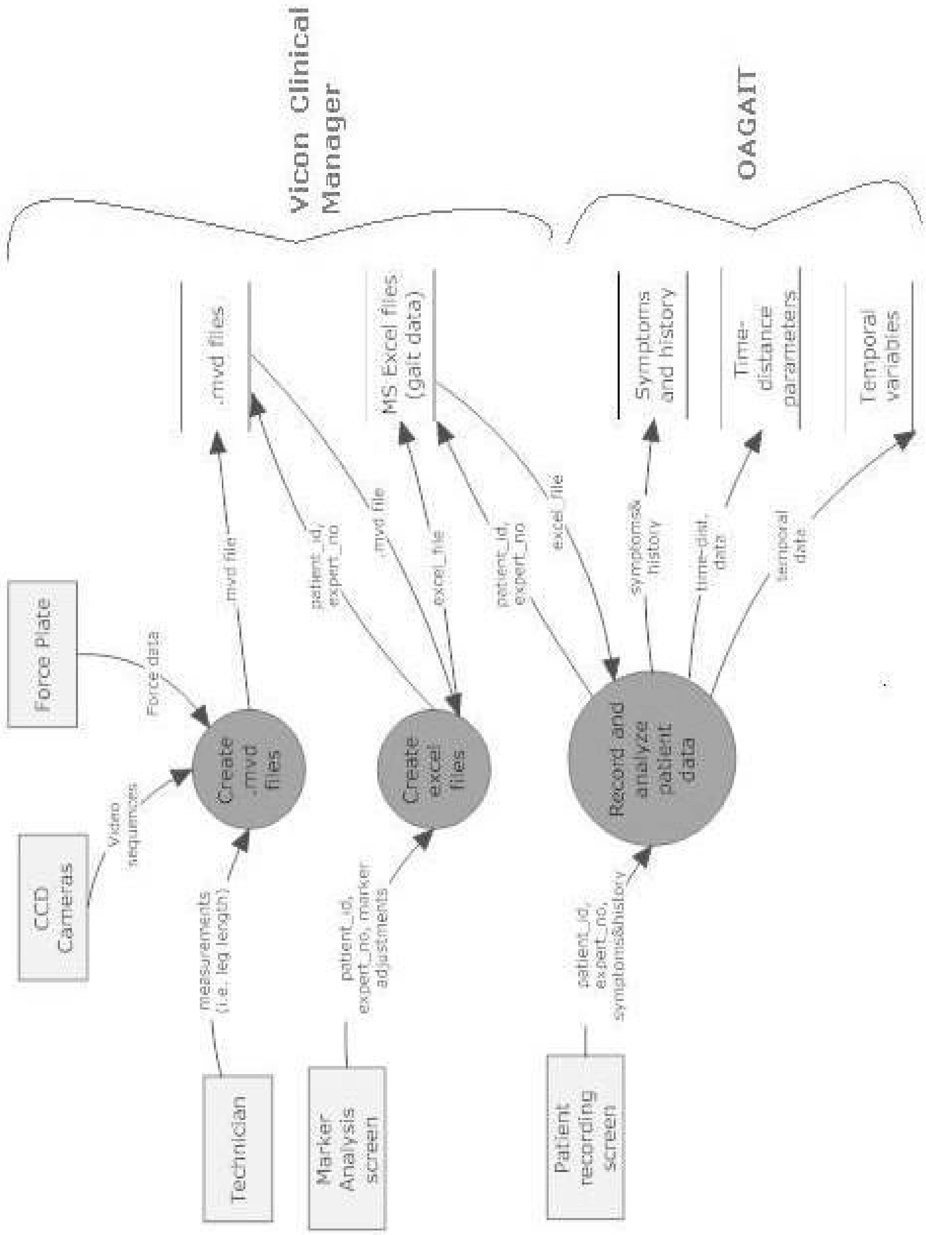


Figure 6.11: Level 1 DFD of the OAGAIT for data recording

After recording the gait data to OAGAIT database, two main screens can be used for tracking the patients and grading their illness. DFD for patients tracking process is shown in Figure 6.12.

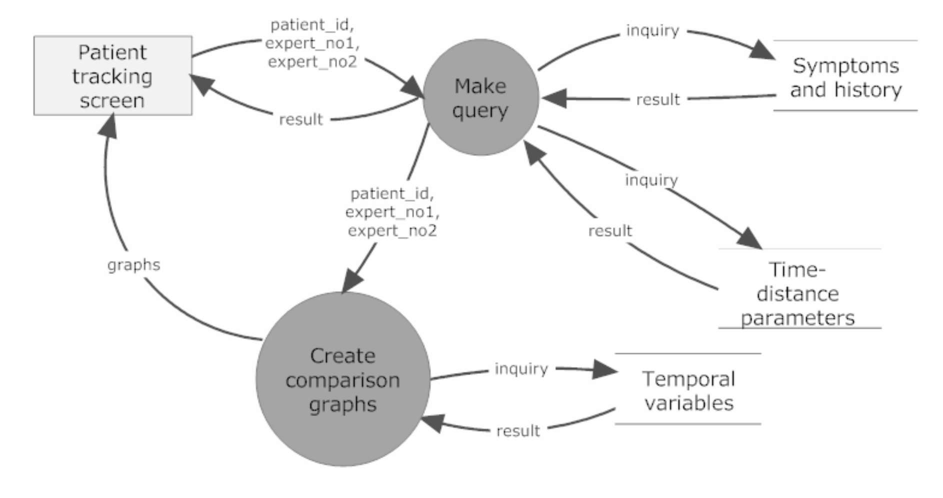


Figure 6.12: DFD for patient tracking process

Actually patient tracking is a database query function which enables the comparison of two different gait experiments of the patient taken in different times. So, the physician can analyze the recovery of the illness.

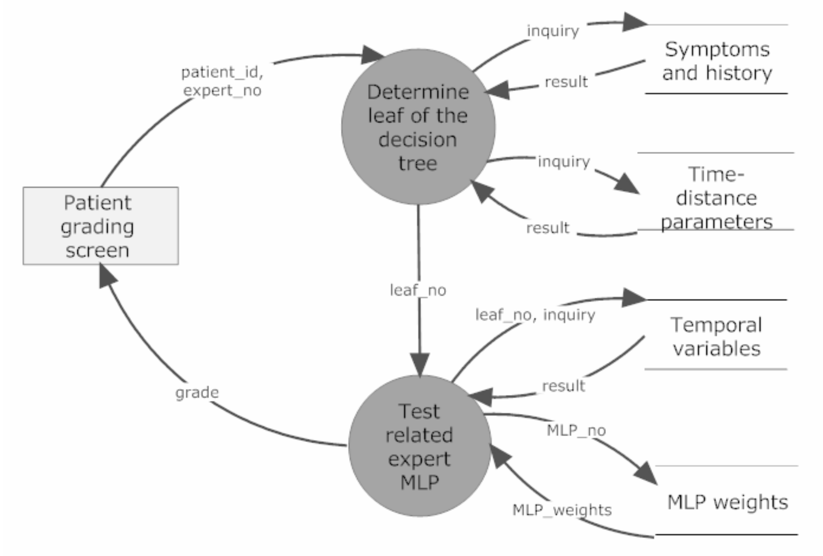


Figure 6.13: DFD for patient grading process

The grading process is mainly composed of two stages as seen in Figure 6.13. Since the classification algorithm is created by combining two different classifiers, these stages represent the testing of them by new gait data. Design of the classification algorithm is detailed in further sections.

6.6. EVALUATION OF OAGAIT AS A CDSS

A comparison table is created for evaluation of OAGAIT system as a CDSS. The features of a good CDSS suggested by Kuwamoto **[79]** are searched for OAGAIT system and the existing features are explained shortly as shown in Table 6.2. It can be seen that almost all features are supplied by OAGAIT system.

Table 6.2: Features of OAGAIT compared to the ones suggested in [79]

Features of a good CDSS	Features of OAGAIT
General system features	
Integration with charting or order entry system to support workflow integration	Integration of OAGAIT and VICON system
Use of a computer to generate the decision support	Fully computerized decision support
Clinician-system interaction features	
Automatic provision of decision support as part of clinician workflow	When the subject's gait data is entered to the database the grading info is automatically displayed on the screen.
No need for additional clinician data entry	All needed data is entered to the database before processing
Request documentation of the reason for not following CDSS recommendations	There is a additional notes entry in all forms
Provision of decision support at time and location of decision making	The grading results are shown on the screen just after the walking of subject
Recommendations executed by noting agreement	Not applicable

Table 6.2 (cont.)

Communication content features

Provision of a recommendation, not just an assessment	OAGAIT supply reasoning for the assessment to help creation of treatment plans
Promotion of action rather than inaction	The system should advise something rather than a blank screen. OAGAIT produces most probable two classes as a result rather than "not classified" message.
Justification of decision support via provision of reasoning	OAGAIT shows assessment stages to support reasoning
Justification of decision support via provision of research evidence	The decision tree property of OAGAIT supplies research evidences for provision of OA

Auxiliary features

Local user involvement in development process	An expert physician is included in the development process as both an knowledge expert and end user
Provision of decision support results to patients as well as providers	Physician is responsible for delivering the results to the patients
CDSS accompanied by periodic performance feedback	Not applicable
CDSS accompanied by conventional education	A short training is given to the physicians and/or other laboratory staff

CHAPTER 7

CONCLUSION AND FUTURE DIRECTIONS

7.1. PROPERTIES OF OAGAIT SYSTEM

Within the scope of this study, a CDSS was implemented to help physicians for grading and further analysis of the knee OA. Main function of OAGAIT is to interpret gait data almost as close as an expert's. This interpretation is done by using expert knowledge on gait and other features in pattern recognition. Main features of the implemented CDSS can be summarized as following:

- OAGAIT supports the function of radiographic films for grading of OA and/or other diseases.
- OAGAIT is a fully computerized system, so incorrect decisions by experts as a result of non-experienced interpretations is minimized.
- It provides a base for patient follow up in time, which helps physicians to make more accurate treatment plans.
- It provides a graphical representation (the decision tree) of the grading process which brings a description to the decision in addition to classification.

- OAGAIT has a complete gait database integrated with the data collection software VICON. This database combined all new and old gait data in an easy access and portable environment. Moreover, this database is convenient to use for further studies about other diseases or integration to other software systems.
- It has easy-to-use user interfaces, so a short training is enough for the physicians and/or other laboratory staff
- It has a short processing time; the grading results are shown on the screen just after walking of the subject. So immediate feedback to the physician and the patient is supplied.

7.2. EVALUATION OF THE GRADING ALGORITHM

Combining classifiers produced premising results for many areas in pattern recognition. Since we deal with a multi-class problem in this study, expert classifiers for different classes are combined for better generalization accuracy. The implemented combination schema is expected to increase success rates of similar medical problems.

Since the classification algorithm was developed using a method similar to spiral development methodology, the result of one stage is important for design of the next one. This means, feature reduction/selection and classification algorithms are selected in a series of successive trials of increasing complexity. Each stage of the algorithm design produced valuable information about further recognition of the selected disease, helping treatment plans.

A grading algorithm is created by combining decision trees and MLPs by considering the results of previous experiments. The symptoms and history information of the subjects in addition to gait data are also included in the combination. This data is used to train a decision tree, which gives an opportunity of the reasoning of the results. The gait data is used to train 3 different MLPs with binary classifications that are used at the leaves of the decision tree. The feature selection processes prior to decision tree and MLP training give us information about relations between the grade of the illness and the affected body parts. Namely, while the subjects with low grade of the disease (grade 1 or 2) have deformation in knee joint, the ones with high grade of the disease (grade 3) have more deformation in hip joint. Although we analyze a knee disease, we see that other parts of the body may be affected and so data from these parts should also be included in the classification processes. Deriving this kind of hidden information in data provides better clinical recognition of the illness while contributing the classification accuracy.

Comparing the accuracy of the implemented classifier with a single multi-class one, it can be concluded that combining a set of binary classifiers produce better results than a single multi-class one. Since the classes are not easily distinguishable creating different experts for different subsets of the feature sets produce better results. But still classification accuracy of the expert MLPs may be improved. Especially for detection of third grade of the illness some more detailed analysis may be helpful.

As a final comment, the results and analysis of the classifier both for accuracy and descriptiveness produced satisfactory results and aimed to be used in gait laboratories. The combined decision tree-MLP approach is also expected to be applicable to similar type of medical decision making processes, where both disease characteristics and clinical measurements and tests are to be combined.

7.3. LIMITATIONS OF THE STUDY

In pattern recognition studies the curse of dimensionality is a significant reason for poor generalization ability of classifiers. In practice it is often observed that the added features may degrade the performance of a classifier if the number of training samples that are used to design the classifier is small relative to the number of features. This generalization is valid for our study, too. Even though the amount of collected data far exceeds the data size used by other studies, it is still not fully satisfactory. If the number of samples was arbitrarily large then more features would be included for both creating decision tree and training MLPs. Most probable, including more features would produce better generalization accuracies.

As said before the implemented combination algorithm may be used for detection and grading of similar type of diseases. Because of time and budget restrictions data collection process limited to only one disease. The algorithm could not be tested by any other set of data.

Another limitation of the study was about data collection process. In most studies OA is scaled to five grades (0-4) according to Kellgren-Lawrence radiographic method. Since most of the fourth grade patients are not able to walk in laboratory and treated in bed in orthopedics department of the hospitals, to collect their gait data is difficult. Therefore, in this study data from fourth grade of the OA is ignored in classification process.

7.4. SUGGESTIONS FOR FUTURE WORK

OAGAIT is currently proposed for use in Ankara University Medicine Faculty Gait Laboratory. It may be installed to other gait laboratories and tested by many experts in the future. Testing results may be used for improvement of the system. If data from other diseases like cerebral palsy (CP) can be added to the database, it will be preferred by more laboratories in the future. If system is used widely, a web based collective database system may be designed to help the data sharing between laboratories. Fast increase in number of samples will lead to design different classifiers and further analysis of the diseases. Moreover, large amounts of data may allow the data mining studies by which hidden knowledge in medical data may be discovered.

In addition to the detection and grading of the diseases, patient follow up property which is significant for clinical decision making may be added to the system. To achieve this, the subjects should have gait data collected in determined time intervals. Therefore a subject should be called to the gait laboratory after some periods of time like 6 months, 1 year etc. For the design of this function, extrapolation methods may be tried by including “time” parameter in the feature set.

Combining pattern classifiers is one of the recent, popular research areas of machine learning. Lots of new combining methods and approaches are implemented everyday. In this study we could try some of them with existing amount of data. However, some other feature reduction/selection methods or different combination models may also be tried to optimize the classification accuracies. Also combination of pattern recognition algorithms and image processing methods may be tried for evaluating gait analysis data and XR images for grading.

REFERENCES

1. L. Lee, W. Grimson, "**Gait analysis for recognition and classification**", in *Proceedings of the IEEE Conference on Face and Gesture Recognition*, 2002, pp. 148-155

2. R. Begg,y, J. Kamruzzaman, **"A Comparison of Neural Networks and Support Vector Machines for Recognizing Young-Old Gait Patterns**", *Proceeding of IEEE TENCON Conference*, 2003, pp. 354-358

3. C. Y. Yam, M. S. Nixon, J. N. Carter, "**Automated person recognition by walking and running via model-based approaches**". *Pattern Recognition,* vol. 37, pp. 1057-1072 , 2004

4. M. S. Nixon, J. N. Carter. "**Advances in Automatic Gait Recognition**," in *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition,* 2004, pp. 39-144

5. S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, K. W. Bowyer, "**The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis**", *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 27, pp. 162-177, 2005

6. L. Wang, W. Hu, T. Tan, "**A New Attempt to Gait-Based Human Identification**," in *Proceedings of the International Conference on Pattern Recognition*, 2002, pp. 115-118

7. R. Collins, R. Gross, J. Shi, "**Silhouette-Based Human Identification from Body Shape and Gait**," in *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 366-371

8. L. Herda, P. Fua, R. Plankers, R. Boulic, D. Thalmann **"Skeleton-Based Motion Capture for Robust Reconstruction of Human Motion",** in *Computer Animation* , 2000, pp. 77-82

9. R. Urtasun, P. Fua, "**3D Tracking for Gait Characterization and Recognition**", in *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition,* 2004, pp. 17-22

10. A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, R. Chowdhury, V. Krüger, R. Chellappa, "**Identification of Humans Using Gait**", *IEEE Transactions On Image Processing*, vol. 13, pp.1163-1173, 2004

11. R. B. Davis, P. A. DeLuca, M. J. Romness, "**Clinical Gait Analysis and Its Role in Treatment Decision-Making**", *Medscape General Medicine* vol. 1, 1999

12. K. R. Kaufman, "**SECTION FOUR: Future Directions in Gait Analysis**", *RRDS Gait Analysis in the Science of Rehabilitation*, May 2005, http://www.vard.org/mono/gait/kaufman.htm,

13. R. Baker, "**Gait Analysis Methods in Rehabilitation**", *Journal of Neuro-Engineering and Rehabilitation,* vol. 3, pp. 1-10, 2006

14. C. L Vaughan, B. L. Davis, J. C. O'Conner, *Dynamics of Human Gait*. Champaign, IL: Human Kinetics, 1991

15. G. Yavuzer, "**The use of computerized gait analyses in the assessment of Neuromusculoskeletal disorders**", *FTR Bil Der - J PMR Sci,* vol. 2, pp. 43-45, 2007

16. S. Salarian, H. Russmann, F. J. G. Vingerhoets, C. Dehollain, Y. Blanc, P. R. Burkhard, K. Aminian, "**Gait Assessment in Parkinson's Disease: Toward an Ambulatory System for Long-Term Monitoring**", *IEEE Transactions On Biomedical Engineering*, vol.51, pp. 1434-1443, 2004

17. R. E. Cook, I. Schneider, M.E. Hazlewood, S.J. Hillman, J.E. Robb, "**Gait analysis alters decision-making in cerebral palsy**". *J Pediatr Orthop*, 23-3: pp.292-295, 2003

18. Salazar, A.J. De Castro, O.C. Bravo, R.J., "**Novel approach for spastic hemiplegia classification through the use of support vector machines**", *Proceedings of the 26th Annual International Conference of the Engineering in Medicine and Biology Society,* 2004, pp. 446-469

19. F. Dobson, M. E. Morris, R. Baker, H. K. Graham, "**Gait classification in children with cerebral palsy: A systematic review**", *Gait and Posture,* vol. 25, pp. 140-52, 2007

20. H. Gök, S. Ergin, G. Yavuzer, "**Kinetic and kinematic characteristics of gait in patients with medial knee arthritis**" *, Acta Orthop Scand*, vol. 73, pp. 647–652, 2002

21. JH Kellgren, JS Lawrence, '**Radiological assessment of osteoarthritis'**, *Ann Rheum Dis,* vol. 16, pp. 494-501, 1957

22. *Osteoarthritis (OA): Bone, Joint, and Muscle Disorders: Merck Manual Home Edition* October 2007, http://www.merck.com/mmhe/sec05/ch066/ch066a.html

23. *Osteoarhritis- Risk Factors,* December 2006 http://adam.about.com/reports/000035_4.htm

24. K. J. Deluzio, J. L. Astephen, "**Biomechanical features of gait waveform data associated with knee osteoarthritis: An application of principal component analysis**", *Gait and Posture,* vol. 25, pp. 86-93, 2007

25. *Kaufman, K., Hughes, C., Morrey, B., Morrey, M., An, K.,* "**Gait characteristics of patients with knee osteoarthritis***", Journal of Biomechanics,,* vol. 34, pp. 907–915, 2001

26. Z. Bejek, R. Paroczai, A. Illyesi, L. Kocsis, R. M. Kiss, "**Gait Parameters Of Patients With Osteoarthritis Of The Knee Joint**", *Physical Education and Sport,* vol. 4, pp. 9-16, 2006

27. F Petersson, T. Boegard, T. Saxne, A. J. Silman, B. Svensson, "**Radiographic osteoarthritis of the knee classified by the Ahlbäck and Kellgren & Lawrence systems for the tibiofemoral joint in people aged 35–54 years with chronic knee pain**", *Annals of the Rheumatic Diseases,* vol. 56, pp. 493–496, 1997

28. N. Samancı, C.Kaçar, M. Sayın, T. Tuncer, "**Primer Diz Osteoartritinde Metabolik, Endokrin Ve Sosyo-Kültürel Risk Faktörleri Ve Radyolojik Bulgularla İlişkisi**", *Romatizma,* vol.18, 2003

29. L. Sharma, S. Cahue, J. Song, K. Hayes, Y. C. Pai, D. Dunlop, "**Physical Functioning Over Three Years in Knee Osteoarthritis" Role Of Psychosocial, Local Mechanical, And Neuromuscular Factors**", *Arthritis & Rheumatism*, vol. 48, pp 3359–3370, 2003

30. K. S. Al-Zahrani, A. M. O. Bakheit, "**A Study of the Gait Characteristics of Patients with Chronic Osteoarthritis of the Knee**", *Disability and Rehabilitation*, vol. 24, pp. 275-280, 2002

31. T. Chau, "**A review of analytical techniques for gait data, Part 1: Fuzzy, statistical and fractal methods**," *Gait and Posture*, vol. 13, pp. 49-66, 2001.

32. T. Chau, "**A review of analytical techniques for gait data, Part 2: Neural network and wavelet methods**", *Gait and Posture*, vol. 13, pp. 102-120, 2001.

33. G. Barton, P. Lisboa, A. Lees, S. Attfield, "**Gait quality assessment using self-organising artificial neural networks**", *Gait and Posture*, vol. 25, pp. 374-379, 2007

34. M. Kohle , D. Merkl , J. Kastner, "**Clinical gait analysis by neural networks: issues and experiences**", in *Proceedings of the 10th IEEE Symposium on Computer-Based Medical Systems* 1997, pp.138

35. W. Wu, F. Su, Y. Cheng, Y. Chou, **"Potential of the Genetic Algorithm Neural Network in the Assessment of Gait Patterns in Ankle Arthrodesis**", *Annals of Biomedical Engineering,* vol. 29, pp. 83–91, 2001

36. J. G. Barton, A. Lees, "**An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams"** *Gait and Posture,* vol. 5, pp. 28-33, 1997

37. M. Kohle , D. Merkl, "**Analyzing human gait patterns for malfunction detection**", in *Proceedings of the ACM symposium on Applied computing,* 2000, pp. 41-45

38. N. Şen Köktaş, N. Yalabık, "**A Neural Network Classifier for Gait Analysis**", in *Proceeding of International Symposium on Health Informatics and Bioinformatics, 2005,* pp. 174-179

39. N. Sen Koktas, N. Yalabik, G. Yavuzer, "**Ensemble Classifiers for Medical Diagnosis of Knee Osteoarthritis Using Gait Data**", in *Proceeding of IEEE International Conference on Machine Learning and Applications,* 2006, pp. 225-230

40. R. Lafuente, JM. Belda, J. Sanchez-Lacuesta, C. Soler, J. Prat J. "**Design and test of neural networks and statistical classifiers in computer aided movement analysis: a case study on gait analysis**", *Clinical Biomechanics*, vol. 13, pp. 216–20, 1997

41. R. K. Begg*,* M. Palaniswami*,* B. Owen, "**Support Vector Machines for Automated Gait Classification**", *IEEE Transactions On Biomedical Engineering*, vol. 52, pp. 828-838, 2005

42. L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004

43. J. Kittler, M. Hatef, R. P. W. Duin, J. Matas, "**On Combining Classifiers**", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, pp. 226-239, 1998

44. R. Ranawana, V. Palade, "**Multi classifier systems: Review and Roadmap for developers**", *International Journal of Hybrid Intelligent Systems*, vol. 3, pp. 35-61, 2006

45. M. I. Jordan, R. A. Jacobs, "**Hierarchical mixtures of experts and the EM algorithm**", *Neural Computation*, vol. 6, pp. 181-214, 1994

46. J. Mao, "**A case study on bagging, boosting and basic ensembles of neural networks for OCR**", in *Proceeding of IJCNN*, 1998, pp.1828-1833

47. J. Kim, J. Ahn., S. Cho, "**Ensemble competitive learning neural networks with reduced input dimension**", *International Journal of Neural Systems*, vol. 6, pp 133-42, 1995

48. A. J. C. Sharkey, "**Types of multinet system**", in *Proceedings of the Third International Workshop on Multiple Classifier Systems,* 2002, pp. 108-117

49. Z. H. Zhou, J. Wu, and W. Tang, "**Ensembling neural networks: many could be better than all**", *Artificial Intelligence*, vol. 137, pp.239-263, 2002

50. L. K. Hansen, P. Salamon, "**Neural Network Ensembles**", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 12, pp. 993-1001, 1990

51. N. Ueda, "**Optimal Linear Combination of Neural Networks for Improving Classification Performance**", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 207-215, 2000

52. G. Fumera, F. Roli, "**A Theoretical and Experimental Analysis of Linear Combiners for Multiple Classifier Systems**", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, pp.942-956, 2005

53. T. K. Ho, J. J. Hull, S. N. Srihari, "**Decision Combination in Multiple Classifier Systems**", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 66-75, 1994

54. R. E. Banfield, L. O. Hall, K. W. Bowyeri, D. Bhadoria, W. P. Kegelmeyer, S. Eschrich, "**A Comparison of Ensemble Creation Techniques**" , in *Proceeding of the International Conference on Multiple Classifier Systems, 2004,* pp. 223-232

55. T.G. Dietterich. "**Ensemble methods in machine learning**", in *Proceeding of Multiple Classifier Systems*, 2000, pp. 1-15

56. A. C Stasis, E. N Loukis, S. A Pavlopoulos, D Koutsouris, "**A multiple decision trees architecture for medical diagnosis: The differentiation of opening snap, second heart sound split and third heart sound**", *Computational Management Science,* vol. 1, pp. 245-274, 2004

57. S. Armand, E. Watelain, M. Mercier, G. Lensel, F. Lepoutre, "**Linking clinical measurements and kinematic gait patterns of toe-walking using fuzzy decision trees**", *Gait and Posture,* vol. 25, pp. 475-484, 2006

58. R. Kohavi, "**Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid**", in *Proceeding of International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 202-207

59. V. Podforelec, P. Kokol, "**Evolutionary construction of medical decision trees**", in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 1998, pp. 1202-1205

60. K. Tsujino, "**Hybrid Knowledge Acqusition by Integrating Decision Trees and Neural Networks**", in *Proceeding of IEEE International Conference on Neural Networks,* 1995, pp. 1379-1383

61. J. R. Quinlan, "**Learning Decision Tree Classifiers**", *ACM Computing Surveys*, vol. 28, pp. 71-72, 1996

62. *Decision Tress and Data Mining*, May, 2007, http://www.decisiontrees.net/

63. F. Esposito, D. Malerba and G. Semeraro, "**A comparative analysis of methods for pruning decision trees**", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, pp. 476–491, 1997

64. Anil K. Jain , Robert P. W. Duin , Jianchang Mao, "**Statistical Pattern Recognition: A Review**", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, pp. 4-37, 2000

65. R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification.* John Wiley and Sons, New York, 2001

66. M. Skurichina, R.P.W Duin, "**Combining Feature Subsets in Feature Selection**", in *Proceeding of the International Conference on Multiple Classifier Systems*, 2005, pp. 165–175

67. R.P.W Duin, PRTOOLS (version 4). A Matlab toolbox for pattern recognition. Pattern Recognition Group, Delft University of Technology, February 2004

68. A.K. Jain and D. Zongker, "**Feature Selection: Evaluation, Application, and Small Sample Performance**," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19, pp. 153-158, 1997

69. M. Kudo, J. Sklansky, "**Comparison of algorithms that select features for pattern classiffers**", *Pattern Recognition*, vol. 33, pp. 25-41, 2000

70. M. Dash, H. Liu, "**Feature selection for classification**", *International Journal of Intelligent Data Analysis*, vol. 1, pp. 131-156, 1997

71. A. R. Webb, *Statistical Pattern Recognition*, Second Edition, Wiley, 2002

72. *The free encyclopedia*, January 2008, http://en.wikipedia.org

73. *Decision Support Systems*, January 2008, http://www.openclinical.org/dss.html

74. R. A. Miller, "**Medical diagnostic decision support systems--past, present, and future: a threaded bibliography and brief commentary**", *Journal of the American Medical Informatics Association, vol.* 1, pp. 8-27, 1994

75. R. A. Miller, F.E. Masarie, "**Use of the Quick Medical Reference (QMR) program as a tool for medical education**", *Methods of Information in Medicine,* vol. 28, pp.340-5, 1989

76. E. Coiera. *The Guide to Health Informatics (2nd Edition)*. Arnold, London, October 2003.

77. *DXPlain*, January 2008, http://lcs.mgh.harvard.edu/projects/dxplain.html

78. D. L. Hunt, R. B. Haynes, S. E. Hanna, K. Smith, "**Effects of Computer-Based Clinical Decision Support Systems on Physician Performance and Patient Outcomes: A Systematic Review",** *The Journal of the American Medical Association,* vol. 280, pp. 1339-1346, 1998

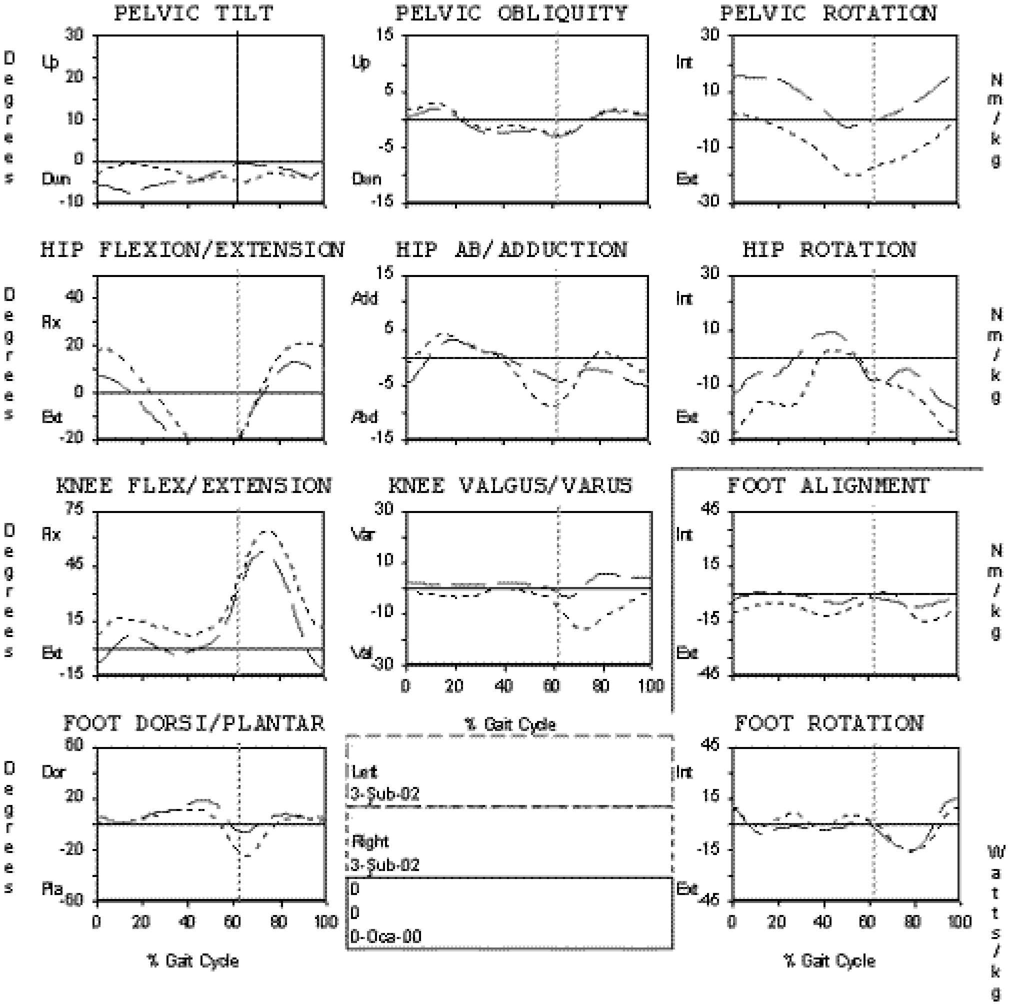
79. K. Kawamoto, "**Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success**", BMJ, vol. 330, pp. 330-765, 2005

APPENDICES

APPENDIX A. An example to excel file of a subject

VICON Clinical Gait Analysis Report

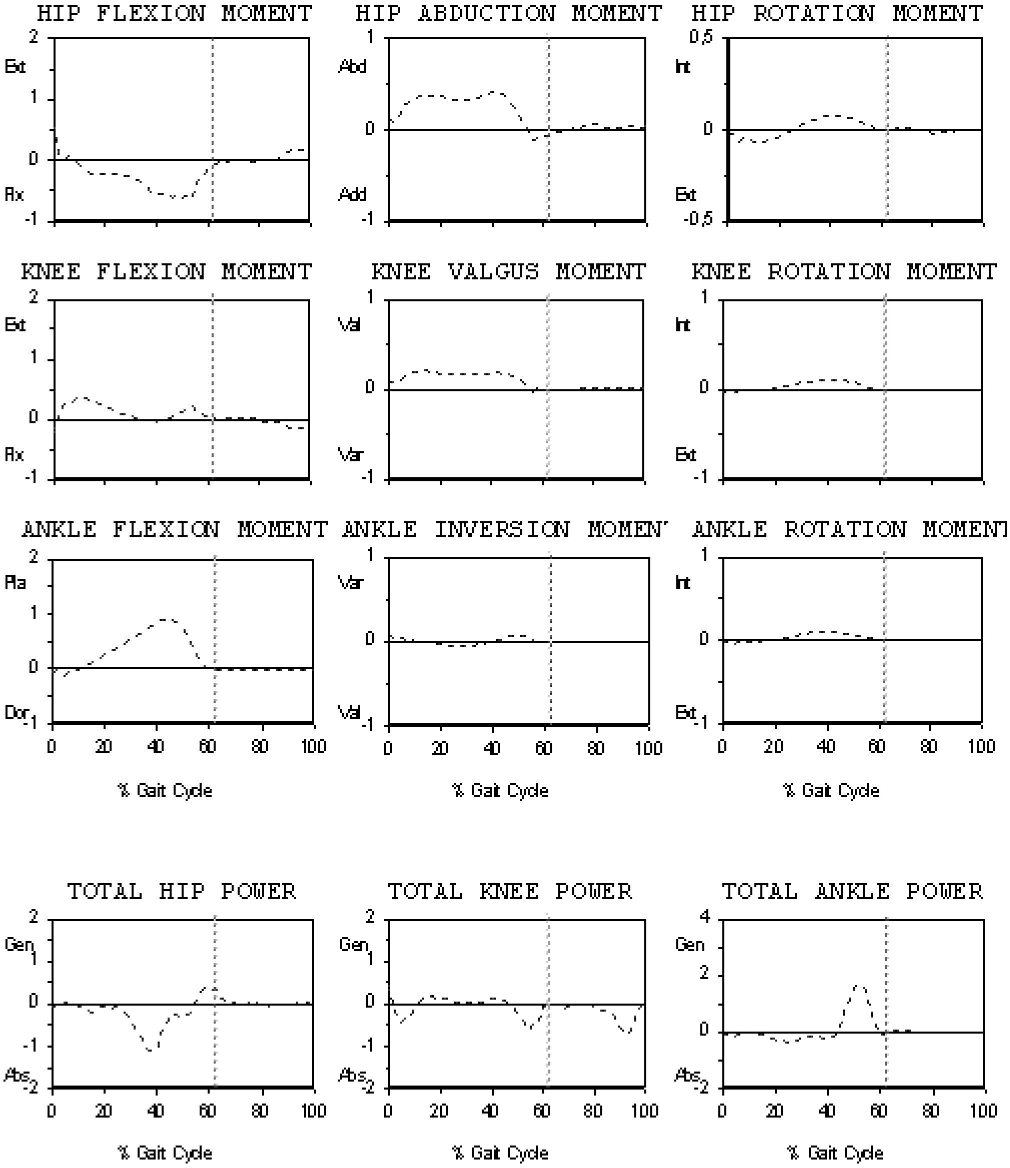
Joint Rotation Angles



Cadence (steps/min)	117,00	115,32	0,00
Walking Speed (m/s)	0,96	0,94	0,00
Stride Time (s)	1,02	1,04	#DIV/0!
Step Time (s)	0,50	0,51	#DIV/0!
Single Support (s)	0,40	0,41	#DIV/0!
Double Support (s)	0,24	0,24	#DIV/0!
Stride Length (m)	0,98	0,97	0,00
Step Length (m)	0,48	0,49	0,00

VICON Clinical Gait Analysis Report

Joint Net Moments and Powers



Diri, Pelin	821	Diri, Pelin	821	0	0
Left	3-Şub-02	Right	3-Şub-02	0	0-Oca-00

VITA

PERSONAL INFORMATION

Surname, Name:	Şen Köktaş, Nigar
Nationality:	Turkish (TC)
Date and Place of Birth:	March 25, 1977, Rize
Marital Status:	Married
Phone:	+ 90 312 210 5552
Fax:	+ 90 312 210 4745
E-mail:	nigar@ceng.metu.edu.tr
	senkoktas@gmail.com

EDUCATION

Degree	**Institution**	**Year of Graduation**
MS	METU Information Systems	2003
BS	METU Mathematics Education	2000

WORK EXPERIENCE

Year	**Place**	**Enrollment**
2006- Present	METU, Department of Computer Engineering	Project Assistant
2000-2006	METU, Informatics Institute	Research Assistant

PUBLICATIONS

1. N. S. Koktas, N. Yalabik, G. Yavuzer, R.P.W. Duin, "**A Decision Tree-MLP Multiclassifier For Grading Knee Osteoarthritis Using Gait Analysis**", (in progress)

2. N. S. Koktas, N. Yalabik, G. Yavuzer, V. Atalay, E. Civek*, "***Combining Decision Trees And Neural Networks For Grading Knee Osteoarthritis***", Proceeding of International Symposium on Health Informatics and Bioinformatics, 2007*

3. N. Sen Koktas, R.P.W. Duin, "**Statistical Analysis of Gait Data Associated with Knee Osteoarthritis**", (in progress)

4. N. Sen Koktas, N. Yalabik, G. Yavuzer, "**Ensemble Classifiers for Medical Diagnosis of Knee Osteoarthritis Using Gait Data**", *Proceeding of IEEE International Conference on Machine Learning and Applications*, 2006

5. N. Sen Koktas, N. Yalabik, G. Yavuzer, "**Combining Neural Networks for Gait Classification**", *Proceeding of Iberoamerican Congress on Pattern Recognition*, 2006

6. N. Sen Koktas, N. Yalabik, "**A Neural Network Classifier for Gait Analysis**", *Proceeding of International Symposium on Health Informatics and Bioinformatics*, 2005

7. N. Sen Koktas, N. Yalabik, "**Self-Examination for an Activity Planning and Progress Following Tool**", Proceeding of the IASTED International Conference on Web-based Education, 2004

8. N. Sen, N. Yalabik, "**An Activity Planning and Progress Following Tool for Self-Directed Distance Learning**", Computer and Information Sciences, 2003

RESEARCH INTERESTS

- Pattern recognition
- Combining classifiers
- Feature selection approaches
- Neural networks
- Gait analysis