FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Screening device for knee osteoarthritis based on Vibroarthrography

Filipa dos Santos Castro Pereira

DISSERTATION

Mestrado Integrado em Bioengenharia - Ramo de Engenharia Biomédica

Supervisor at FEUP: Aníbal Ferreira, PhD

Supervisors at TUDelft: Armagan Albayrak, PhD and Matthijs Netten, MSc

July 27, 2018

Screening device for knee osteoarthritis based on Vibroarthrography

Filipa dos Santos Castro Pereira

Mestrado Integrado em Bioengenharia - Ramo de Engenharia Biomédica

Faculdade de Engenharia da Universidade do Porto

July 27, 2018

Abstract

The knee plays an essential role in the musculoskeletal system, being one of the strongest and most important joints in the human body. Knee stability and pain-free range of motion are important in maintaining daily function. However, due to the complex and fragile anatomic components of the joint, it is easily injured. Osteoarthritis, characterized by progressive cartilage degeneration, is one of the most common knee joint disorders that essentially affects elderly adults. Currently used techniques by orthopedists for the evaluation of knee joint status include imaging techniques or semi-invasive procedures. These methods can be either complex, expensive, not able to detect small and progressive changes in the knee joint or even carry risks associated with radiation exposure or with the procedure itself. However, their main problem lies on the fact that they are not accessible by General Practitioners, whose patients require assistance when they feel the first symptoms. Thus, there is lack of accurate diagnostic methods at primary healthcare.

Crepitus-based techniques represent an innovative approach, being radiation-free and inexpensive. Namely, Vibroarthrography detects, recurring to an accelerometer, vibratory signals arising from the defected knee joint. These signals bear diagnostic information related to the roughness, softening, breakdown or the state of lubrication of the articular cartilage surfaces. OASense was created to be used as a new diagnostic method based on crepitus detection. This device collects and stores the VAG signals, which are further processed to enable the differentiation between unaffected and OA affected knees. With the purpose of obtaining the most discriminant features and the binomial classification model that can obtain the best performance results, an open-source dataset was firstly used, which is composed of 58 healthy and 31 pathological knee joint signals. Spatiotemporal, statistical and time-frequency features were evaluated with several types of classifiers, such as k-nearest neighbour, support vector machine, neural networks, bayesian and even ensemble of classifiers, in order to find the model that provides the best performance results. The best classification was obtained by four different models with three selected time-frequency features with an overall accuracy of 96.15%. These models were the BPNN, an ensemble of SVM and k-NN, an ensemble of BPNN and Random Forest and an ensemble of BPNN and k-NN.

OASense was then used to collect a dataset of 30 normal and 18 OA signals. The same three selected time-frequency features were computed and tested with the 4 models which obtained the best classification. An accuracy value of 93% and a sensitivity of 96.67% were obtained with the BPNN model. These results show that OASense is a promising non-invasive and low-cost technique to be used as a diagnostic method at primary healthcare. A simple knee joint extension/flexion movement needs to be performed by the patient, while the GP uses the device to record VAG signals. The encouraging results motivate the use of the device to discriminate between several stages of OA or to distinguish between different knee joint pathologies, such as Rheumatoid Arthritis and Osteoarthritis.

Resumo

O joelho assume um papel essencial no sistema musculoesquelético, sendo uma das articulações mais importantes e resistentes do corpo humano. No entanto, esta articulação é facilmente afetada devido ao caráter complexo e frágil dos seus componentes anatómicos. A Osteoartrite (OA), caracterizada pela degeneração progressiva da cartilagem, é uma das doenças mais comuns da articulação do joelho, afetando essencialmente pessoas idosas. A avaliação do estado desta articulação é feita pelos ortopedistas através de técnicas de imagiologia ou de procedimentos semi-invasivos. Estes métodos ou são complexos, caros, não são capazes de detetar alterações pequenas e progressivas ou acarretam riscos associados à exposição a radiação ou ao próprio procedimento. No entanto, o maior problema é o facto de não estarem acessíveis aos médicos de clínica geral, os quais são consultados pelos pacientes quando sentem os primeiros sintomas. Assim, há uma falta de métodos de diagnóstico precisos ao nível de saúde primário.

Métodos baseados em crepitação representam uma abordagem inovativa, sendo acessíveis e livres de radiação. Nomeadamente, a vibroartrografia deteta, através do uso de acelerómetros, sinais vibratórios resultantes da articulação afetada. Estes sinais carregam informação diagnóstica relacionada com a rugosidade, suavidade, degradação ou estado de lubrificação da cartilagem articular. OASense foi criado para ser usado como um novo método de diagnóstico baseado na deteção de crepitação. Este dispositivo recolhe e guarda os sinais VAG, que são posteriormente processados, o que torna possível a diferenciação entre joelhos afetados e não afetados por OA. Tendo como propósito a obtenção das características mais discriminantes e o modelo de classificação binomial que permite alcançar os melhores resultados de performance, um conjunto de dados *open-source* foi, primeiramente, usado, o qual é composto por 58 sinais saudáveis e 31 sinais patológicos. Características espaciotemporais, estatísticas e no domínio tempo-frequência foram avaliadas com diferentes classificadores, tal como *k-nearest neighbour*, *support vector machine*, *neural networks*, *bayesian* e ainda assemblagem de classificadores, de forma a encontrar o modelo que confere os melhores resultados. A melhor classificação foi obtida com quatro modelos diferentes usando três características no domínio tempo-frequência, com uma precisão global de 96.15%. Estes modelos foram BPNN, uma assemblagem de SVM com k-NN, assemblagem de BPNN com Random Forest e assemblagem de BPNN com k-NN.

OASense foi depois usado para a recolha de uma base de dados constituída por 30 sinais normais e 18 sinais com OA. As mesmas três características foram extraídas e testadas com os quatro modelos que obtiveram os melhores resultados. Foi obtido um valor de precisão de 93% e uma sensibilidade de 96.67%, usando o modelo BPNN. Estes resultados comprovam que OASense é um método de diagnóstico promissor para ser usado ao nível de saúde primário, sendo não-invasivo e acessível. Um simples movimento de extensão/flexão do joelho necessita de ser efetuado pelo paciente, enquanto o médico usa o dispositivo para gravar os sinais VAG. Os resultados bastante positivos motivam, assim, o uso do dispositivo para discriminar entre diferentes estados de severidade de OA ou ainda para distinguir entre diferentes patologias do joelho, como a Osteoartrite e a Artrite Reumatóide.

Agradecimentos

A realização desta dissertação contou com importantes apoios e incentivos sem os quais não se teria tornado realidade.

Em primeiro lugar, gostaria de agradecer ao Professor Aníbal Ferreira, o meu orientador da FEUP por todo o apoio fornecido e esclarecimento de qualquer dúvida que pudesse ter, mesmo que à distância. Aos meus orientadores da TUDelft e Zimmer Biomet, Armargan Albayrak e Matthijs Netten, pela sugestão de um projeto entusiasmante como este e por toda a supervisão e sugestões ao longo do semestre.

Obrigada a Hilbrand Bodewes pelo incentivo por parte da Zimmer Biomet e ao Dr. Hennie Verburg, Nina Matthijsen e Nicole de Esch por parte do Hospital Reinier de Graaf por toda a ajuda e dedicação primeiramente para validação do estudo, seguida do processo de seleção dos pacientes e finalizando com a aquisição dos dados. Aos estudantes Diego Renting, Nathalja Wiersma, Charlotte Kemp, Laura Heikamp, Sharon Kwant e Mucahit Aydin pela ajuda na aquisição dos dados em pacientes que não falavam inglês. E, acima de tudo, agradeço a participação de todos os voluntários (pacientes e saudáveis) no estudo.

Por último, mas não menos importante, obrigada a todos os meus amigos que, em Portugal ou noutras partes do Mundo, nunca me deixaram de apoiar nos momentos em que mais me apetecia desistir e ainda me proporcionaram todos os momentos de descontração tão necessários. Em especial, obrigada ao Erotik e à Bella que, apesar de longe, sempre me apoiaram. Obrigada à minha colega de casa e grande amiga Tintol por todos estes 5 meses e obrigada aos meus amigos Silvana, Iolanda, Javier, Erica e Sérgio pelos tão bons momentos em Delft.

Obrigada aos meus pais, Rui e Susana, e à minha irmã Marta, por me proporcionarem mais uma incrível experiência internacional e por sempre estarem lá para me ouvir, nos bons e nos maus momentos.

Filipa Pereira

"The real voyage of discovery consists not in seeking new lands but seeing with new eyes."

Marcel Proust

Contents

List of Figures

List of Tables

List of Abbreviations

ACD	Advanced Concept Design
ACL	Anterior Cruciate Ligament
ACLT	Anterior Cruciate Ligament Tear
ADC	Analog to Digital Converter
AE	Acoustic Emission
AED	Advanced Embodiment Design
ANN	Artifical Neural Network
ApEn	Approximate Entropy
AUC	Area Under the ROC Curve
BMI	Body Mass Index
BN	Bayesian Network
BP	Back Propagation
CMA	Cascade Moving Average
COMP	Cartilage Oligomeric Matrix Protein
CT	Computed Tomography
CTX-II	C-Telopeptide regions of type II collagen
DAG	Directly Acyclic Graph
DB2	Daubechies 2 wavelet
DFA	Detrended Fluctuation Analysis
dGEMRIC	delayed Gadolinium-Enhanced Magnetic Resonance Imaging of Cartilage
DTW	Dynamic Time Wrapping
EAG	Electroarthrography
EEMD	Ensemble Empirical Mode Decomposition
EP	Energy Parameter
ESP	Energy Spread Parameter
FD	Fractal Dimension
FF	Form Factor
FFT	Fast Fourier Transform
FLDA	Fisher's Linear Discriminant Analysis
FP	Frequency Parameter
FPR	False Positive Rate
FSP	Frequency Spread Parameter
FT	Fourier Transform
GAG	Glycosaminoglycans
GP	General Practicioner
H	Entropy
IMF	Intrinsic Mode Function
JAAS	Joint Acoustic Analysis System

JSW	Joint Space Width
KL	Kellgren-Lawrence
k-NN	k Nearest Neighbors
KU	Kurtosis
LiPo	Lithium Polymer
LMS	Least Mean Squared
MP	Matching Pursuit
MRI	Magnetic Resonance Imaging
NB	Naive Bayes
OA	Osteoarthritis
PC	Principal Component
PCA	Principal Component Analysis
PCB	Printed Circuit Broad
PDF	Probability Density Function
PET	Positron Emission Tomography
Phono-A	Phonoarthrography
QMF	Quadrature Mirror Filter
qMRI	quantitative Magnetic Resonance Imaging
RBFN	Radial Basis Function Network
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
RQA	Recurrence Quantification Analysis
SampEn	Sample Entropy
SBS	Sequential Backward Selection
SFS	Sequential Forward Selection
SK	Skewness
SPECT	Single-Photon Emission Computed Tomography
STFT	Short Time Fourier Transform
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TC	Turns Count
TFD	Time Frequency Distribution
TKR	Total Knee Replacement
TPR	True Positive Rate
TUDelf	Delft University of Technology
US	Ultrasound
UTE-T2*	Ultrashort echo-Time Enhanced T2*
VAG	Vibroarthrography
VMS	Variance of the mean-squared
WPD	Wavelet Packet Decomposition
WPT	Wavelet Packet Transform
WT	Wavelet Transform
WVD	Wigner-Ville Distribution

Chapter 1

Introduction

The knee joint is the largest and most complex joint in the human body. It supports nearly the whole weight and its stability depends almost entirely upon its associated ligaments and muscles. Therefore, it is most vulnerable both to injury and the development of Osteoarthritis (OA) (Wu, 2015).

Being one of the most frequent injured sites, this joint is particularly interesting for research because without its well-functioning, human motion could be compromised along with the capability to perform simple daily tasks (Wu, 2015).

Osteoarthritis is one of the most common diagnosed diseases in clinical practice resulting from the progressive degeneration of joint constituents including cartilage and subchondral bone (Kim et al., 2005). It is the most common joint disease, affecting nearly 10% of the population worldwide (Cooper, 2013). Specifically, in the USA, knee Osteoarthritis accounted for more than 80% of OA's total burden in 2017, affecting 19% of adults aged 45 years and older (Wallace et al., 2017). These numbers are projected to increase substantially in the near future due largely to an aging population and an ever-increasing prevalence of obesity (Johnson and Hunter, 2014).

Knee OA is the most dominant agent of functional incapacity, resulting in more difficulties with climbing stairs and walking than any other disease. Patients with OA experience a 50% increase in all-cause mortality rates (death from any cause) (Alekseeva and Nasonov, 2013; Nüesch et al., 2011).

1.1 Motivation

OA diagnosis starts at the primary healthcare level, when patients consult General Practicioners (GPs) because of symptoms, joint pain being the most common one (Martel-Pelletier et al., 2016). GPs listen to patient's complaints and perform physical examination, such as palpation, to analyse the joint. In case the GP suspects of some joint disease, the patient is referred to the orthopedist.

Orthopedists represent the secondary healthcare level, who have access to more accurate diagnostic methods. Image-based non-invasive techniques can be used, such as Radiography, which

is the current gold standard because of its wide availability and low-cost profile, Magnetic Resonance Imaging (MRI), Ultrasound (US) and Computed Tomography (CT) (Guermazi et al., 2013). Arthroscopy is the most used non-imagiologic diagnostic tool for the evaluation of cartilage condition. However, such technique is not suitable for all patients due to its invasive nature, required anesthesia and surgery related risks (Madry et al., 2016).

However, the main problem is indeed at the primary level, due to the lack of an accurate diagnostic technique, which may lead to a first misdiagnosis. Therefore, a new low-cost and easily accessible diagnostic method to be used by GPs is needed.

Some techniques are arising, where quantitative MRI (qMRI), Biomarkers and Electroartrography (EAG) are some examples (Guermazi et al., 2013; Zhu et al., 2017). However, more studies are needed to confirm their accuracy in detecting cartilage degeneration over time. Also, qMRI is not low-cost and biomarkers are difficult to validate.

Based on the detection of crepitus (the sound produced by a joint), a new diagnostic technique emerged, named Phonoarthrography (Phono-A) (Bassiouni et al., 1995, 2011). This concept was followed by the development of two other methods, based also on the detection of crepitus, which are Vibroarthrography (VAG) (Mascaro et al., 2009) and Acoustic Emission (AE) (Shark et al., 2011).

These sound-based techniques have great potential in screening knee OA and require low-cost equipment. Thus, the gap at the primary care level may have its solution on these methods (Bassiouni et al., 1995; Mascaro et al., 2009; Shark et al., 2011).

1.2 Goals

Knowing the potential of diagnostic techniques based on crepitus detection, the primary goal of this dissertation is to create a powerful algorithm that is able to classify VAG signals according to the status of the knee joint (affected/unaffected by OA). This requires the understanding of the most adequate pre-processing techniques, the identification of the most relevant and differenciating signal features to be extracted as well as the model that returns the best performance results in this two-class problem.

The ultimate goal of this dissertation is to create a new, accurate and non-invasive knee joint screening test based on crepitus detection in order to be used at the primary care level. Additionally, it could be used for monitoring during OA progression or even during recovery after surgery.

This test is composed of both the previously described algorithm and a handheld device that enables the acquisition and storage of trustful data from patients. The implementation of this device requires understanding where it should be positioned on the knee surface, as well as the most appropriate movement to be executed in order to accurately detect crepitus.

1.3 Document structure

Besides the introduction, which explains the motivation and the main goals of the proposed work, this dissertation includes six more chapters. Chapter 2 places the project, reviewing the knee anatomy and biomechanics, the main fundaments of Osteoarthritis and the diagnostic tools currently used for OA management as well as the emerging ones, with an emphasis on the crepitus-based methods.

Chapter 3 includes the history of the OASense device, focusing on the initial concept and the several developments, regarding both the electronics and the classification algorithm. The main problems are identified and the new concept developed within this dissertation is explained. In Chapter 4, the literature is reviewed, which includes all the steps needed in order to perform VAG signal analysis. Therefore, the most used techniques in signal acquistion, pre-processing, feature extraction and selection and classification are explained in detail.

Chapter 5 describes the used approach to the analysis of VAG signals belonging to an open-source dataset, consisting of 89 signals. The approach includes all sections mentioned in Chapter 4 that integrate a machine learning classification system. Chapter 6 describes the data collection using OASense and the extraction of the most discriminant features as well as the training of the model that returned the best performance results in Chapter 5.

Finally, Chapter 7 summarizes the main outcomes and conclusions of this dissertation and presents some suggestions for future improvements.

Chapter 2

Contextualization

2.1 Knee anatomy and biomechanics

The knee joint (Figure 2.1) includes three articulating bones, namely the femur, patella and tibia, which originate three functional compartments: the patellofemoral articulation, lateral femorotibial articulation and medial femorotibial articulation. All compartments contribute to the synovial characteristic of the knee joint, as they are bathed in synovial fluid which helps lubricate and cushion the joint without any friction in motion (Wu, 2015).

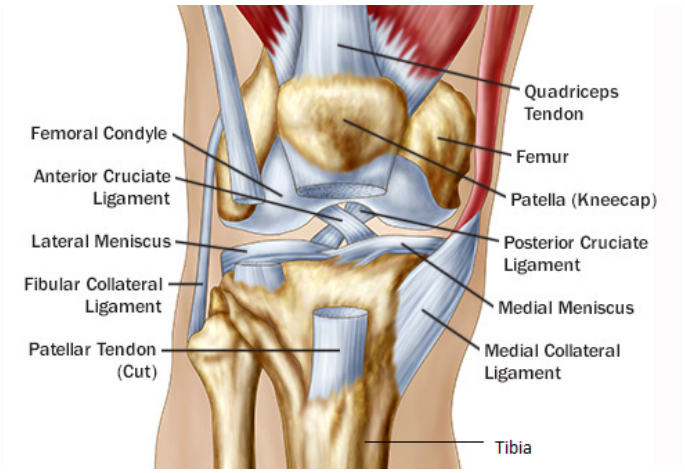


Figure 2.1: Representation of the knee joint and its main components (WebMD, 2018).

The first compartment is a synovial gliding joint, connecting the patella posteriorly with the two femoral condyles of the anterior side of the femur, being able to create a frictionless transfer over the knee of the forces generated by contraction of the quadriceps femoris muscle (Moreira, 2015; Abulhasan and Grey, 2017; Wu, 2015). The two femorotibial compartments are synovial hinge joints formed by the pair of condyles from the distal end of the femur and the proximal end of the tibia, bearing most of the body weight (Orthopod, 2016; Abulhasan and Grey, 2017; Wu, 2015).

With this configuration, the knee permits a typical rolling-gliding mechanism under a great spectrum of loading conditions, offering a six degree of freedom range of motion (Figure 2.2). The knee allows up to 160º of flexion from up to 5º of hyperextension. In addition to that, it permits 25-30º of rotation, 6-8º of valgus and varus in extension, 5-10mm of anteroposterior translation and 1-2mm of mediolateral translation. The articular cartilage and the menisci also allow 2-5mm of joint compression (Goldblatt and Richmond, 2003; Shenoy et al., 2013).

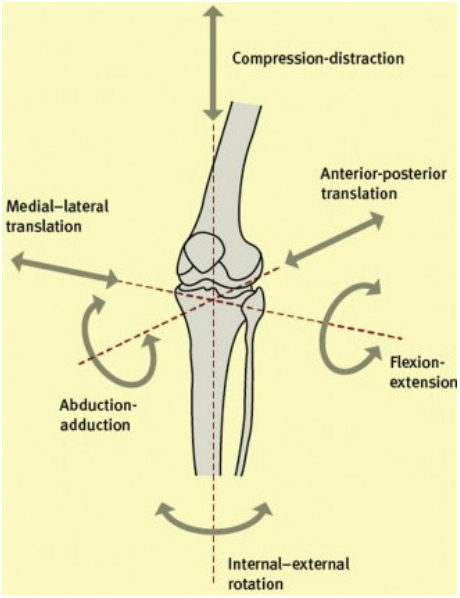


Figure 2.2: Representation of the six degree of freedom range of motion in the knee (Shenoy et al., 2013).

Two different types of cartilage can be found in the knee joint: the articular cartilage which covers the articular surfaces, namely, the femoral condyles, the flattened plateaus of the tibia and the facets of the patella; and the menisci, two fibrocartilaginous structures located between the first two articular surfaces (Krishnan, 1999; Moreira, 2015). These C-shaped articulated disks act as shock absorbers for the body load and dynamic movements, providing structural integrity to the knee when it undergoes tension and torsion as well as dispersing the load and reducing friction over the articular surfaces (Orthopod, 2016; Abulhasan and Grey, 2017; Wu, 2015).

The knee joint is stabilised by both primary and secondary stabilisers. Knee ligaments constitute the primary stabilisers, while muscles around the knee play the secondary role (Abulhasan and Grey, 2017; Orthopod, 2016).

Ligaments are fibrous bands of tissue that connect bone to bone. The knee set of ligaments include the medial collateral, lateral collateral and the anterior and posterior cruciate ligaments. While collateral prevent the knee from great displacements in the medial-lateral direction, cruciate ligaments control the excessive anterior-posterior and varus-valgus motion of the tibia in relation to the femur (Orthopod, 2016; Abulhasan and Grey, 2017). There are additional small ligaments that surround the knee and aid in maintaining overall knee stability. The anterior cruciate ligament

(ACL) is considered the main stabiliser of the knee, contributing to about 85% of the joint stabilisation and enabling smooth and stable flexion and rotation of the knee (Abulhasan and Grey, 2017).

The secondary stabilisers of the knee joint are all the surrounding muscles, whose primary function is to produce motion for all the six degrees of freedom (Abulhasan and Grey, 2017). The anterior aspect of the knee consists predominantly of the quadriceps muscles, which straighten the knee, whereas the hamstring group of muscles, located in the posterior part, function as knee flexor. The medial and lateral musculature of the knee both aid in the joint flexion (Moreira, 2015; Orthopod, 2016; Abulhasan and Grey, 2017).

Contrary to ligaments, which connect bones to bones, tendons attach muscles to bones and also provide stabilization to the joint. There is one main tendon in the knee joint, the patellar quadriceps tendon, which can be divided in two parts. The first part is the largest and connects the patella to the tibia, while the second establishes the connection of the quadriceps muscle to the patella (Abulhasan and Grey, 2017).

2.2 Osteoarthritis

2.2.1 Overview

OA is a degenerative joint disease, characterized by the local and progressive loss of articular cartilage along with simultaneous changes in the bone underneath the cartilage (Hochberg, 2013; Martel-Pelletier et al., 2016). It causes joint pain, tenderness, crepitus, movement limitation and variable degrees of local inflammation (Woolf and Pfleger, 2003; Felson et al., 2000).

2.2.2 Pathology

Under normal physiological conditions, a dynamic remodelling process of the cartilage occurs, in which there is a balance on the levels of degradative and synthetic enzyme activities, resulting in a maintenance of the cartilage's volume (Lohmander et al., 1992).

In OA cartilage, matrix degrading enzymes are overexpressed, which leads to loss of collagen and proteoglycan, components that provide tensile strength and compressive resilience, respectively (Martel-Pelletier et al., 2016; Lohmander et al., 1992). In reaction to this loss, chondrocytes, the main cellular component, proliferate and synthesize increased quantities of proteoglycan and collagen at a first stage. However, with the progression of the disease, these reparative efforts are surpassed by progressive cartilage degradation (Lohmander et al., 1992).

In the initial stages of OA, the superficial cartilage layer is affected by phenomena of fibrillation, erosion and cracking. With the development of the disease, they start to affect deeper layers, resulting in clinically observable erosions (Lohmander et al., 1992).

Alterations on the articular cartilage are accompanied by damages to the bone, due to alterations in bone remodelling mechanisms. These changes are characterized by a progressive increase in the subchondral bone plate thickness, a modification in the architecture of subchondral

trabecular bone and formation of new bone at the joint margins (osteophytes) (Martel-Pelletier et al., 2016). In addition, synovitis (inflammation of the synovium) is a common feature of OA. It involves thickening of the synovial lining cell layer and increased vascularity and inflammatory cell infiltration of the synovial membranes (Martel-Pelletier et al., 2016).

The most important alterations are depicted in Figure 2.3.

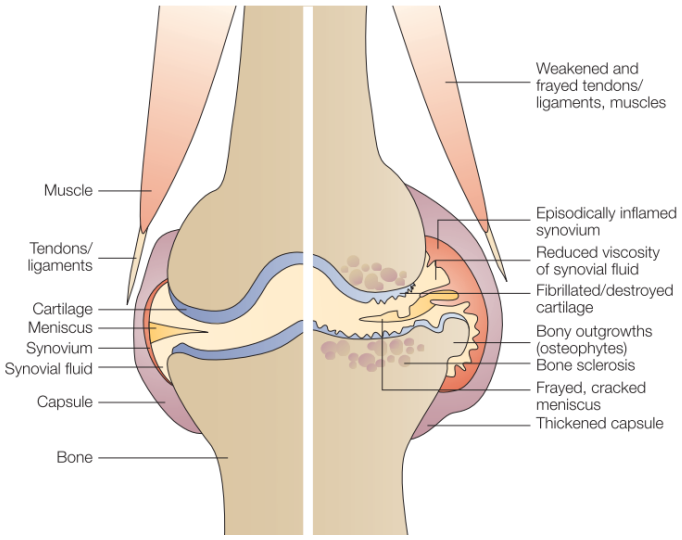


Figure 2.3: Schematic comparison between a healthy and an OA knee joint. The healthy joint (left) has normal cartilage without any fissures and no signs of synovial inflammation. In the OA joint (right), it is possible to see the fibrillation of the cartilage as well as remodelling of bone (Wieland et al., 2005).

2.2.3 Risk factors

Age is the main risk factor of OA, due to numerous reasons, such as oxidative damage, thinning of cartilage and muscle weakening. Fundamental cellular mechanics, which control tissue homeostasis, drop with age, resulting in a defective response to stress or joint injury and resultant joint tissue destruction and loss (Litwic et al., 2013). Along with age, obesity is one of the strongest risk factors of OA, particularly for knee OA. Excess weight increases the load on the joint and causes lesions in bone and muscle tissue (Alekseeva and Nasonov, 2013). Also, muscle weakness, namely quadriceps weakness, leads to less stabilization of the knee during physical activity (Cooper, 2013).

Some traumatic knee injuries can cause rapid joint degeneration, such as intra-articular fracture, dislocations and anterior cruciate ligament tear (ACLT). These injuries can provoke joint instability, articular cartilage damage, among others, which increase the risk of developing OA (Chu et al., 2012). Occupations that involve repetitive knee bending and squatting also increase the risk of knee OA (CDC, 2018).

2.2.4 Management

Currently, no immediate cure is available since the osteoarthritic process is irreversible. It is only possible to alleviate symptoms and postpone the almost inevitable total knee replacement (TKR). There are several options for this, starting on the safest and least invasive nonpharmacologic therapy which is followed by the pharmacologic and alternative options (Sinusas, 2012).

Nonpharmacologic therapy often starts with exercise (Sinusas, 2012). Physical activity including aerobic, resistance land-based or aquatic exercises which provoke local muscle strengthening, is recommended for OA patients. Orthotic devices such as special footwear, insoles, knee bracing, and canes are also an option (Roubille et al., 2013).

Regarding pharmacologic treatment, there are some drugs available, which differ mainly in their efficacy and adverse effects (Roubille et al., 2013). Acupuncture and transcutaneous electrical stimulation have been reported to show some short-term pain relief efficacy (Sinusas, 2012).

Ultimately, when pain and loss of function continue after the patient is subjected to the previous therapies, surgery is the solution. TKR is the most effective intervention since the prosthesis lasts for 10-15 years (Sinusas, 2012; Roubille et al., 2013). OA accounts for more than 90% of all TKR procedures (Lohmander, 2013).

2.3 Diagnostic tools

2.3.1 Primary healthcare level

At the primary healthcare level, symptoms reported by patients and signs noted during physical examination guide GPs in identifying patients affected by knee OA. Pain is one of the most reported symptoms, together with redness and swelling. However, they are non-specific symptoms, related to pathology in several structures within the knee joint, which include synovitis, subchondral bone marrow lesions and joint effusion. Further, they cannot be directly measured or assessed during physical examination (Crema et al., 2011).

2.3.2 Secondary healthcare level

At the secondary level, orthopedists have more accurate diagnostic techniques available. The most relevant are explained below.

2.3.2.1 Radiography

Radiography, commonly known as X-Ray, is the simplest and least expensive imaging technique, being the gold standard to detect and monitor OA progression (Kumar et al., 2011). Figure 2.4 shows a radiography of the knee joint.

It is possible to detect OA-associated bony features and calculate the joint space width (JSW), manually or semi-automatically using computer software, which is an indirect representation of cartilage thickness and meniscal integrity. Thus, the narrowing of JSW is the criterion used for

Figure 2.4: X-ray of the knee joint in (A) anteroposterior and (B) lateral views showing (1) joint space narrowing and (2) osteophyte formation (Sinusas, 2012).

the assessment of OA progression and the complete loss of JSW is a strong indicator for joint replacement (Guermazi et al., 2013).

However, a precise measurement of the articular structures mentioned above and the detection of other pathologic features associated with OA is not possible by means of radiography, which constitute major limitations of this technique. Moreover, this modality is limited to detect OA at later stages only and it uses X-ray, which is an ionizing radiation (Kumar et al., 2011).

2.3.2.2 Magnetic Resonance Imaging

MRI, which uses magnetic field and radio frequency pulses (Kumar et al., 2011), has a tomographic viewing perspective, providing cross-sectional images of the knee's anatomy. Contrary to radiography, it is able to directly represent all the components of the joint and their pathologies, including the articular cartilage, menisci, ligaments, among others. Thus, the joint can be evaluated as a whole organ, providing more details of the changes associated with OA. MRI can detect complications of the disease at a much earlier stage than radiography (Guermazi et al., 2013).

However, MRI requires expensive instrumentation, time and facilities and is contraindicated in people who have implanted devices, such as pacemakers (Ahmed et al., 2015).

2.3.2.3 Ultrasound

US is a technique that enables multiplanar and real-time imaging using sound waves. It is able to image soft tissue and to detect synovial pathology, contrary to radiography. It is relatively low cost and allows the observation of cartilage integrity, inflammation, irregularities and thickness (Kumar et al., 2011).

However, US usage is still limited because of images' poor resolution and the operator-dependency. Further, it can not be applied to deeper articular structures or to the subchondral bone, due to physical properties of the sound (Guermazi et al., 2013).

2.3.2.4 Computed Tomography

CT is an important technique for the characterization of OA, since it produces cross-sectional images of a specific area, combining several computer-processed radiographies. Both cortical bone and soft tissue are better represented than on MRI. However, as well as radiography, it delivers high doses of radiation and provides low soft-tissue contrast (Guermazi et al., 2013).

2.3.2.5 PET and SPECT

All of the above imaging techniques measure structural changes that are a result of the disease process. Thus, there is a need to develop methods that can image the activity of the disease process itself. Functional nuclear imaging techniques, such as Positron Emission Tomography (PET) and Single-Photon Emission Computer Tomography (SPECT), have been reported for monitoring the inflammatory process of OA (Kandahari et al., 2015).

PET is able to detect radioactivity emitted after a small amount of radioactive tracer is injected into a peripheral vein reflecting, for example, the blood flow or metabolism of target organs (Berger, 2003). In OA's case, it can detect, for instance, the glucose metabolism of chondrocytes, which maintains the cartilage matrix (Nakamura et al., 2007). SPECT's tracer stays in the blood stream rather than absorbed by surrounding tissues, offering a cheaper and more readily available solution than higher resolution PET scans (MayfieldClinic, 2016).

Although these techniques can provide valuable information for the diagnosis of OA, they are expensive and not widely available. More tests are needed to prove their efficacy (Berger, 2003; MayfieldClinic, 2016).

2.3.2.6 Arthroscopy

Contrary to the previous imaging techniques, arthroscopy is an invasive diagnostic method, which constitutes a major disadvantage. However, the use of needle arthroscopes permits direct visualization and palpation of intra-articular structures (Balint and Szebenyi, 1996). Thus, articular cartilage lesions can be easily seen and palpated by this tool (Chu et al., 2012).

2.3.2.7 CT arthrography

The needle arthroscopes can be combined with CT, giving rise to a diagnostic technique called CT arthrography. This method permits an indirect visualization of cartilage and other intrinsic joint structures, especially in the knee joint (Guermazi et al., 2013). It is used mainly to assess cartilage surface but does not provide soft tissue information. This technique continues invasive and is insensitive to changes of the deep layers of cartilage without surface alterations (Braun and Gold, 2012).

2.3.3 Emerging

Other diagnostic methods are emerging, which try to solve some drawbacks of the previous techniques.

2.3.3.1 Quantitative MRI

qMRI is able to detect biochemical, spatial and relational changes of the primary components of articular cartilage (menisci, water, type II collagen and proteoglycans), through quantifying MRI signal changes. The already studied qMRI techniques include delayed gadolinium-enhanced magnetic resonance imaging of cartilage (dGEMRIC), T2 and T1rho mapping and ultrashort echo-time enhanced T2* (UTE-T2*) mapping (Chu et al., 2012).

dGEMRIC is a technique used to map the relative concentration and spatial distribution of cartilage proteoglycan content. OA is characterized by a loss in tissue glycosaminoglycan, reflected by the decrease in the dGEMRIC index (Tiderius et al., 2003). Figure 2.5 represents a dGEMRIC scan of the knee.

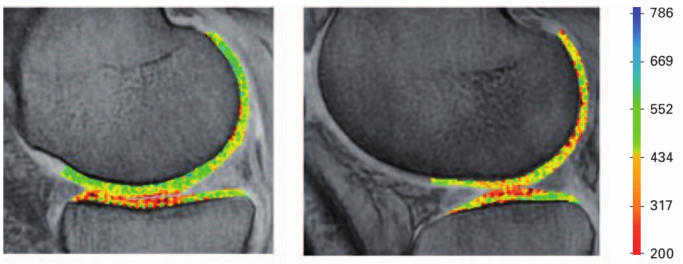


Figure 2.5: dGEMRIC scan of the knee joint showing slices from the medial (left) and lateral (right) compartments. The colour scale represents the dGEMRIC index, where higher values represent increased glucosaminoglycan (GAG) content. The scan demonstrates the physiological reduction in concentration of GAG as one moves from the deep to superficial cartilage zones (Pollard et al., 2008).

T2 is sensitive to tissue hydration and collagen matrix organization, whose values increase with cartilage degeneration. Thus, high levels of T2 have been detected in subjects at risk for developing OA (Menezes et al., 2004).

Similarly to T2, T1rho is sensitive to tissue hydration and matrix macromolecular architecture. It is thought to relate also with the proteoglycan content of the tissue (Menezes et al., 2004). T1rho values tend to be higher in OA patients and this technique is reported to be more sensitive to cartilage degeneration than T2, although further work is needed (Li et al., 2007).

UTE-T2* has the potential to measure short-T2 MRI signals from the meniscus and deep layers of articular cartilage which are not well captured by conventional MRI or by T2 mapping. It reflects the integrity of collagen matrix (Williams et al., 2010).

As a new and innovative area, more studies are needed in order to answer some questions. For instance, the reliability of dGEMRIC use in deep cartilage remains unclear and more work

is required in both T2 and T1rho to demonstrate their specificity for molecular changes to the proteoglycan and/or collagen components of articular cartilage (Chu et al., 2012).

2.3.3.2 Biomarkers

A biomarker is a "characteristic that is objectively measured and evaluated as an indicator of normal biological process, pathogenic processes, or pharmacological responses to a therapeutic intervention" (Glyn-Jones et al., 2015).

Many biomarkers have been investigated. As OA affects the metabolism of bone, cartilage and synovium, potential candidates include the matrix components, their propeptides which are cleaved off during the conversion to the functioning protein, their breakdown products, cytokines and proteases of these tissues. Other less studied candidates are synovial, inflammatory and genetic markers (Pollard et al., 2008).

There are a number of requirements for the validation of a biomarker. For instance, its biology must be understood in terms of its origin, tissue specificity, spatial tissue distribution, metabolism and release from the tissue. Its appraisal should be robust, reproducible, readily available, convenient and cheap (Pollard et al., 2008).

CTX-II (C-telopeptide regions of type II collagen) has been extensively studied, which correlates well with the burden of osteophytes (Patra and Sandell, 2011), being able to predict OA severity and progression. However, it is still unclear whether it is the marker of choice for OA, since its tissue origins remain uncertain (Pollard et al., 2008).

Another example is COMP (cartilage oligomeric matrix protein). Elevated levels of COMP have been proved to occur in OA of the knee and are associated with reduced cartilage volume and disease progression. However, it lacks specificity since it is also present in other connective tissues, such as the synovium and meniscus, making it possible to also reflect synovitis, for example (Pollard et al., 2008).

Sources for biomarkers are blood and other fluids, like plasma or urine. However, secreted molecules are often diluted and may become undetectable and systemic biomarkers may be confounded with other physiological processes originating from other tissues. These problems, together with the difficulty for validation (Pollard et al., 2008), in addition to their no linear correlation with structural changes and not well defined distinction from age-related changes, represent the main limitations of biomarkers' use as a diagnostic tool (Madry et al., 2016; Pollard et al., 2008).

Despite the high research efforts, no single OA biomarker stands out as a gold standard and has been sufficiently validated for a systematic use, with only a few of them presenting extensive research and yet limitations, like the previously stated CTX-II and COMP (Madry et al., 2016).

2.3.3.3 Electroarthrography

Knee EAG consists on "recording electrical potentials using electrodes applied on the surface of the knee during mechanical loading of the knee joint". The so called "streaming potentials",

that can be recorded directly on the cartilage surface, are produced by the displacement of ions against the fixed charge density within the cartilage, which is induced by the pressure gradient during compression. The streaming potentials are sensitive to cartilage degeneration, leading to a decrease in EAG potentials in OA (Zhu et al., 2017; Préville et al., 2013).

Préville et al. (2013) proved that the potentials recorded from non-OA knees were higher compared to the OA group. However, further studies are needed to confirm the possible use of EAG for diagnosis of OA, as this is a non-invasive technique and requires low-cost instrumentation (Préville et al., 2013).

2.3.4 Crepitus based

Crepitus is defined as a "crackling or grinding sound on joint movement with a sensation on the joint", according to Bassiouni et al. (2011). However, the first step on its detection started with Heuter (1889) study which envolved the evaluation of joint sounds, as he described hearing and localizing loose bodies within the knee, using a stethoscope. More research was done on the auscultation of the knee (Blodgett, 1902; Ludloff, 1906; Walters, 1929). Later, Erb (1933), replaced the stethoscope by a microphone to reduce subjectivity.

2.3.4.1 Phonoarthrography

In 1986 (Bassiouni, 1986), a detailed report was published on phonoarthrography, which is described as the recording of knee sounds in a fixed lapse of time by a microphone in the audible frequency range (20Hz to 20kHz) located on the centre of the patella during both flexion and extension of the knee (Bassiouni, 1986; Mascaro et al., 2009).

In this study, a microphone, an amplifier, a recorder tape and a memory oscilloscope were used. However, this device had some disadvantages, such as tape noise imposed on the knee signal, limited memory storage of the oscilloscope, which did not allow storage of the full motion of the knee, and finally incapacity of auto-analysis of the obtained signals. Later, the same authors developed another device, which consists mainly of a computerised central unit that converts for analysis the captured sound wave by means of a microphone (Bassiouni et al., 1995).

Figure 2.6 represents the acquired phonoarthrographic signals. It can be concluded that the signals correlate well in amplitude with the state of the disease, with the severe case of OA presenting the higher amplitude value.

Being an innovative diagnostic tool, Phono-A was compared with standard techniques, namely radiography and ultrasound, in order to find out its potential. Two studies by Bassiouni et al. (1995, 2011) tested phonoarthrography against radiography. For each phonoarthrographed knee, they had a radiological Kellgren-Lawrence (KL) classification grade. Both studies found that the average amplitude of the Phono-A signal rose steadily with higher radiological grades. Once again, phonoarthrography and US were performed for the same knees (Bassiouni et al., 2011). Using US, the cartilage thickness from the four condolytes was measured. Phonoarthrography inversely correlated with the thickness of cartilage, as expected.

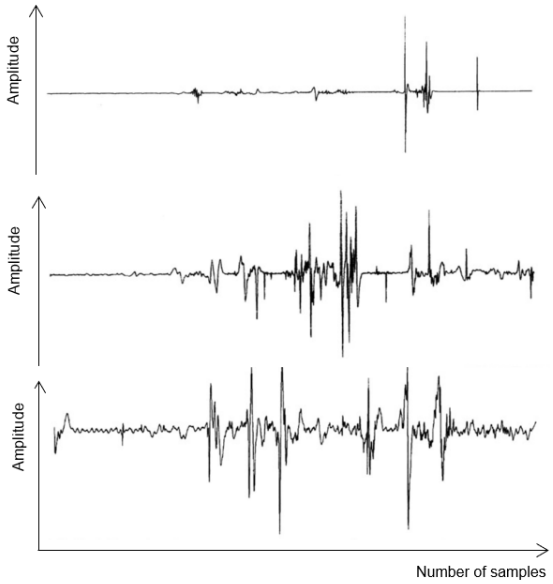


Figure 2.6: Phonoarthrographic signals of a control knee, a moderate case of OA and a severe case of OA (from top to bottom), with an average amplitude of 27.55, 47.53 and 64.45 units/recording, respectively (Bassiouni et al., 1995).

2.3.4.2 Vibroarthrography

The concept of phonoarthrography was followed by the development of vibroarthrography, which uses miniature accelerometers to record crepitus in a lower frequency range, below 1kHz (Mascaro et al., 2009).

VAG signals were recorded during standing up and sitting down movements (Figure 2.7) performed based on timing instructions by an examiner, since leg movement might contribute with noise to the VAG signals. A three axis accelerometer and a potentiometer were used to record angular changes of the knee joint (Tanaka and Hoshiyama, 2012).

Figure 2.7 shows that tiny signals are visible on the individual axis' traces in healthy young subjects, while knees in the senior and patient groups showed clearly visible signals in the traces, being greater in the knees of OA patients.

2.3.4.3 Acoustic emission

AE is a natural phenomenon of high frequency sound that can be generated by structure under loading or surface interaction (Shark et al., 2011). Piezo-electric contact sensors are used to detect these sound waves with frequencies in the ultrasonic domain (20kHz-200kHz) (Mascaro et al., 2009; Shark et al., 2011). For instance, a knee affected by OA with poorly lubricated cartilage surface produces these acoustic signals (Shark et al., 2011).

Mascaro et al. (2009); Shark et al. (2011) developed a knee acoustic emission system. In both studies, the measurement of knee AE was performed using the Joint Acoustic Analysis System

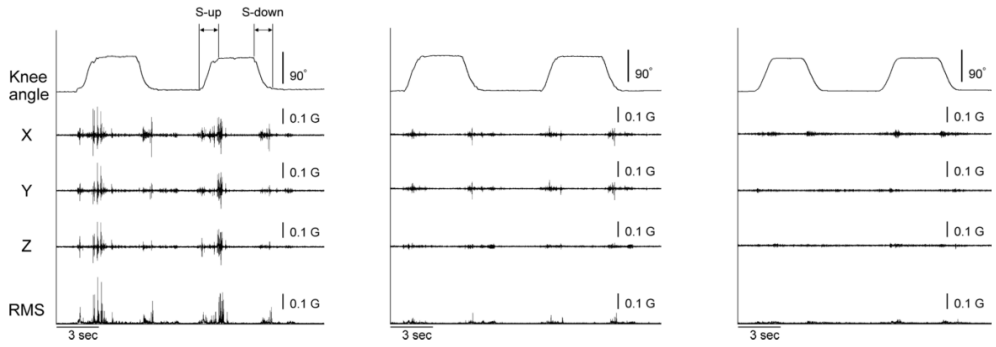


Figure 2.7: Angular changes and VAG signals obtained from a representative knee in the (a) OA diagnosed, (b) healthy elderly and (c) healthy young groups. The upper traces indicate angular change of the knee joint, where the arrows indicate the duration of standing-up (S-up) and sitting-down (S-down) movements. Lower traces (X-Y-Z) represent processed VAG signals for each of the axes of the accelerometer. The bottom trace indicates the values of the root mean square (RMS) in the three axis of the VAG signals (Tanaka and Hoshiyama, 2012).

(JAAS) developed by the authors, combining a traditional AE system with an electronic angle measurement system to provide joint angle based AE data.

The traditional AE system consists of two piezoelectric sensors to attach to each knee joint, two amplifiers, an AE acquisition board and a computer for measurement control and signal processing. The electronic angle measurement system consists of two electrogoniometers and an amplification unit.

In both studies, participants were divided in five groups: early, mild, late adulthood to study joint ageing and middle, late adulthood with OA to understand joint degeneration with OA. They were asked to do repeated sit-stand-sit movements at a fixed speed.

The average, maximum and minimum number of AE hits or events, which refer to the AE signals picked up by the sensor, are seen to increase based on age and knee condition, increasing from the early adulthood to the late adulthood healthy group, followed by the middle to the late adulthood OA group.

2.4 Summary

In addition to being hardly accessible to GPs, the techniques presented in Section 2.3.2 have some other problems, which include high costs, invasiveness, radiation exposure, long acquisition times.

In order to solve these shortcomings, some diagnostic techniques for knee OA are emerging. qMRI seems promising due to its unique detection of biomechanical changes in articular cartilage. However, as well as MRI, it implies high costs. Biomarkers, despite able to monitor some components from the matrix, are still poorly explored and cannot be easily validated. The same happens with electroarthrography, to which more research is needed for the evaluation of its potential as a diagnostic technique.

Crepitus based techniques have been reported to be powerful tools for the screening of knee OA. Therefore, due to their low-cost status, non-invasiveness and no radiation exposure, their use as diagnostic techiques at the primary healthcare level should definitely be explored.

Table 2.1 was built considering all the previous stated diagnostic tools. The clinical usage was evaluated having as reference the gold standard radiography that represents the maximum level (5). The emerging techniques have a 0 in this parameter, since they are only used in research.

Table 2.1: Comparison of the different diagnostic techniques

Technique	Clinical usage (0-5)	Detection	Advantages	Disadvantages
Radiography	5	Cartilage thickness (JSW)	Low cost, Fast, Simple, Easily applicable	Indirect measurements, 2D images, Difficult to detect disease progression, Radiation exposure
MRI	2	Articular cartilage, Menisci and ligaments assessment	Direct measurements, 3D images, No ionizing radiation	High cost, Long acquisition times, Time-consuming analysis
Ultrasound	3	Cartilage integrity, Inflammation, Irregularities and thickness	Low cost, Real-time and multiplanar imaging, Direct measurements, No ionizing radiation	User dependent, Tissue depth dependent on sound's physical properties
CT	2	Articular cartilage assessment	Multiplanar 3D images, Good resolution	High exposure to radiation, Low soft tissue contrast
PET and SPECT	1	Inflammatory process	Continuous monitoring	High cost, Not widely available, More tests needed to prove efficacy
Arthroscopy	2	Direct visualization and palpation of articular cartilage lesions and meniscal and cruciate ligament tears	Low cost, Real-time visualization and palpation	Invasive, Long examination times
CT Arthrography	1	Cartilage surface assessment	Real-time, Multiplanar 3D images, Good resolution	No soft tissue information, Invasive, Insensitive to deep layers changes
qMRI	0	Biomechanical, spatial and relational changes in articular cartilage detection	Molecules concentration continuous monitoring, Deep layers and meniscus assessment	High cost, More tests needed
Biomarkers	0	Matrix components monitoring	Osteophytes analysis, OA severity and progression evaluation	Difficulty in validation, Confusion with other physiological processes, No linear correlation with structural changes
Electroarthrography	0	Cartilage degeneration detection	Low cost instrumentation, Sensitive to changes in the molecular structure of articular cartilage	More tests needed
Phonoarthrography	0	Crepitus detection (sound)	Low cost, Fast, Simple, Sensitive to small alterations, Direct measurements	More tests needed
Vibroarthrography	0	Crepitus detection (vibration)		
Acoustic emission	0	Crepitus detection (high frequency sound)		

Chapter 3

OASense

3.1 Concept

OASense is an innovative device to be used by GPs as a diagnostic tool for knee Osteoarthritis. It is based on the detection of knee sounds (crepitus).

It has been discussed that the gold-standard diagnostic tool for knee OA is radiography, which is requested by an orthopedist for a specific patient. However, when a patient feels the first symptoms of knee OA, he contacts the GP, who doesn't have access to X-ray equipment. This is the reason why OASense has GPs as the target group, so it can be used as a screening tool in primary healthcare (Cherryman, 2006).

With this clear market need, OASense arose as a project from a collaboration between Delft University of Technology (TUDelft), Zimmer Biomet^{TM} (ZimmerBiomet, 2018) and Reinier de Graaf hospital. Zimmer Biomet^{TM} is a global leading company in musculoskeletal healthcare. Their products and solutions help treat patients suffering from disorders of, or injuries to, bones, joints or supporting soft tissues. This seems an excellent starting point for the development of this type of device, with the adequate company giving the needed financial support. Students and researchers from TUDelft developed the first concept in September 2014 in the course of Advanced Concept Design (ACD). In February 2015, the concept was redesigned in the course of Advanced Embodiment Design (AED), with the first prototype being produced and tested in patients from Reinier de Graaf hospital. Later in September of the same year, the project was part of a Demonstrator Grant, where the previous prototype was optimized for further research.

3.2 Development timeline

3.2.1 ACD course

The first concept, which resulted from ACD, is depicted in Figure 3.1. At this stage, no prototype was produced, only conceptualized. The system is composed of an active sensor (piezoelectric or contact microphone) or a electrostethoscope which is attachable to a band in order to fix it on the knee. The signal is forwarded to a processing module, where it is processed. The signal is sent

back to the system, and it returns a light feedback to the GP, with the red light indicating knee damage and the green light showing that the system is on.

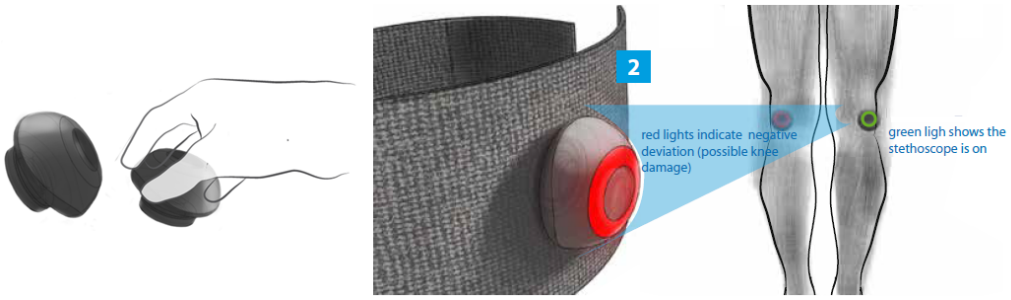


Figure 3.1: ACD product concept

3.2.1.1 AED course

The piezoelectric microphone was selected over the stethoscope due to its smaller size and more simple function. The final sensor chosen was the contact microphone CM-01B (TEconnectivity, 2017).

Regarding the processing unit, the Arduino Uno™ (Arduino, 2018) would be the easiest choice, because of the availability of tutorials and libraries. However, it has only an 8-bit processor. Better alternatives in terms of memory, speed and ADC capabilities, are the Raspberry Pi™ (RaspberryPi, 2018), the Arduino Due™ (Arduino, 2018) and the Teensy™ (Sparkfun, 2018b).

Concerning the Raspberry Pi™, it is not compatible with other operating systems, such as Windows. The Arduino Due™ has a sufficient processing power (32-bit), but it is big and not really suitable for a final prototype. The Teensy™ also has a 32-bit processor and it is smaller than the Arduino Due™ so Teensy™ is the best choice.

In order to make the product fully handheld and portable, as initially intended, it should be able to carry and use a battery for all its subsystems, preferably a rechargeable one. Lithium Polymer (LiPo) batteries are usually selected for such electronic products. In this case, the battery is connected to a micro-USB charger breakout, so it can be charged externally of the product casing.

The device also has an audio shield and earphones so that the doctor can listen to the signal in real time and a SD card to store the data. Two buttons are present (an on-off and a recording button), as well as a multicolour LED to provide light feedback. These components are all represented in Figure 3.2.

As OASense is intended to be a handheld device, it is important to consider how it will fit in the hand of the user and in what way it should be held against the patella. It was concluded that a longer shape provided support for more of the whole hand as opposed to only the finger, which made it more comfortable to hold the device in one place. A shape that was curved, rising in the

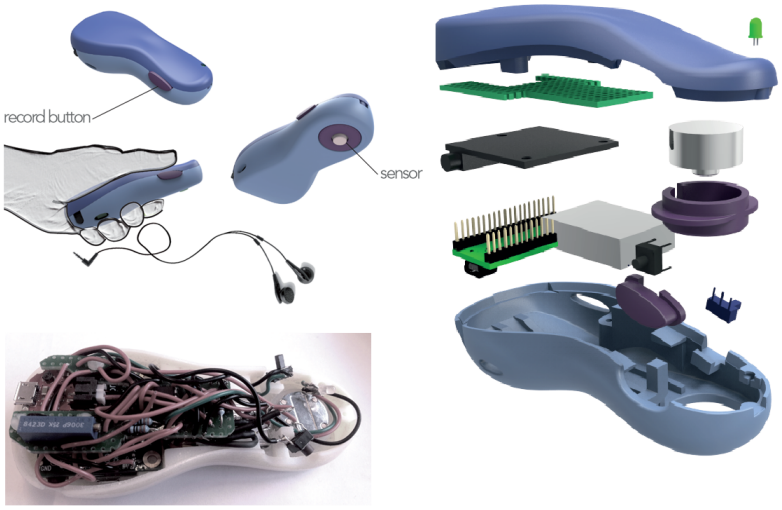


Figure 3.2: AED final product

back, increased this support further by increasing the contact area both from the finger and palm of the hand.

3.2.1.2 Demonstrator Grant

Looking at the prototype produced in AED, a key difference between it and contemporary electronic devices is noticeable, which is a printed circuit board (PCB). This enhancement can imply several benefits for the product, such as a reduced price, a significantly reduced total size of the electronic part and an improvement of the manufacturing and assembly of the product. The final product prototyped for the Demonstrator Grant was basically an optimized version of the previous one in terms of the electronics, as shown in Figure 3.3. A mesh of components was turn into a well-designed PCB which ended up not fitting into the 3D printed ergonomic casing.

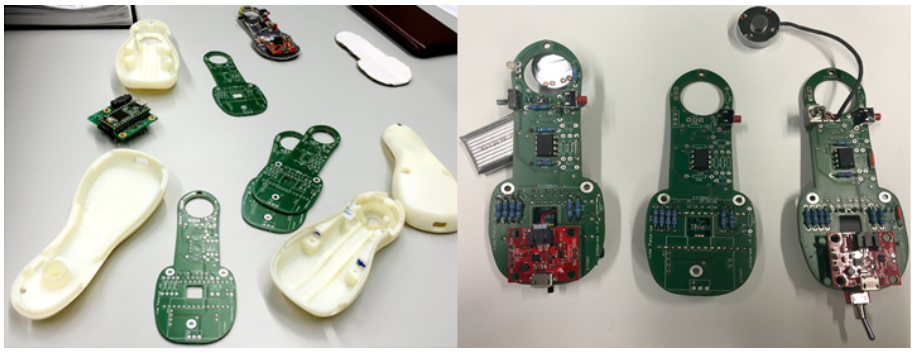


Figure 3.3: Demonstrator Grant final product

3.2.2 Data analysis

The algorithm for the data collection of 20 healthy subjects and 7 people affected by OA was implemented into the Teensy™ (Sparkfun, 2018b). The obtained audio files were clustered and analysed in the time-amplitude domain with the MATLAB® (MATLAB, 2018b) tool Treebagger, which is a machine learning algorithm based on an ensemble of decision trees (MATLAB, 2018e).

This algorithm returned a false positive rate of 5,21%, a false negative rate of 5,21% and 95,82% of true cases. Therefore, the percentage to classify an OA person as healthy is small (3,15%), as well as the classification of a healthy person as OA affected (5,21%).

A machine learning algorithm, like Treebager (MATLAB, 2018e), offers the advantages of dealing with uncertainty in the data gathering and adapting to various scenarios arising from different experimental set up, contrary to some other used statistical analysis tools, such as Boxplot (MATLAB, 2018a). However, it has some drawbacks, such as requiring a considerable sample of data for learning in order to have entrusted results, which is not the case. Thus, a wider number of data samples are required for training, which will allow the algorithm to be generic and work with more datasets.

3.3 Main problems

After all these iterative developments, there are some good things, but also some problems about OASense. The concept is well thought as well as OASense's aesthetic, shape and size and materials used, which make it a user friendly and an ergonomic device.

However, the product could not be validated because of the small dataset acquired and the signal processing and classification algorithms poor robustness. In addition, although the final PCB board is very professional and adequate for commercial use, it ended up not fitting in the 3D printed ergonomic casing. The final version of the PCB also did not work correctly, not allowing data acquisition.

3.4 New concept

In terms of hardware, the main piezoelectric sensor (contact microphone CM-01B (TEconnectivity, 2017)) was kept, which allows the doctor to listen to crepitus in real time. The CM-01B uses a sensitive piezo film combined with a low-noise electronic preamplifier to provide a sound with buffered output, making it ideal for detecting body sounds. Its design minimizes external acoustic noise while offering extremely high sensitivity to vibration applied to the central rubber pad. A new sensor was added, namely a three-axis accelerometer (ADXL335 (AnalogDevices, 2018)), which detects crepitus through vibration. It operates in a dynamic range of -3g to +3g.

Therefore, the device collects a phonoartrographic (with the microphone) and a vibroarthrographic signal (with the accelerometer). It is actually this second one (VAG signal) that will be further processed, due to the narrower frequency range (below 1kHz), the reliable available

dataset of 89 VAG signals broadly present in research (which will be used for understanding the best classification system) and the well described extracted features, processing and classification algorithms in the literature.

The microprocessor used was also kept, Teensy™ 3.1 (Sparkfun, 2018b), as well as the audio shield that allows the listening of the signal in real time and the storage of both Phono-A and VAG signals in a SD card, while being recorded, for further analysis.

A LiPo battery of 500mAh was added to the device, together with a charger. One switch button and one pushbutton are present. The first is used to turn the device on and off and the second, when pressed, starts the recording and, when pressed again, stops it. A RG (red-green) led was also installed, which transmits feedback to the user. When green, it means the device is on, but not recording. When red, it means the device is on and in recording mode.

For the first prototype (Figure 3.4), the casing was kept as it was. However, since now two sensors have to be directly in contact with the subject's knee for the acquisition of both Phono-A and VAG signals, an external 3D-printed part had to be added to the casing.

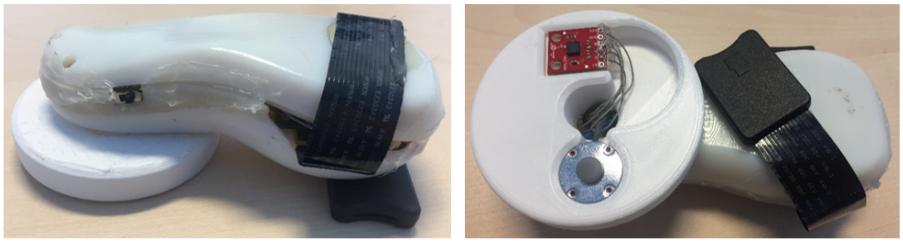


Figure 3.4: First prototype of the new concept of OASense. The external part incorporates both the accelerometer and the microphone. The SD card reader had also to be kept outside the casing.

The PCB board previously designed for the device could not be used, since it didn't fit in the casing and it didn't work properly. Therefore, a new, functional, but less robust electronic system was built (see Appendix A.1 for the schematics), with the hardware mentioned above, and inserted into the casing.

For the second prototype, a new casing as well as a new functional PCB (based on the same electronic schematics of Appendix A.1) were designed. In Figure 3.5, it is possible to see that the two sensors are now closer, which allows for a clearer listening of crepitus by the doctor. However, a compromise had to be done and the shape of the device was changed, so that all the components could fit in a hand-size device.

In terms of hardware, a code was written using Arduino® (Arduino, 2018) environment that allows the Teensy™ (Sparkfun, 2018b) to collect and store data as well as display the light feedback, assuming the state of both buttons. New and powerful processing and classification algorithms were implemented in MATLAB® (MATLAB, 2018b), which will be described in Chapter 5.

Figure 3.5: Bottom, top and side views of the second prototype. On the bottom view, the two sensors can be seen. On the side views, both buttons, the earphones, the SD card and the micro USB for charging outlets can be observed.

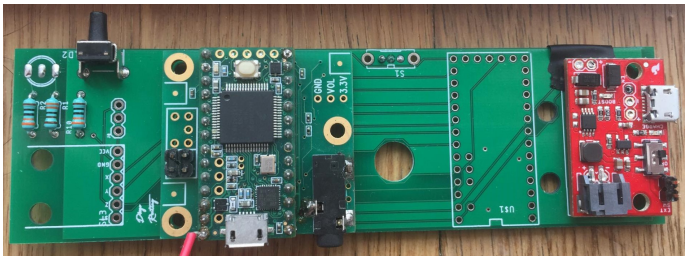


Figure 3.6: PCB of the second prototype with some of the components represented. Namely, the battery, the accelerometer and the switch button are missing.

Chapter 4

Studies on vibroarthrography

A literature review of the most relevant VAG related studies for knee OA detection is provided in this chapter, focusing on the current available strategies for OA signal classification, its advantages and major drawbacks.

Firstly, the raw time series VAG signal is collected and is subjected to several pre-processing techniques, which aim at removing the artifacts in the raw signal. Such artifacts mainly include baseline wander, random noise, and periodical power-line interference.

Several computational methods can then be employed to study the signal in the time or frequency domain. Thus, several features are extracted, whose discriminant capacity is analysed in the selection phase. Finally, the classification model can be trained with these features, whose goal is to accurately distinguish between normal and abnormal VAG signals.

4.1 Signal acquisition

The first step consists on the VAG signal acquisition. In order to obtain a reliable signal which describes and fully characterizes the knee joint status, several considerations must be taken into account. Among them, it must be considered the type of sensor used, its inherent limitations (e.g. sampling frequency, noise), its placing on the knee joint and the movement executed by the subject.

4.1.1 Sensors

4.1.1.1 Accelerometer

Accelerometers are sensors broadly used (Krishnan et al., 2001; Rangayyan and Wu, 2008, 2009a; Cai et al., 2013; Rangayyan and Wu, 2010; Wu et al., 2013, 2014; Liu et al., 2014; Befrui et al., 2018; Kręcisz and Bączkowicz, 2018) for the purpose of VAG signal acquisition, when in direct contact with the skin surface. They are electromechanical devices that sense either gravity, vibrations or movement (Sparkfun, 2018a).

These sensors measure acceleration, in meters per second squared (m/s^2) or in G-forces (g) (1g=9.8m/s^2 on planet Earth). This can be done on one, two or three axis. The most sensitive and stable are the capacitive accelerometers, which contain capacitive plates internally. Some of these are fixed, while others are attached to minute springs that move internally as acceleration forces act upon the sensor. As these plates move in relation to each other, the capacitance between them changes. From these changes in capacitance, the acceleration can be determined (Sparkfun, 2018a; Wong et al., 2007). Other accelerometers can be centered around piezoelectric materials. These tiny crystal structures output electrical charge when placed under mechanical stress (Wong et al., 2007).

4.1.1.2 Stethoscope

Stethoscopes have been used to listen to heart sounds for diagnostic purposes since 1815 when it was invented by René Leannec (Grais, 2013). They have three main components: the chest piece, tubes and headset. The chest piece is composed of two sides: a diaphragm (plastic disc) responsible for the transmission of high frequency sounds and a bell (hollow cup) for the low frequency waves. When in contact with the patient, the chest piece senses the vibration caused by the body sounds creating acoustic pressure waves which travel through the tubes to the listener's ears (Singh et al., 2013).

The common acoustic stethoscopes have the disadvantage of a reduced sound amplification, becoming difficult for the doctor to listen to it clearly. On the other hand, the electronic stethoscopes, mostly known as electrostethoscopes, convert the sound to an electric signal which can be transmitted in a wireless way to an external platform for posterior analysis. These devices allow a better listening of the body sounds, since the signal can be digitally amplified and filtered for noise reduction (Singh et al., 2013).

Among the literature studies revised, the acoustic stethoscopes are actually used together with the signal acquisition by an accelerometer (Krishnan et al., 2001; Rangayyan and Wu, 2008, 2010; Cai et al., 2013; Wu et al., 2013), in order to obtain a qualitative description of sound intensity and to facilitate auscultation of important diagnostic events.

In turn, VAG signals can be recorded by an electrostethoscope without the need of using an accelerometer (Kim et al., 2009b,c,a; Lee et al., 2012; Lin et al., 2014; Wu et al., 2014; Liu et al., 2014; Seng et al., 2013).

4.1.1.3 Electrogoniometer

It is important to record the knee joint angle at the same time the VAG signals are being collected, in order to segment the leg movement, which gives precise information about the phase of the movement the VAG signal belongs to. While it is possible to calculate it with accelerometer data, the electrostethoscope does not allow it. Therefore, a flexible electrogoniometer can be used together with this device to measure the knee joint angle on the lateral aspect of the knee with the axis of rotation at the joint line (Kim et al., 2009b,a; Lee et al., 2012; Lin et al., 2014).

4.1.2 Experimental setup

4.1.2.1 Sensor positioning

It is crucial to position the sensor correctly on the skin surface in order to obtain an accurate measurement of the signals. The optimal location is generally accepted to be slightly below the midline of the patella (medial condyle on the patella), being the closest point to the contact area between moving knee joint surfaces (Figure 4.1). This position is relatively stable, since it is not affected by the actual knee joint movement. All the analysed studies which perform signal acquisition record the signals in this exact position (Krishnan et al., 2001; Rangayyan and Wu, 2008; Kim et al., 2009b,c; Rangayyan and Wu, 2009a; Kim et al., 2009a; Rangayyan and Wu, 2010; Lee et al., 2012; Cai et al., 2013; Wu et al., 2013; Lin et al., 2014; Wu et al., 2014; Liu et al., 2014; Befrui et al., 2018; Kręcisz and Bączkowicz, 2018; Seng et al., 2013).

Figure 4.1: Optimal sensor position for crepitus detection (Krishnan et al., 2001).

4.1.2.2 Experimental protocol

One of the most common performed tasks for the collection of VAG signals is the knee movement from full flexion to full extension, as shown in Figure 4.2. In the full flexion position, the knee forms an angle of 90º between the femur and the tibia, whereas in the full extension, the angle formed is 0º. The motion from the full flexion to full extension and back to full flexion again (90º-0º-90º) is defined as one cycle (Wu, 2015).

During the recording, each subject is sat on a rigid table in a relaxed position with the leg under test freely suspended in air while executing the movement. Some training should be done prior to the experiment in order to warm up the knee joint as well as for the establishment of a constant velocity during the whole exercise (usually 2 or 4s/cycle). In addition, the experiment is normally divided in periods of 20-30s and between each period the subject should have a rest period between 1-3 minutes in order to avoid muscle fatigue (Rangayyan and Wu, 2008; Kim et al., 2009c). The velocity of the movement should remain constant in order to not affect the frequency characteristics of the VAG signals (Rangayyan and Wu, 2008).

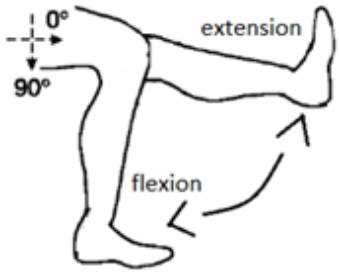


Figure 4.2: Leg flexion-extension movement (Wu, 2015).

4.2 Pre-processing

In clinical application, it is essential to record high-quality VAG signals for computer-aided diagnostic analysis of knee joint pathology (Wu, 2015). However, the acquired VAG signals are of a very complex nature as they result from the combination of several components, such as random or muscular noise, ambient interference and baseline wander (Moreira, 2015; Wu, 2015). Therefore, the component of interest of a VAG signal consists on the low amplitude remaining signal after the removal of the components mentioned above, which corresponds, ideally, to the knee joint vibrations.

The muscular noise is commonly induced by concurrent muscular contractions during the knee extension-flexion movement. Random noise due to the thermal effect in the cables is inevitable. However, this mechanism is complex and random, so the range of signal-to-noise ratio of VAG signals cannot be determined a priori (Wu, 2015; Wu et al., 2010). The baseline wander could be caused by some additional movement, like trembling of the leg or skin friction (Wu, 2015).

Several approaches are documented in the literature in order to remove some of these interferences. The simpler strategies consist on the minimization of the noise generation during the data collection itself. Thus, an optimal placement of the sensor as well as an use of conductive gels or double-sided tape between the probe surface and the skin may reduce in great scale the friction between the skin and the sensor (Krishnan et al., 2001; Wu et al., 2014). Other techniques are used, such as band-pass filtering, signal normalization, adaptive filtering and signal segmentation (Rangayyan and Wu, 2008; Kim et al., 2008, 2009c; Rangayyan and Wu, 2009a).

4.2.1 Normalization

Each collected signal has its own range of amplitude, that may differ among subjects or trials. Therefore, amplitude normalization is one of the pre-processing procedures applied to VAG signals, where the amplitudes are scaled so as to fall within a specified range, such as -1.0 to 1.0 or 0.0 to 1.0, with the purpose of providing a more general comparison between signals (Visalakshi and Thangavel, 2009).

The simplest and most common used amplitude normalization method is known as minimum-maximum (min-max) normalization, in which each signal's amplitude value can be scaled in order

to have a new amplitude within a certain and pre-defined interval of values (Patel and Mehta, 2011). The min-max normalization method can be obtained, for a signal x, according to equation 4.1.

$$x_{norm}[n] = (x[n] - x_{min}) * \frac{new_{max} - new_{min}}{x_{max} - x_{min}} + new_{min} \tag{4.1}$$

4.2.2 Filtering

In signal processing, the function of a filter is to remove unwanted parts of the signal, such as random noise, or to extract useful parts of the signal, such as the components lying within a certain frequency range. Analog and digital are the two main types of filtering available. While an analog filter produces the required effect by inserting specific components like resistors, capacitors and amplifiers in the electronic circuit, a digital filter acts after the analog input signal is sampled and digitised (UCDavis, 2014).

Digital filtering is most commonly applied on VAG signals processing (Krishnan et al., 2001; Rangayyan and Wu, 2008; Kim et al., 2008, 2009b,c; Rangayyan and Wu, 2009a; Kim et al., 2009a; Rangayyan and Wu, 2010). On one hand, fixed filtering is related to a low-pass, band-pass or high-pass filter, with specific cut-off frequencies and pre-defined characteristics. On the other hand, adaptive filtering involves the changing of filter parameters adapting to the signal changing characteristics, making it more complex but with better results than fixed filtering, in the case of VAG signals, which are non-stationary (UCDavis, 2014; Douglas, 1999).

4.2.2.1 Fixed Filtering

VAG signals are usually processed by a band-pass filter with a bandwidth of 10Hz to 1kHz, when a sampling frequency of 2kHz is used, in order to remove low-frequency movement artifacts, high-frequency noise and muscle contraction interferences (Krishnan et al., 2001; Rangayyan and Wu, 2008, 2009a; Kim et al., 2009a; Rangayyan and Wu, 2010; Nalband et al., 2016; Rangayyan et al., 2013; Wu et al., 2014; Befrui et al., 2018).

However, when the artifacts overlap the VAG signal spectrum, fixed filtering does not present a correct performance. Thus, some more appropriate and adaptable filters are required.

4.2.2.2 Adaptive Filtering

For the removal of baseline wander and random noise, adaptive filtering techniques are commonly used (Wu, 2015; Kim et al., 2009b; Cai et al., 2013; Wu et al., 2013). While baseline wander provokes oscillations in the signal from the straight baseline line over short periods of time, random noise adds random fluctuations to the signal.

A cascade moving average filter may be used to remove the baseline wander, by first estimating the artifact and then subtracting it from the raw signal. This type of filter consists of a hierarchical structure that combines two-layer moving average operators (Figure 4.3), which average the samples in a temporal moving window to produce the output at various points of time.

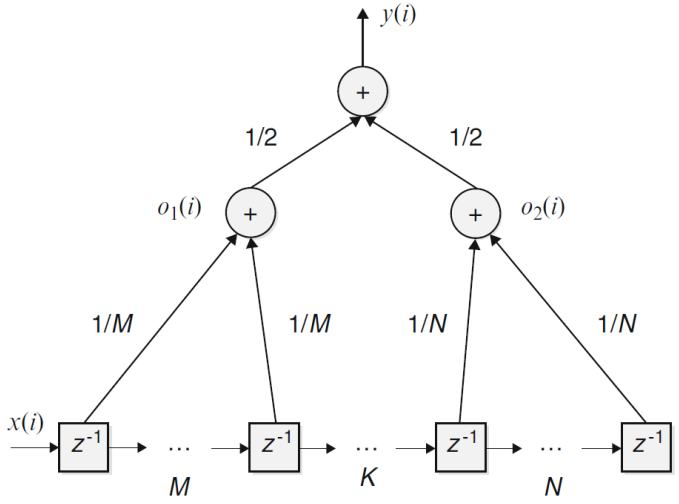


Figure 4.3: Hierarchical structure of the cascade moving average filter (Wu, 2015).

The first layer consists on an M-order and an N-order successive moving average operators. The K inputs in the tail end of the M-order operator overlap with the beginning inputs of the N-order operator (Cai et al., 2012; Wu, 2015). The output of the M-order operator and N-operator can be expressed as $o_1(i)$ and $o_2(i)$, respectively.

$$o_1(i) = \frac{1}{M}[x(i-1) + ... + x(i-M)]$$

$$= \frac{1}{M}\sum_{m=1}^{M} x(i-m) \tag{4.2}$$

$$o_2(i) = \frac{1}{N}[x(i-M+K) + ... + x(i-M+K-N)]$$

$$= \frac{1}{N}\sum_{n=1}^{N} x(i-M+K-n) \tag{4.3}$$

The second layer of the cascade filter aims at smoothing the piecewise linear trends obtained from the outputs of the two operators in the first layer (Cai et al., 2012; Wu, 2015). Therefore, the final output of the filter is given by

$$y(i) = [o_1(i) + o_2(i)]/2$$

$$= \frac{1}{2M}\sum_{m=1}^{M} x(i-m) + \frac{1}{2N}\sum_{n=1}^{N} x(i-M+K-n). \tag{4.4}$$

For the removal of random noise, a recent and effective method that combines the ensemble empirical mode decomposition (EEMD) and detrended fluctuation analysis (DFA) algorithms has

been used (Wu et al., 2014; Liu et al., 2014; Wu, 2015) after being proposed by Wu and Huang (2009). Firstly, the EEMD divides the signal into several intrinsic mode functions (IMFs) that represent the fast and slow oscillations in the signal. Then, the DFA algorithm is applied to identify the inherent correlation property of each IMF, through the computation of the fractal scaling index that describes the subtle fluctuations associated with intrinsic correlations of the dynamics in the signal. Finally, the IMFs that contain the dominant artifacts of random noise can be removed (Wu, 2015). Although giving good results, this method is computationally expensive, so it becomes difficult to use with considerable amounts of data.

4.2.3 Segmentation

In order to minimize the effect of the variation in the knee angular velocity on the frequency characteristics of the VAG signals, the signals obtained during one motion cycle can be segmented according to the recorded knee joint angle. This technique is also useful for the extraction of the segments that are less affected by noise (Rangayyan and Wu, 2010; Lee et al., 2012).

One knee motion cycle comprises an angular range of approximately 90º. In fact, several studies reported the exclusion of the initial and end phase movement segments, namely the initial 0º-25º and the last 75º-90º. Only the segments correspondent to the middle phase of the movement, i.e., segments with angles between 25º-75º were kept for further analysis. This exclusion can be explained due to the ambient noise related to acceleration/deceleration of knee joint being more pronounced during this period. Also, the patella is not fully contacting the femur surface in these segments (Rangayyan and Wu, 2010; Lee et al., 2012).

4.3 Feature extraction

Feature extraction is a vital step for the development of an accurate and robust classifier that is capable to discriminate between classes, which means, in this case, distinguish an OA affected knee joint from an unaffected one. The major goal of feature extraction is to obtain the most relevant information from the original data, extracting from it an array of features which maximizes the classification results with the least amount of elements (Moreira, 2015; Kumar and Bhatia, 2014).

This step can normally be performed either in time and frequency domain, resulting in a different set of characteristics which fully describe the signal. However, the VAG signal is non-stationary, which means it exhibits time-varying spectral characteristics, and it cannot be accurately represented by common signal processing techniques such as Fourier transform (FT). Therefore, it is useful to use a joint time-frequency approach for VAG analysis, extracting a new set of time-frequency based features (Moreira, 2015; Kim et al., 2009c).

4.3.1 Spatiotemporal features

Three different spatiotemporal features were extracted from each VAG signal, namely the form factor (FF), the variance of the mean-squared (VMS) values and the turns count (TC).

The FF value can be computed with the activity and mobility parameters (Wu, 2015). The activity parameter is defined as the variance σ_x^2 of the given signal x. The mobility parameter, M_x, is computed as the square root of the ratio of the variance of the first derivative $x^{'}$ of the signal over the variance of the original signal, i.e.,

$$M_x = \left[\frac{\sigma_{x^{'}}^2}{\sigma_x^2}\right]^{\frac{1}{2}} = \frac{\sigma_{x^{'}}}{\sigma_x}. \tag{4.5}$$

FF is then defined as the ratio of the mobility of the first derivative of the signal over the mobility of the original signal (Wu, 2015), which is expressed as

$$FF = \frac{M_{x^{'}}}{M_x} = \frac{\sigma_{x^{''}}}{\sigma_x}, \tag{4.6}$$

where σ_x, $\sigma_x^{'}$ and $\sigma_x^{''}$ represent the standard deviation of the signal x, of the first derivative $x^{'}$ and of the second derivative $x^{''}$, respectively.

VMS values are relevant to measure the local variations of a given VAG signal (Rangayyan and Wu, 2009a; Wu et al., 2013; Liu et al., 2014; Kręcisz and Bączkowicz, 2018). These values are normally calculated in the nonoverlapping fixed-duration segments of 5ms each (Wu, 2015).

Finally, the signal variability over a certain threshold can be measured with the turns count parameter, which accumulates the number of significant changes in direction for a given signal. The value obtained for the TC over the time series represents the degree of fluctuation dynamics in the signal (Moreira, 2015; Wu, 2015).

A sample can be considered as a turn if it satisfies, simultaneously, the following two conditions (Wu, 2015):

- It represents an alteration in direction in the signal, i.e., a change in the sign of the derivative.

$$TC_{candidate}[k] = [x(n) - x(n-1)][x(n+1) - x(n)] < 0 (2 \leq n \leq N-1) \tag{4.7}$$

- The difference (absolute value) between the amplitude of the current alteration and that of the preceding alteration is greater than a threshold.

$$TC_{final} = |TC_{candidate}[k] - TC_{candidate}[k-1]| \geq Th \tag{4.8}$$

where $x[n]$ is the nth sample, with n=0,1,...,N, $TC_{candidate}[k]$ the extracted sample candidates that satisfy the first condition, Th the predifined threshold and TC_{final} the final TC sequence.

The value for the threshold can be predefined with the fixed or adaptive style. Rangayyan and Wu (2009b) report the use of a fixed threshold value of 0.2, considering an amplitude-normalized VAG signal in the range [-1,1]. Alternatively, the threshold value can be estimated as 0.5σ, i.e., taking into account the standard deviation of the signal (Wu, 2015).

4.3.2 Statistical Features

Some statistical parameters, such as mean (μ), standard deviation (σ), variance (σ^2), skewness (SK), kurtosis (KU) and entropy (H) can be also helpful in providing discriminant features of the VAG signals (Rangayyan and Wu, 2008, 2009a, 2010; Lee et al., 2012; Cai et al., 2013; Lin et al., 2014). To derive these parameters, the histogram is necessary as a reference of the probability density function (PDF) of the signal. The histogram is often estimated with L bins, which helps to calculate the probability of occurrence with L containers of equal length in the amplitude range of the signal (Wu, 2015).

Skewness is related to the asymmetry of the PDF. Kurtosis measures the tendency of the PDF to have peaks, i.e., evaluates whether data is peaked or flat in relation to a normal distribution. Entropy represents the nature and spread of a PDF, measuring the inherent randomness observed (Wu, 2015; Rangayyan and Wu, 2008). SK, KU and H are computed as follows

$$SK = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^3}, \tag{4.9}$$

$$KU = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^4}, \tag{4.10}$$

$$H = -\sum_{L=0}^{L-1} p_x(x_l) log_2[p_x(x_l)], \tag{4.11}$$

where n is the total number of samples, x_i the sample with the i index and $\bar{x}$ the mean value of all samples. $p_x(x_l)$ means the PDF of the given signal with x_l, l=0,1,2,...,L-1 representing the L bins used to represent the range of value of the signal x.

4.3.3 Time-frequency features

The frequency domain contains the information about the inherent rhythms and periodicity of a given VAG signal. The Fourier transform (FT) is a method often used to transform signals between time domain and frequency domain, allowing their frequency analysis (Wu, 2015). In turn, the Fast Fourier Transform (FFT) is an efficient algorithm for the estimation of the FT (in the discrete domain), as represented in Equation 4.12, with N representing the number of samples.

$$F(k) = \sum_{n=0}^{N-1} x(n) e^{-2\pi ikm/N} \tag{4.12}$$

However, FT is independent of time, i.e., the frequency decomposition is averaged over the total duration of the signal by assuming that the random process behind the signal creation is stationary and does not change with time (Moreira, 2015). Therefore, the inability to describe the varying frequency components in response to the elapsed time is certainly a major disadvantage of the frequency analysis.

Knowing that the VAG signal is non-stationary and with time-varying spectral properties, time-frequency distribution (TFD) analysis methods arise as a solution that is able to overcome the frequency analysis disadvantage (Wu, 2015). After FT poor time resolution, Short Time Fourier Transform (STFT) came up and is considered more suitable for non-stationary signals. STFT basically consists on the application of the FFT algorithm over a window that slides over time (Moreira, 2015). STFT is defined in Equation 4.13, where the $w(n-m)$ is the window function.

$$X(n,k) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-2\pi ikm/N} \tag{4.13}$$

Wigner-Ville distribution (WVD), similarly to the STFT, is a time-frequency representation of signals. WVD correlates the input signal with a time-translated and frequency-translated version of itself, i.e., the window consists on a shifted version of the same input signal (Castiglioni, 2015). The WVD can be calculated as it follows

$$W(t,f) = \frac{1}{2}\int x^*(t-\frac{1}{2}\tau)x(t+\frac{1}{2}\tau)e^{-2\pi if\tau}d\tau, \tag{4.14}$$

where $x^*(t)$ is the complex conjugate of $x(t)$ and f represents a certain frequency band. However, WVD may lead to a difficult interpretation of the result, particularly for multi-component signals. Therefore, as a way to overcome this problem, the wavelet-based analysis appeared. It is defined as mathematical functions that cup data into different frequency components, and then study each component with a resolution matched to its scale (Hoang, 2014; Lee and Yamamoto, 1994).

Wavelet Transform (WT) processing is different from other signal-processing methods, such as the Fourier based-transforms which break down the input signal into its harmonics, because of the unique properties of the local basis function wavelets. They can represent signals sparsely, capture the transient features of signals and enable signal analysis at multiple resolutions (Wu, 2015; Hoang, 2014).

WT consists on computing coefficients, which are inner products of the signal and a family of wavelets. Thus, WT decomposes a signal into localized contributions labelled by a scale and a position parameter. The spatial localization of wavelets makes them more suited to present a large class of signals (i.e. spectra can be represented by far fewer wavelets than sinusoids obtained from FT) (Hoang, 2014). Generally, the Wavelet Transform of a continuous signal can be written as

$$W(\tau,s) = \frac{1}{\sqrt{|s|}}\int_{-\infty}^{\infty} x(t)\Psi(\frac{t-\tau}{s})dt, \tag{4.15}$$

where τ and s are the so-called translation (or time location) factor and the scaling (or dilation) factor, respectively. Factor $\sqrt{|s|}$ is for energy normalization across the different scales, whereas $\Psi_{\tau,s}(t)$ can be obtained by dilations and translations of a single function $\Psi(t)$, the so-called "mother wavelet" (Hoang, 2014), as follows

$$\Psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}}\Psi\left(\frac{t-\tau}{s}\right). \tag{4.16}$$

Despite the flexible time-frequency resolution properties of WT, the frequency resolution could be poor in the high-frequency region. Therefore, there are some difficulties in discriminating between signals having close high-frequency components. A more generalist approach of the wavelets has arisen, called wavelets packets or wavelet packet decomposition (WPD) (Gokhale and Khanduja, 2010). These wavelet packets are formed by taking linear combination of the common wavelets functions, preserving some properties such as orthogonality, smoothness and location from their corresponding mother wavelets (Gokhale and Khanduja, 2010; Romero et al., 2005).

The full WPD of a discrete-time signal is calculated by applying both low and high-pass filters followed by the decimation procedure (usually downsampling by a factor of 2) to the input signal and then recursively to each intermediate signal's version until the last level of decomposition is reached (Romero et al., 2005). The equation describing the discrete Wavelet Packet Transform (WPT) can be represented by

$$W_{i,k}^{n}(t) = 2^{i/2} W^{n}(2^{i}t - k), \tag{4.17}$$

where the index n is called the modulation or oscillation parameter. The first two wavelet packet functions are the usual scaling function and motherwavelet function, respectively

$$W_{0,0}^{0}(t) = \phi(t), \tag{4.18}$$

$$W_{0,0}^{1}(t) = \Psi(t). \tag{4.19}$$

Wavelet packet functions for $n = 2, 3, \ldots$ are then defined by the following recursive relationships

$$W_{0,0}^{2n}(t) = \sqrt{2} \sum_{k} h(k) W_{1,k}^{n}(2t - k), \tag{4.20}$$

$$W_{0,0}^{2n+1}(t) = \sqrt{2} \sum_{k} g(k) W_{1,k}^{n}(2t - k), \tag{4.21}$$

where $h(k)$ and $g(k)$ are the quadrature mirror filters (QMF). Specific time-frequency information measurements of a signal can be obtained by taking the inner product of the signal and that particular basis function, resulting in a set of wavelet packets coefficients of the given function $f(t)$ according to the following equation

$$W_{i,n,k} = \left\langle f W_{i,k}^{n} \right\rangle = \int f(t) W_{i,k}^{n}(t) dt. \tag{4.22}$$

The major advantage of the WPT approach over the WT is that the latter only decomposes the low-frequency components of the intermediate signals at each node while the WPT decomposes not only the low but also the high-frequency components of all signals (see Figures 4.4 and 4.5), which allows a more consistent analysis of the input signal, providing signal's characterization

for both stationary and repetitive events but also the punctual and non-stationary (Gokhale and Khanduja, 2010; Romero et al., 2005).

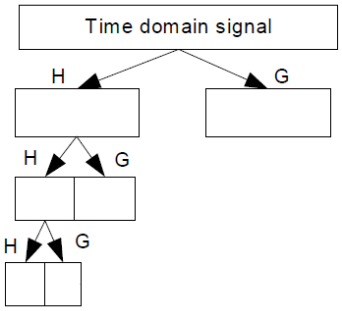


Figure 4.4: Wavelet decomposition of time-domain signal (Romero et al., 2005)

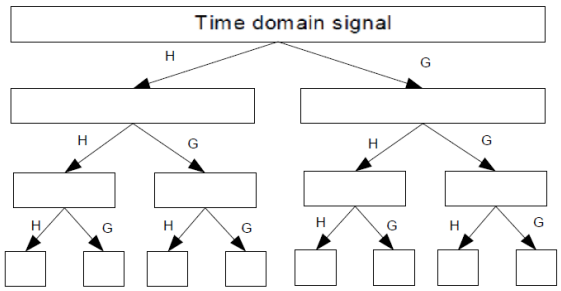


Figure 4.5: Wavelet packet decomposition of time-domain signal (Romero et al., 2005)

An equivalent approach for time-frequency decomposition of a signal was introduced by Mallat and Zhang (1993), using the Matching Pursuit (MP) algorithm, also known as "greedy" algorithm. This method successively approximates a signal $x(t)$ of N samples onto elements from a dictionary $D = \{d_r(t)\}_{r\in\Gamma}$ of P vectors, in which $\|d_r\| = \sqrt{[\int d_r^2(t)dt]} = 1$. These dictionaries can be built up using Gabor function, local cosine trees or wavelet packets (Wu, 2015).

The projection of VAG signal $x(t)$ using the dictionary of wavelet packet bases $d_{r_m}(t)$, calculated with a Daubechies wavelet filter, can be formulated as

$$x(t) = \sum_{m=0}^{M-1} a_m d_{r_m}(t), \tag{4.23}$$

where a_m are the expansion coefficients and M denotes the iterations of decomposition.

To implement the wavelet MP decomposition, $x(t)$ is first projected on a vector $d_{r_0}(t) \in D$ and the residue $R^1x(t)$ is computed, i.e.,

$$x(t) = \langle x, d_{r_0}\rangle d_{r_0}(t) + R^1x(t), \tag{4.24}$$

where $\langle x, d_{r_0}\rangle$ denotes the inner product.

The MP iterates this procedure by subdecomposing the residue. Thus, the VAG signal $x(t)$ after M iterations of decomposition is expressed as

$$x(t) = \sum_{m=1}^{M-1} \langle R^m x, d_{r_m} \rangle \, d_{r_m}(t) + R^M x(t). \tag{4.25}$$

A straightforward way to stop the iterative decomposition process it to predefine the number M of the time-frequency atoms.

Some relevant studies (Kim et al., 2008, 2009b,c,a) extracted four time-frequency parameters from VAG signals, namely the energy parameter (EP), energy spread parameter (ESP), frequency parameter (FP) and frequency spread parameter (FSP). The features are represented by the average and standard deviation of each of the parameters.

EP represents the energy variation with time, being calculated as the mean of the signal along each time slice. ESP measures the spread of energy over the frequency range across time and is computed as the standard deviation of the signal along each time slice. FP, also known as the instantaneous mean frequency, represents the first moment of the signal along each time slice. FSP is given by the second central moment of the signal along each time slice, measuring the spreadability of frequencies across time (Wu, 2015; Kim et al., 2009a).

EP, ESP, FP and FSP are defined as it follows

$$EP(t) = \frac{\sum_{f=0}^{fm} X_e(t,f)}{f_m}, \tag{4.26}$$

$$ESP(t) = \left[\frac{\sum_{f=0}^{fm} |X_e(t,f) - EP(t)|^2}{f_m} \right]^{1/2}, \tag{4.27}$$

$$FP(t) = \frac{\sum_{f=0}^{fm} f X_e(t,f)}{\sum_{f=0}^{fm} X_e(t,f)}, \tag{4.28}$$

$$FSP(t) = \left[\frac{\sum_{f=0}^{fm} |X_e(t,f) - EP(t)|^2}{\sum_{f=0}^{fm} X_e(t,f)}, \right]^{1/2} \tag{4.29}$$

where $X_e(t,f)$ represents the value of the matrix derived from the TFD of the VAG signal at a given time t and frequency f and f_m is the maximum frequency in the signal.

4.4 Feature selection and Dimensionality reduction

Feature selection is the process of detecting the relevant features and discarding the irrelevant and redundant ones, representing a major step in machine learning. Irrelevant features may lead to overfitting, specially when the number of features is high but the number of samples is small (Bolón-Canedo et al., 2013; Liu and Motoda, 2007). Reducing the number of irrelevant/redundant features drastically reduces the running time of a machine learning algorithm and

yields a more general concept (Dash and Liu, 1997). Ideally, this method would search through all the 2^N candidate subsets (N being the number of features) and try to find the best one, according to some evaluation criterion. However, this has a high computational cost, even for a medium-sized feature set, and there is also the risk of overfitting (Dash and Liu, 1997; Liu and Motoda, 2007).

There are two major approaches in feature selection: individual evaluation and subset evaluation. The former, also known as feature ranking, assesses individual features by assigning them weights according to their relevance. The latter, which is much more used, produces candidate feature subsets based on a certain search strategy. While the individual evaluation is incapable of removing redundant features since they are likely to have similar rankings, the subset evaluation can handle feature redundancy with feature relevance (Bolón-Canedo et al., 2013).

With regard to the relationship between a feature selection algorithm and the inductive learning method used to infer a model, three models can be distinguished: filters, embedded and wrapper methods. Filters rely on the general characteristics of training data and carry out the feature selection process as a pre-processing step, independently from the learning algorithm. Embedded perform feature selection in the process of training. Wrappers consist on the optimization of the model as part of the selection process. However, these methods have the highest computational costs (Bolón-Canedo et al., 2013).

The most commonly used techniques are sequential forward selection (SFS), sequential backward selection (SBS), bidirectional search and principal component analysis (PCA).

Sequential search methods such as the SFS, SBS or bidirectional search create, iteratively, a subset of features (by adding and/or removing features) that is evaluated according to a certain criterion, usually, predictive accuracy through a cross-validation method. The SFS method starts with an empty feature set and sequentially adds one feature at a time. Once a feature is retained in the subset, it cannot be discarded in the next searching step. On the other hand, the SBS method begins with all available features and sequentially excludes one feature that makes least contribution to the chosen criterion. In general, SBS method is more computationally expensive than the SFS approach. The bidirectional search starts with a certain set of features and simultaneously adds and removes features (Wu, 2015; Moreira, 2015).

PCA is the most popular dimensionality reduction method currently used in the field of data science. It generates new features which are a linear combination of the original ones, mapping each instance of the given dataset present in a d dimensional space to a k dimensional subspace such that $k < d$ (Vasan and Surendiran, 2016).

This set of k new generated dimensions are called the Principal Components (PC), and each PC is directed toward maximum variance. Therefore, the first PC covers the maximum variance and each component that follows covers less value of variance. The Principal Components can be represented as it follows

$$PC_i = a_1X_1 + a_2X_2 + ... + a_dX_d, \tag{4.30}$$

where PC_i represents the principal component i, X_j the original feature j and a_j the numerical

coefficient for X_j. It is often chosen, in most machine learning problems, to keep at least 95% of the original's set variance (Vasan and Surendiran, 2016; MATLAB, 2015).

4.5 Classification

Classification can be defined as the problem of assigning a given instance in a dataset to a class or label. Therefore, its goal is to accurately predict the label of a new observation using a model that was built with the extraction of relevant information from a dataset composed of similar examples. If these instances are given with known labels (the corresponding correct outputs) then it is a supervised classification. On the other hand, unsupervised classification deals with instances which are unlabeled (Wu, 2015; Moreira, 2015; Kotsiantis et al., 2006).

For this specific work with VAG signals (binomial classification), only supervised classification techniques will be used. A supervised learning algorithm takes a known set of data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data.

In order to predict the outcome of unlabeled data, a supervised learning algorithm usually requires five steps: data preparation, model selection, classifier training, classifier evaluation and finally output prediction. Data preparation was mentioned in the previous section and includes the processes of feature extraction and selection. Then, the next three steps are normally integrated, with the choice of the model being made according to its performance with the training dataset. The input of the classifier consists on the selected features and labels from the training set. Finally, the trained model is used to make the predictions for the unlabeled output, returning a class for each instance (Kotsiantis et al., 2006).

4.5.1 Decision trees

Decision Trees is a widely used and one of the easiest to understand classification algorithms in the data mining field due to its practical and feature-guided nature. This algorithm uses a tree-like shape model (with nodes, branches and leaves) to predict the classes of an unlabeled set of instances by sorting them according to their feature value (Wu, 2015; Kotsiantis et al., 2006).

Each node represents a feature in an instance to be classified and each leaf represents a value that the node can assume. The attribution of a given label is achieved by going through the path from the root node to the final leaf. The most discriminant feature that best divides the training data is the root node of the tree. The path since the root node to the final leaf is constituted by branches that represent the set of features and the possible range of values used to predict the final class label. At each node, a given feature is evaluated and each branch stands for the possible outcome, the final leaf being the predicted class label (Wu, 2015; Moreira, 2015; Kotsiantis et al., 2006). In Figure 4.6, a simple decision tree for a binomial classification problem is represented.

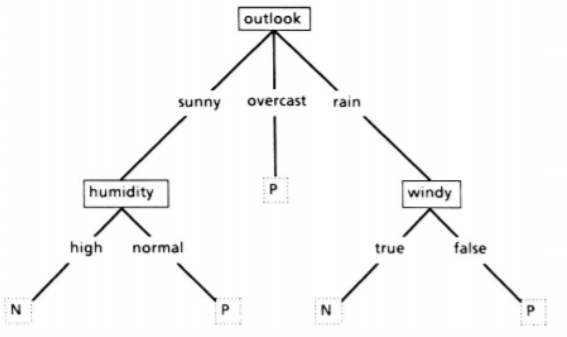


Figure 4.6: Example of a decision tree for a binomial classification. It classifies a given morning as suitable or not (P or N) for playing tennis (Quinlan, 1986).

4.5.2 k-nearest neighbors

k-NN is one of the simplest and trivial used classifiers in the data mining field. It is an instance-based learning algorithm, being characterized by its laziness as it delays the induction or generalization process until classification is performed. These type of algorithms require less computation time during the training phase than decision trees, neural networks or Bayesian networks, but more computation time during the classification process itself (Kotsiantis et al., 2006; Wu, 2015).

k-NN is based on the principle that the data points within a dataset will generally exist in close proximity to other points that have similar properties. Therefore, the label of an unclassified instance can be determined by observing the class of its nearest neighbors (assumed by some distance metric, normally Euclidean distance). Specifically, the algorithm locates the k nearest instances to the query instance and determines its class by identifying the single most frequent class label, i.e., a new instance is classified by a majority vote of its neighbors (Wu, 2015; Vaish, 2016; Kotsiantis et al., 2006). In Figure 4.7, an example can be found.

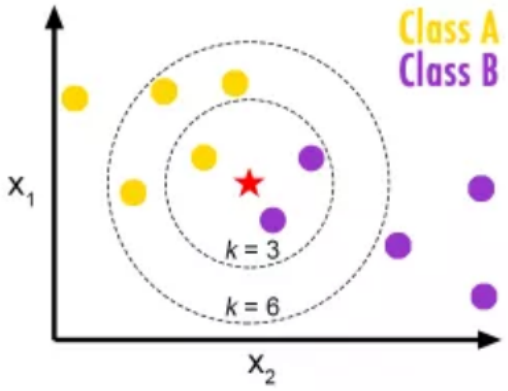


Figure 4.7: Example of a k-NN algorithm. The unlabeled object (represented by a star) is assigned to class B if k=3 and to class A if k=6 (Vaish, 2016).

In order to obtain more accurate results, some algorithms use weighting schemes which alter the distance metrics and voting influence of each instance (Kotsiantis et al., 2006).

k-NN is one of the easiest classifiers to understand and implement as its algorithm is straightforward and intuitive, given that it only requires a given integer number of neighbours k, a training dataset and a distance metric for decision. . However, it has some limitations, such as having large storage requirements and lacking an established way to choose k. This choice is normally done through cross-validation or a similar computationally expensive technique (Wu, 2015).

4.5.3 Support vector machine

SVMs revolve around the notion of a "margin" – either side of a hyperplane that separates two data classes. Maximizing the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it is SVM's main goal. If data is linearly separable, once the optimum separating hyperplane is found, data points that lie on its margin are known as support vector points and the solution is represented as a linear combination of only these points (Hearst et al., 1998).

Therefore, the model complexity of an SVM is unaffected by the number of features selected from the training data. So, one of the advantages of SVM is the good performance results obtained in learning tasks where the number of features is large with respect to the number of training data points (Steinwart and Christmann, 2008).

However, the SVM may not be able to find any separating hyperplane for the dataset. This can be dealt with by using a soft margin (by introducing a set of slack variables) which accepts some misclassifications of the training instances.

After this, data points may remain non-separable for which no separable hyperplane exists, i.e., it is not possible to separate the positive from negative instances in the training set. Thus, data can be mapped onto a higher-dimensional space where the hyperplane will be defined through a nonlinear kernel function (polynomial, gaussian or radial basis) (Hearst et al., 1998; Steinwart and Christmann, 2008).

A kernel function $\kappa(\cdot)$ is defined as the inner product of the nonlinear mapping operator $\varphi(f)$ as

$$\kappa(f_j, f_k) = \varphi(f_j)^T \varphi(f_k). \tag{4.31}$$

The selection of an appropriate function is important, since it will define the feature space in which the training set will be classified. It is common to use cross-validation over the training set to find the best function. For this, a limitation of SVMs is the low speed of the training (Wu, 2015; Kotsiantis et al., 2006).

Given a set of training data f_n, y_n, with the class label $y_n \in -1, +1$ and a set of slack variables $\xi_n \geq 0$, the boundaries of the margin (see Figure 4.8) for a binary classification can be expressed as

$$w^T \varphi(f_n) + b \geq 1 - \xi_n, y_n = +1, \tag{4.32}$$

$$w^T \varphi(f_n) + b \leq -1 + \xi_n, y_n = -1. \tag{4.33}$$

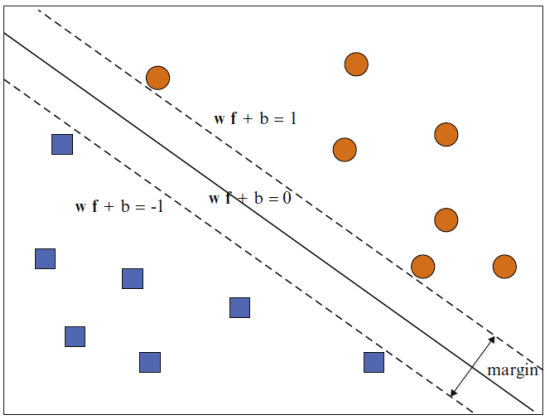


Figure 4.8: Decision boundary margin of a support vector machine. The blue samples have label $y_n = -1$ and the orange samples $y_n = 1$ (Wu, 2015).

4.5.4 Artificial neural network

ANN is a perceptron-based classification technique. The perceptron algorithm computes the sum of weighted input $\sum_i x_i w_i$, knowing that x_1 through x_n are input feature values and w_1 through w_n are connection weights/prediction vector (typically real numbers in the interval [-1 1]). Then, the weighted sum goes through an adjustable threshold, which returns 1 as output if it is above the threshold or 0 if it is below. The perceptron algorithm is run repeatedly through the training set until it finds a prediction vector which is correct for all the instances. This prediction rule is then used for predicting the labels on the test set (Freund and Schapire, 1999).

However, perceptrons can only classify linearly separable sets of instances. Thus, ANN arose to try to solve this problem.

Multi-layer neural networks have been widely used in practical applications, which consist on a large number of units (neurons) joined together in a pattern of connections (Figure 4.9). These neurons are usually organized in one input layer which receive information to be processed, one or more hidden layers that process the information and one output layer where the results of the processing are found (Hopfield, 1988).

Generally, it is difficult to determine the number of adequate hidden layers. Underestimation can lead to poor approximation and generalization capabilities and an excessive number can result in overfitting and eventually make the search for the global optimum more difficult (Ripley, 2007).

ANN is dependent on three important aspects: input and activation functions of the unit, network architecture and weight of each input connection. The first two aspects are fixed, so the

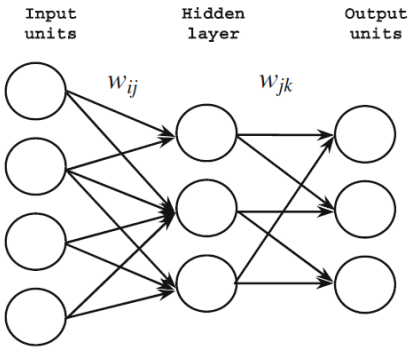


Figure 4.9: Feed-forward ANN, which allows signals to travel one way only, from input to output. w_{ij} represents the connection weight from input unit i to hidden layer j, whereas w_{jk} is the connection weight from hidden layer j to output unit k (Kotsiantis et al., 2006).

behavior of the network is defined by the weight values. Initially, weights are set to random values which are repeatedly adjusted after the output of the network is compared with the desired output for each instance (Ripley, 2007; Khan et al., 2001).

There are several algorithms with which a network can be trained, but the best known and widely used is the Back Propagation (BP) algorithm, whose general rule for updating weights is $\Delta W_{ji} = \alpha \delta_j O_i$, where

- α is a positive number, called the learning rate, that determines the step size in the gradient descent search. A large value makes back propagation to move faster to the target weight configuration, but it also increases the chance of never reaching this target.
- O_i is the output computed by neuron i.
- $\delta_j = O_j(1-O_j)(T_j - O_j)$ for the output neurons, where T_j is the target output for the neuron j.
- $\delta_j = O_j(1-O_j)\sum_k \delta_k W_k j$ for the hidden neurons.

The signal at the input units propagates all the way through the net to determine the activation values at the output units. Firstly, every input unit, which has an activation value representing some feature external to the net, sends this value to each of the hidden units. Each of these hidden units computes its own activation value (according to an activation function), which is passed on to output units. This activation function sums the contributions of all sending units, where the contribution of a unit is defined as the weight of the connection between the sending and receiving units multiplied by the sending unit's activation value. This sum is further modified, adjusting it to a value between 0 and 1 and/or by setting the activation value to 0 unless a threshold level for that sum is reached (Hecht-Nielsen, 1992; Zhang et al., 2007).

4.5.4.1 Radial basis function network

RBFNs are another type of networks that have been widely applied in data mining, especially in the analysis of VAG signal patterns. They typically contain only one hidden layer, as represented in Figure 4.10.

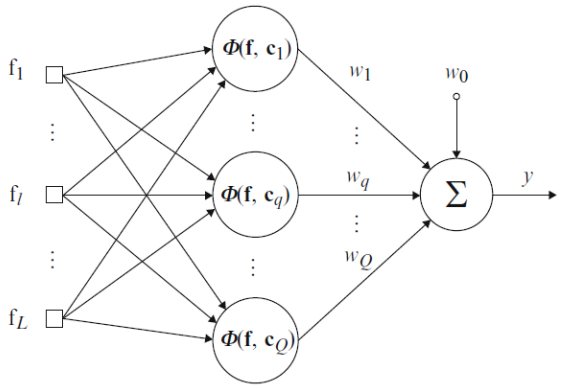


Figure 4.10: Structure of a RBFN. f_l represents input feature l, $\Phi(f,c_q)$ is the radial basis kernel function of the qth hidden node with c_q as center vector and w_q is the weight connecting the qth hidden node and the network output (Wu, 2015).

The training procedure can be divided in two stages. Firstly, the centers and widths of the hidden layer are determined by clustering algorithms. Then, the weights connecting the hidden layer with the output layer are determined by Singular Value Decomposition (SVD) or Least Mean Squared (LMS) algorithms (Vt and Shin, 1994).

Given that the training data set consists of N signals, each of which is represented with the pairs of L-dimensional input features $f^n = [f_1^n, f_2^n, ..., f_L^n]^T$ and desired class label y_d^n, in the form of $\left\{f^n, y_d^n\right\}_{n=1}^N$, the RBFN with Q hidden neurons produces the output response as

$$y^n = \sum_{q=1}^{Q} w_q \phi(f^n, c_q) + w_0, \tag{4.34}$$

where $\phi(\cdot)$ represents the radial basis kernel function, c_q is the center vector for the qth hidden node, w_q denotes the weight that connects the qth hidden node and the network output, and w_0 is the bias (Schilling et al., 2001). The RBFN centers are selected from the training dataset and is an important factor that may affect the classification performance.

4.5.5 Bayesian network

A BN is a statistical multivariate model for a set of variables $X = X_1, \ldots, X_n$, which is defined in terms of two components:

- Qualitative component (see Figure 4.11): a directed acyclic graph (DAG) where each vertex represents one of the variables in the model, and so the presence of an edge linking two variables indicates the existence of statistical dependence between them.

- Quantitative component: a conditional distribution $p(x_i|pa(x_i))$ for each variable $x_i, i = 1,\ldots,n$ given its parents in the graph denoted as $pa(x_i)$.

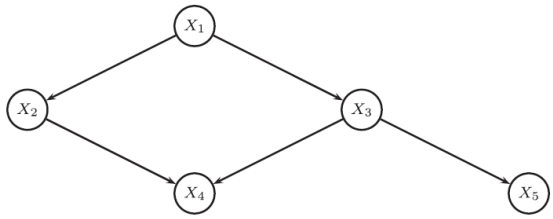


Figure 4.11: Bayesian network with five variables. This graph represents the qualitative component of the network for variables $x_1,\ldots,x_5$. In this case, the distributions are $p(x_1), p(x_2|x_1), p(x_3|x_1), p(x_4|x_2,x_3)$ and $p(x_5|x_3)$ (Aguilera et al., 2011).

Given a training set with features $x_1, x_2, \ldots x_n$, a BN classifier returns the label c which maximizes the posterior probability $p(c|x_1,x_2,\ldots,x_n)$ (Friedman et al., 1997).

One advantage of a Bayesian network is that the structure of the DAG determines the dependence and independence relationships among the variables. Therefore, it is possible to see which variables are relevant/irrelevant for some other variable of interest without numerical calculations. However, they have an inherent limitation, which is the computational difficulty of exploring a previously unknown network. Also, they are not suitable for datasets with many features (Aguilera et al., 2011; Friedman et al., 1997).

4.5.5.1 Naïve Bayes

NB networks are very simple Bayesian networks which are composed of DAGs with only one parent (representing the unobserved node) and several children (corresponding to observed nodes), with a strong assumption of independence among child nodes in the context of their parent. Therefore, this model is based on the estimation of R as

$$R = \frac{P(i|x)}{P(j|x)} = \frac{P(i)P(x|i)}{P(j)P(x|j)} = \frac{P(i)\prod_{r=1}^{n} P(x_r|i)}{P(j)\prod_{r=1}^{n} P(x_r|j)}, \tag{4.35}$$

where n is the number of features. Comparing these two probabilities, the larger probability indicates the class label value that is more likely to be the true label for the training feature set x (if $R > 1$, the prediction is i, otherwise is j) (Zhang, 2004).

Naive Bayes classifiers are usually less accurate than other learning algorithms, such as ANNs, since the assumption of independence among child nodes is frequently wrong. However, it has the advantage of having short computational time for training.

4.5.6 Fisher's linear discriminant analysis

FLDA is a simple but effective pattern classification tool that searches for a mapping orientation which leads to the best separation among classes. It projects the multidimensional data onto a

linear surface such that the dimensionality of the complex dataset is reduced (Wu, 2015).

Thus, the main goal of the FLDA algorithm is to seek a linear combination of features that leads to the maximization of the discriminant function

$$J(w) = \frac{w^T S_B w}{w^T S_W w}, \tag{4.36}$$

where S_B represents the inter-class scatter matrix and S_W denotes the intra-class scatter matrix, being defined as

$$S_B = (m_N - m_A)(m_N - m_A)^T, \tag{4.37}$$

$$S_W = \sum_{\mu \in \omega_N} (\mu - m_N)(\mu - m_N)^T + \sum_{\mu \in \omega_A} (\mu - m_A)(\mu - m_A)^T, \tag{4.38}$$

where m_N and m_A denote the mean values of μ for the normal and abnormal signal groups, respectively.

4.5.7 Combined classifiers

Combining classifiers is an active area in supervised learning. There are several methods that have been suggested for the creation of a good ensemble of classifiers, which include: using different subsets of training data with a single learning method; using different training parameters with a single training method (e.g. using different initial weights for each neural network in an ensemble); using different learning methods (Kuncheva and Rodríguez, 2014).

4.5.7.1 Different subsets of training data with a single learning method

Bagging is an example of this, which draws e random subsets from a training set of size n with replacement (i.e., using a uniform distribution). These e subsets are learned, and this process is repeated several times. Thus, a classifier is generated in each cycle through the process. Then, the final prediction is generated taking a vote of the prediction of each built classifier (Figure 4.12).

Another method is Boosting, which is structurally similar to Bagging, but it keeps track of the performance of the model and focuses on instances that have not been correctly learned. Therefore, instead of choosing the e training subsets randomly, it chooses them in a way to favor the instances that have not been accurately learned. Several cycles occur, and the prediction is computed by taking a weighted vote of the predictions of each classifier, the weights being proportional to each classifier's accuracy on its training set. AdaBoost is a practical version of the boosting approach (Zhang et al., 2014; Pérez-Ortiz et al., 2016).

The number of sub-classifiers is an important variable to decide in practical applications, keeping in mind that the algorithms should run in reasonable time. Boosting has one big problem: robustness to noise, which is due to the fact that noisy instances tend to be misclassified, increasing their weight for the next iteration (Kuncheva and Rodríguez, 2014).

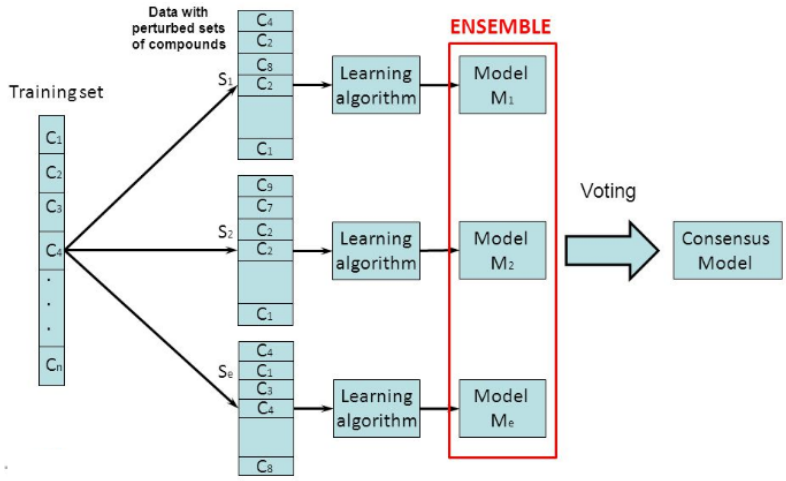


Figure 4.12: Bagging approach schema (of Strasbourg, 2018).

4.5.7.2 Different training parameters with a single training method

These methods generate classifiers, altering their training process, which produce different predictions. For instance, neural networks techniques that have been employed include methods for training with different topologies, different initial weights and different parameters.

Ensemble feature selection is another technique which generate a set of base classifiers. It consists on finding a set of feature subsets for generation of the base classifiers for an ensemble with one learning algorithm (Kuncheva and Rodríguez, 2014).

4.5.7.3 Different learning methods

The simplest method of combining predictions from multiple classifiers is majority voting, in which each learning method contributes a single vote. Thus, the final prediction is decided by the majority of the votes, that is, the class with the most votes constitutes the final prediction.

On the other hand, in weighted voting, the influence of the classifiers on the collective prediction varies, which is relative to their predictive accuracy. Therefore, each classifier has an associated weight, that is determined by its performance on a validation set. The final prediction is then decided by summing up all weighted votes and by choosing the class with the highest aggregate (Kuncheva and Rodríguez, 2014; Zhang et al., 2014).

4.6 Classifier evaluation

Having so many options for learning algorithms, the choice of the most appropriate is a critical step. There are some methods to accomplish this purpose that share the same principle: to build a model in a subset of data, called the training set, and to test it in an unseen subset of data, called the test set.

One method consists in splitting the training set into a training subset and a validation subset that is used to estimate the classifier's performance. Normally, two-thirds of the samples are used for training and the remaining one-third for validation. Other method is named cross-validation, in which the training set is divided into mutually exclusive and equal-sized subsets. Iteratively, one of these subsets is used for validation, while the classifier is trained on the union of all the other subsets. Leave-one-out validation is a particular case of cross validation, in which all validation subsets consist of a single instance. This last type of validation is, obviously, more computationally expensive, but useful when the most accurate estimate of a classifier's error rate is required (Sokolova and Lapalme, 2009).

Generally, the classifier's performance is computed knowing the amount of correctly and incorrectly predicted samples for each class, the result being displayed in a so-called confusion matrix (Table 4.1).

Table 4.1: Confusion matrix for a binary classification

		Predicted	
		positive	negative
Ground Truth	positive	TP	FN
	negative	FP	TN

The true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are defined as

- TP: number of samples predicted as positive, whose true class is also positive.
- TN: number of samples predicted as negative, whose true class is also negative.
- FP: number of samples predicted as positive, whose true class is negative.
- FN: number of samples predicted as negative, whose true class is positive.

Having calculated these variables, it is now possible to compute some performance measures. The most important is the accuracy, which evaluates the overall performance of the classifier by the percentage of correct predictions. Sensitivity or recall corresponds to the proportion of samples belonging to the positive class which were correctly predicted as positive. Precision computes the proportion of positive incidents that were correctly predicted from the total predicted positive samples. Specificity corresponds to the percentage of negative samples that were correctly predicted as negative. F-measure can also be determined, which represents the weighted average of the sensitivity and precision (Sokolova and Lapalme, 2009; Ferri et al., 2009).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{4.39}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{4.40}$$

$$Precision = \frac{TP}{TP+FP} \tag{4.41}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4.42}$$

$$F-measure = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity} \tag{4.43}$$

Another tool for binary classification performance evaluation is the receiver operating characteristic (ROC) curve. This is established by computing the fraction of true positive rate (TPR), which is the same as sensitivity, over the fraction of false positive rate (FPR), which corresponds to $1 - specificity$, at different threshold levels. The area under the ROC curve (AUC or A_z), which is the performance measurement used, represents the probability that a classifier ranks a randomly chosen positive instance higher than a randomly chosen negative one according to a threshold parameter (Bradley, 1997).

4.7 Literature review

Over the years, several studies have implemented different classification methods on features extracted from different VAG datasets. Obtaining a high AUC and accuracy value is, indeed, the ultimate goal of a classification system.

Rangayyan and Wu (2008, 2009a, 2010) have published some studies, which differ on the statistical features extracted. Firstly, the form factor (FF), the skewness (SK), kurtosis (KU) and entropy (H) were computed using the PDF obtained through the normalization of the histogram (Rangayyan and Wu, 2008). Both FLDA and RBFN classifiers were trained and values of 0.72 and 0.82 for the AUC were obtained. Then, new statistical features (TC and VMS) were extracted from the same dataset, but considering the average and the standard deviation of the signal this time, and not its PDF. The same classifiers were used, and the AUC values were better for the RBFN (0.92), but not for the FLDA (0.65) (Rangayyan and Wu, 2009a). Finally, the average, standard deviation, SK, KU and H were obtained through the PDF which was estimated using Parzen windows. However, the classifiers' performance did not improve, with AUC values of 0.63 and 0.83 for the FLDA and RBFN respectively (Rangayyan and Wu, 2010).

Kim et al. (2008, 2009b,c,a) have also published a few relevant studies over the years, using the same dataset. The main differences between Kim's studies is the VAG signal processing and the parameters used for the BPNN classifier, namely the number of nodes in the three layers of the network and the learning rate. The four time-frequency computed parameters were the energy parameter (EP), energy spread parameter (ESP), frequency parameter (FP) and frequency spread parameter (FSP). The best value obtained for the accuracy was 91.4%.

Nalband et al. (2016) used the same dataset in his study as Rangayyan. However, instead of computing statistical parameters, nonlinear features such as recurrence quantification analysis

(RQA), approximate entropy (ApEn) and sample entropy (SampEn) were extracted as features of the VAG signals.

RQA requires no assumption about the linearity or stationary about the data for analysing nonlinear signals. Three different features are computed, based on recurrence plots, which are binary symmetric of NxN arrays in which a spot is placed at (i, j) when the distance between X_i and X_j is less than a prescribed value ε.

$$R_{i,j} = \begin{cases} 0, if \left\| X_i - X_j \right\| < \varepsilon \\ b, if \left\| X_i - X_j \right\| \geq \varepsilon \end{cases} \tag{4.44}$$

ApEn and SampEn measure the "likelihood that runs of patterns that are close remain close on next incremental comparisons". WPD is computed, so the energy in each of the sub-band signals is used as a feature (Nalband et al., 2016). With a SVM learning algorithm (Least-squares SVM) and a decision tree approach (Random Forest), accuracy values of 94.31% and 91.01% were obtained, respectively.

Once again, the same dataset was used by Cai et al. (2013). Some statistical parameters were extracted as features, namely, TC, FF, VMS and the signal's average. Cai brought an innovation, which was the use of a classifier ensemble, a Bagging of LS-SVMs, with a performance accuracy of 78.65%. However, a single LS-SVM gave better results, with an accuracy of 87.64%.

Wu et al. (2013) obtained an accuracy value of 81.33% using a SVM with a polynomial kernel. The extracted features were the VMS and FF.

Later, Rangayyan et al. (2013) computed the fractal dimension for different segments of the VAG signal, making a total of 7 features. Fractals are based on geometric structures that have self-similiarity at different scales of length. They lack any single scale of length and have a noninteger (fractional) dimension, referred to as the fractal dimension (FD). Fractal properties have been observed in VAG signal (Rangayyan et al., 2013). This innovative approach resulted in an accuracy value of 75.10% using a RBFN learning algorithm.

Liu et al. (2014) trained a k-NN model with two statistical parameters, the VMS and the FF. A performance accuracy of 80% was obtained, while for the FLDA algorithm, this value was 77%.

Kręcisz and Bączkowicz (2018) extracted statistical parameters (VMS, TC and FF), recurrence quantification analysis and the fractal dimension. Using a multilayer perceptron and a random forest as learning algorithms, accuracy values of 87.2% and 88.8% were obtained, respectively.

A summary of all these studies is presented in Table 4.2

Table 4.2: Comparison of different knee joint vibrational based classification systems reported in the literature.

Authors	Sensor	Dataset (Healthy-Pathological)	Features	Classification model	Performance
Rangayyan and Wu (2008, 2009a, 2010)	Ax, Stethoscope	51-38	Statistical Parameters	FLDA, RBFN	AUC: 0.72, 0.92
Kim et al. (2008, 2009b,c,a)	Electrostethoscope, Electrogoniometer	20-11	TF Features	BPNN	Acc: 91.4%
Nalband et al. (2016)	Ax	51-38	Recurrence quantification features, sample and approximate entropy, wavelet based energy	LS-SVM, Random Forest	Acc: 94.31%, 91.01%
Cai et al. (2013)	Ax, Stethoscope	51-38	Statistical parameters	LS-SVM, Bagging of LS-SVMs	Acc: 87.64%, 78.65%
Wu et al. (2013)	Ax, Stethoscope	47-28	Statistical parameters	SVM	Acc: 81.33%
Rangayyan et al. (2013)	Ax	51-38	Fractal analysis	FLDA	Acc: 75.10%
Liu et al. (2014)	Ax, Stethoscope	47-28	Statistical parameters	k-NN, FLDA	Acc: 80%, 77%
Kręcisz and Bączkowicz (2018)	Ax	66-121	Statistical parameters, Recurrence quantification features, Fractal analysis	Perceptron, Random Forest	Acc: 88.8%, 87.2%

Ax - accelerometer; TF - Time Frequency; Acc - Accuracy

Chapter 5

Analysis using open-source VAG signals dataset

The dataset consists of 89 signals, with 51 from normal volunteers (22 male, 29 female, age 28± 9.5 years) and 38 from subjects with knee-joint pathology (20 male, 28 female, age 35± 13.8 years). The normals (control) were established by clinical examination and history. The abnormals were collected from symptomatic patients scheduled to undergo arthroscopy. These include chondromalacia of different grades at the patella, meniscal tear, tibia chondromalacia, and anterior cruciate ligament injuries, as confirmed during arthroscopic examination.

This dataset does not permit classification of the signals into various types or stages of pathology. Therefore, it can only be used for screening purposes, i.e., classification of the signals and the corresponding knee joints as normal or abnormal.

Subjects were seated on a rigid table with both of their legs suspended. In order to detect the VAG signal, a Dytran accelerometer (Dytran, 2018) (model 3115a) was placed on the surface of the skin at the mid-patella position of the knee, and the signal was recorded on a Hewlett Packard (HP, 2018) instrumentation recorder during swinging movement of the leg from 90º (full flexion) to 0º (full extension) to 90º in a total time period of approximately 4s. An electrogoniometer was placed on the lateral side of the knee to measure the angle of the limb (Rangayyan et al., 1997).

Although the angle measurements were performed using an electrogoniometer, this data is not available in the open-source dataset in Rangayyan (2006). Thus, the segmentation of the signals in the corresponding ascending/descending movements is not possible.

In Figure 5.1, it is possible to see that the abnormal signal exhibits a higher degree of overall variability and complexity than the normal signal.

The whole signal analysis approach (Figure 5.2) from the signal acquisition until the final binomial classification will be described in the next sections.

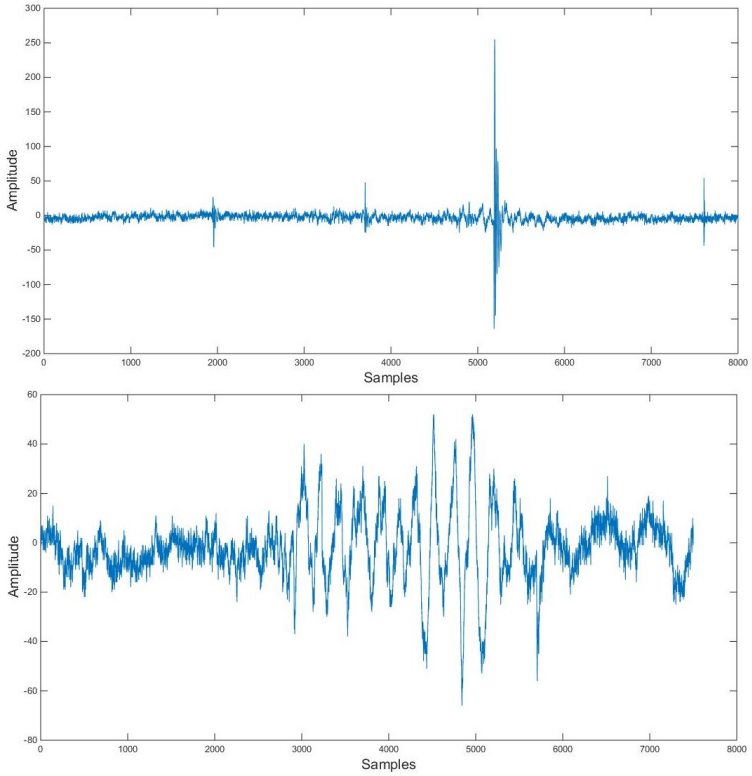


Figure 5.1: Example of a normal (top) and abnormal (bottom) VAG signal.

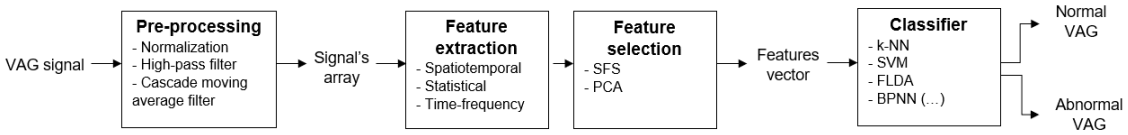


Figure 5.2: Approach for the VAG signal analysis.

5.1 Pre-processing

5.1.1 Normalization

Amplitude normalization was applied to all raw signals, in order to reduce the effect of the discrepancies observed on the signal's amplitude across subjects. Therefore, the normalized amplitude was bounded to the interval [-1,1] using the min-max normalization approach, already described in Section 4.2.1.

In Figure 5.3, it is possible to see that both normal and abnormal signals' original amplitudes were bounded to the interval [-1,1] as pretended, keeping the shape of the waveform, as well as its characteristics. Analysing the normalized histograms, it is evident that the abnormal signal has a broader range of values than the normal signal.

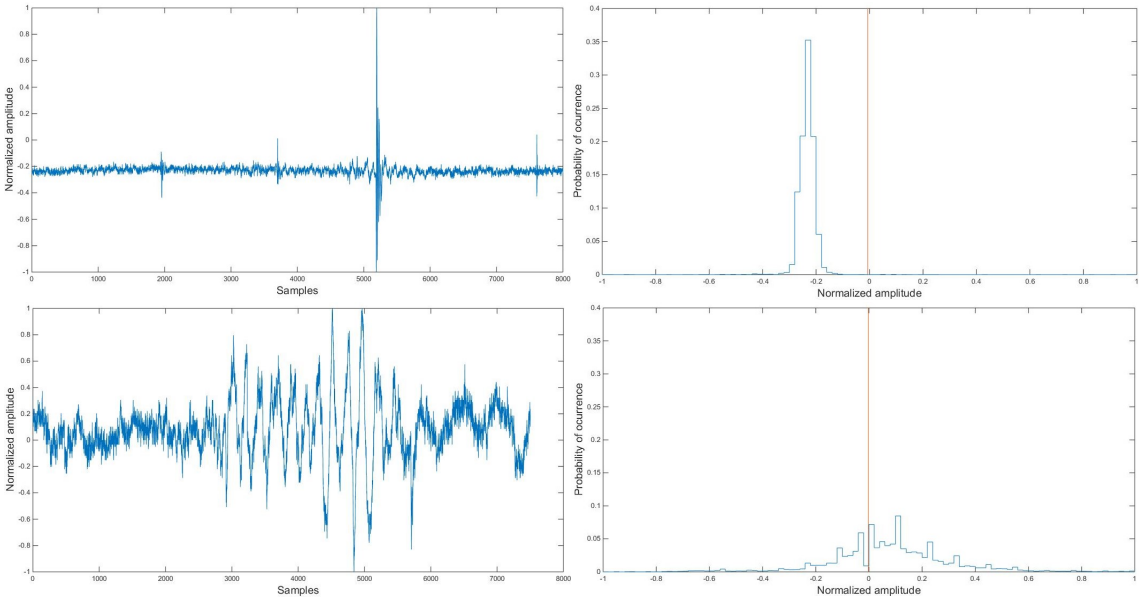


Figure 5.3: Amplitude range normalization of a normal (top) and abnormal (bottom) VAG signal. The normalized histogram (estimated with L=100 bins) is also represented at the right of each signal, where the mean of each signal is shown in orange.

5.1.2 High-pass filtering

A high-pass filter was applied, in order to remove unwanted parts of the signal (such as random noise or muscular interference) that hide the important information that is present in the signal.

The applied filter was a 9th order butterworth high-pass filter with a cut-off frequency of 20Hz. The value always used in the reported literature (10Hz) was tested, but 20Hz lead to better performance results. The butterworth filter was chosen over other filters such as chebyshev, since it is widely used with biomedical signals. Its optimal flatness in the pass-band and rapid response in the time domain are optimal for smoothing movement data (Robertson and Dowling, 2003).

In Figure 5.4 it is noticeable that artefacts arising from movement related issues or muscular interference are successfully removed. Therefore, the obtained signal represents the pure specific knee joint vibrations emitted during the movement.

5.1.3 Baseline wandering removal

Baseline wandering estimation and subsequent removal from the VAG signal was accomplished via adaptive filtering, more specifically, by using a cascade moving average filter (CMA), which was described and explained in Section 4.2.2. According to the literature, the two moving average operators in the first layer of the hierarchical model were configured with equal orders, i.e., M = 20 and N = 20, respectively, due to the symmetry of the leg movement angles (90º-0º-90º). The number of the overlapping inputs of the moving average operators was set to be K = 5 (Cai et al., 2012; Wu, 2015).

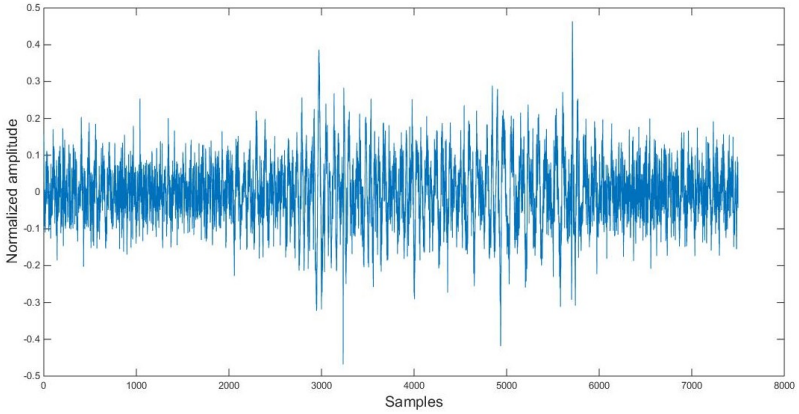


Figure 5.4: High-pass filtered version of the normalized abnormal signal in Figure 5.3.

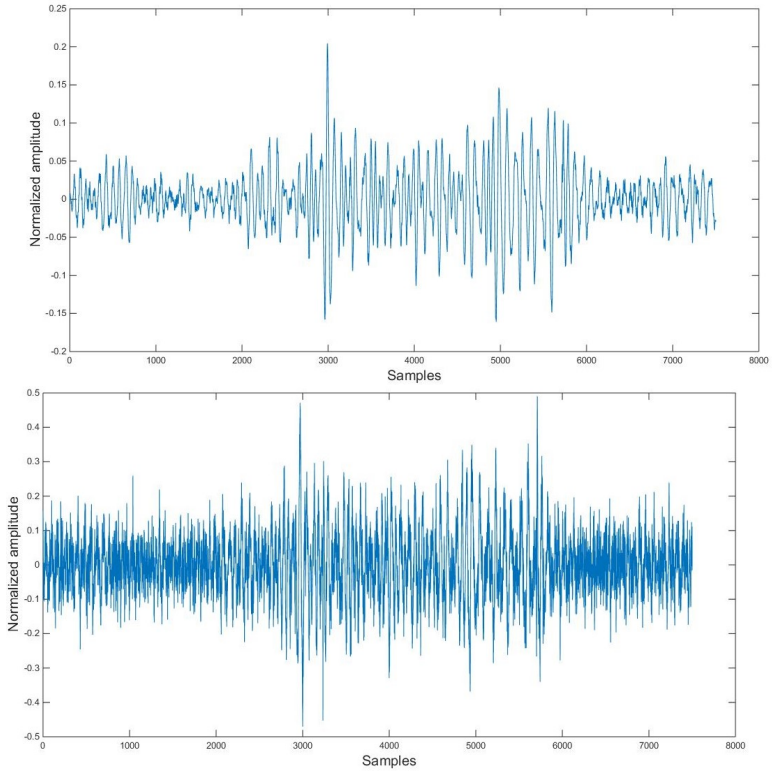


Figure 5.5: The algorithm estimates the mean component (top) of the abnormal signal in Figure 5.3, the bottom signal being the final output given by the cascade moving average filter.

Therefore, the transfer function ($H(z)$) of the CMA can be represented as

$$H(z) = \frac{1}{40}(z^{-1} + ... + z^{-20}) + \frac{1}{40}(z^{-16} + ... + z^{-35}). \tag{5.1}$$

Figure 5.5 represents the application of the filter, which was able to estimate and remove the baseline wandering of the signal. This artifact can mask some important features of the VAG signals, so its removal is important (Kaur et al., 2011).

5.2 Feature extraction and selection

A classification system requires a set of highly discriminant features that characterize the VAG signals. A set of time-domain (namely spatiotemporal and statistical) and time-frequency-domain features were extracted from the provided signals. While the former are mostly based on simple mathematical and statistical metrics which characterize the signal's shape and basic information, the latter contain information from repetitive events that occur over time.

The time-frequency-domain features are important to extract, due to the non-stationary property of the VAG signals. The most recent and accurate approach to perform this analysis is the wavelet packets transform.

5.2.1 Signals' array

In order to increase the number of features that may help in discriminating between the two classes of signals, normal and abnormal, an array of signals was created. This was done because it was observed that some features were more discriminating while extracted from signals that were not subjected to the fully described pre-processing approach (see Appendix B.1).

This array is constituted by seven different signals, which differ in the pre-processing techniques used. Therefore, it contains the normalized signals (VAG_{norm}), the high-pass filtered signals (VAG_{hp}), the CMA filtered signals (VAG_{cma}), the normalized high-pass filtered signals ($VAG_{norm-hp}$), the normalized CMA filtered signals ($VAG_{norm-cma}$), the high-pass and CMA filtered signals (VAG_{hp-cma}) and the CMA high-pass filtered normalized signals ($VAG_{norm-hp-cma}$).

For each signal a total of 126 features were extracted. Thus, a total of 882 features were extracted fom the 7-element signal's array.

5.2.2 Spatiotemporal features

Each one of the following items constitutes one feature that characterizes the signal.

- Form Factor (ST_{FF}): measures the complexity of the signal.
- Variance of the mean-squared values (ST_{VMS}): measures the local variation of the signal, by estimating the average on fixed intervals (5 ms) and computes its variance over the entire duration of the signal.
- Turns Count (ST_{TC}): computes the number of significant "turns" over a certain threshold ($th = 0.5\sigma$).

VAG signals recorded from the symptomatic patients with knee joint pathologies commonly exhibit a higher level of signal fluctuations in the temporal scale. Therefore, it is expected that the values of FF and VMS are larger for the abnormal signals than for the normal ones (see Table 5.1).

Because of the use of an adaptive threshold proportional to the standard deviation of the signal being processed, the abnormal signals have lower TC values than the normal signals. The adaptive

Table 5.1: Mean and standard deviation of the spatiotemporal features for the normal (51) and abnormal (38) VAG signals. The first and second row represent, respectively, the values for the $VAG_{norm-hp-cma}$ and the VAG_{norm} signals.

	Normal	**Abnormal**
$\mathbf{ST_{FF}}$	1.51 ± 0.38 2.39 ± 1.41	1.93 ± 0.74 3.74 ± 2.71
$\mathbf{ST_{VMS}}$	3.74E-4 $\pm$ 8.26E-4 4.30E-3$\pm$ 5.10E-3	4.25E-4 $\pm$ 7.49E-4 9.20E-3$\pm$ 1.06E-2
$\mathbf{ST_{TC}}$	532 ± 438 453 ± 380	397 ± 443 342 ± 362

threshold was preferred over the fixed one ($th = 0.2$), since the latter led to poor discrimination between normal and abnormal VAG signals. This could be due to variations in the signal acquisition procedure and normalization of the signal amplitude.

5.2.3 Statistical features

The histogram was estimated with $L = 100$ bins, according to Rangayyan and Wu (2010). Each one of the following items represents one feature and was calculated based on the PDF of the signal.

- Average (SS_{AVG}): estimates the "central" value for the signal's amplitude.
- Standard deviation (SS_{STD}): measures the spreadability of the signal.
- Variance (SS_{VAR}): also measures the spreadibility of the signal, being the squared value of the standard deviation.
- Skewness (SS_{SK}): quantifies the asymmetry of the PDF of a signal about its mean.
- Kurtosis (SS_{KU}): determines the tendency of the PDF to have peaks (in relation to a normal distribution).
- Entropy (SS_H): assesses the randomness observed in the PDF.

It is evident that the abnormal signal exhibits a higher level of signal fluctuations (with higher amplitudes) in the temporal scale. Thus, the values of μ, σ and σ^2 are larger for the abnormal signals than for the normal ones, except for the variance of the VAG_{norm} signals, whose values are similiar.

Skewness measures the symmetry of the PDF about its mean. If the PDF is symmetric or balanced, the skewness is zero; if the mass of the PDF is concentrated on the left of the mean, the skewness is positive (SK>0); in contrast, if the mass of the PDF concentrates on the right of the mean, the skewness is negative (SK<0). Analysing the signal in the time domain, an equal number of large and small amplitude values leads to a skewness of zero. A time series with many

Table 5.2: Mean and standard deviation of the statistical features for the normal (51) and abnormal (38) VAG signals. The first and second row represent, respectively, the values for the $VAG_{norm-hp-cma}$ and the VAG_{norm} signals.

	Normal	**Abnormal**
$\mathbf{SS_{AVG}}$	1.30E-3±2.06E-2 -1.05±8.44	6.90E-3±9.20E-3 6.80± 7.69
$\mathbf{SS_{STD}}$	3.93±2.22 275±283	4.95±2.63 305±224
$\mathbf{SS_{VAR}}$	20.3±20.0 1.54E5± 3.10E5	31.2±27.2 1.42E5± 1.87E5
$\mathbf{SS_{SK}}$	-3.24E-2± 0.193 2.37E-2± 0.151	9.00E-3± 0.122 -8.02E-2± 0.131
$\mathbf{SS_{KU}}$	0.730 ± 0.954 2.46 E-2± 1.04E-2	0.526 ± 0.882 2.34E-2± 3.90E-3
$\mathbf{SS_{H}}$	-87.8 ± 58.4 -46.4 ± 50.9	-100 ± 67.4 -24.6 ± 39.8

small values and few large values is positively skewed and the skewness value is positive. A time series with many large values and few small values is negatively skewed and the skewness value is negative (Wu, 2015; NationalInstruments, 2018).

For the VAG_{norm} signal, the SK value for the normal signal is positive, as can be seen in Figure 5.3, due to the deviation of the PDF to the left of the mean and the small amount of large values. On the other hand, the SK value for the abnormal signal is justifiably negative. However, the opposite happens for the $VAG_{norm-hp-cma}$ signals. In order to understand the effect of signal processing on the skewness values, their mean for all the normal and abnormal signals are presented in Table 5.3 for the 7 elements of the signal's array.

Table 5.3: Mean of the skewness values for the normal and abnormal signals of the 7 elements signal's array.

	Normal	**Abnormal**
VAG_{norm}	2.37E-2	-8.02E-2
$VAG_{norm-hp}$	2.31E-2	1.44E-2
$VAG_{norm-hp-cma}$	3.24E-2	9.00E-3
VAG_{hp}	0.220	0.189
VAG_{cma}	1.10	-0.463
VAG_{hp-cma}	0.751	0.219
$VAG_{norm-cma}$	9.05E-2	6.33E-2

It is possible to conclude that for the VAG_{norm}, $VAG_{norm-cma}$ and VAG_{cma} signals, SK>0 for the normal and SK<0 for the abnormal signals. On the other hand, when high-pass filtering is applied, the opposite scenario happens, with normal signals having SK<0 and abnormal SK>0 (see Appendix B.2).

As expected, kurtosis is higher for the normal signals, since it represents the "peakedness" of the PDF (also represented in Figure 5.3). The entropy is closer to 0 for an uniform PDF, and has lower negative values for PDFs with narrow ranges of significant probability values (Wu, 2015). Since the PDF of the normal signals often have a peak shape, it was expected that they would have a lower entropy value, which happens for the VAG_{norm} signals. For the $VAG_{norm-hp-cma}$ signals, similiar entropy values were obtained for both the normal and abnormal.

5.2.4 Time-frequency features

For the time-frequency analysis, the wavelet packet transform was used with 4 levels of decomposition and the Daubechies 2 (DB2) wavelet, resulting into a set of 16 (2^4) different frequency bands computed for all time instants, as represented in Figure 5.6. Since the sampling frequency is 2kHz, the first level of the WPT is 0-1kHz, with 1kHz corresponding to the Nyquist frequency (the highest frequency that the sampled signal can unambiguously represent) (Boersma, 1993).

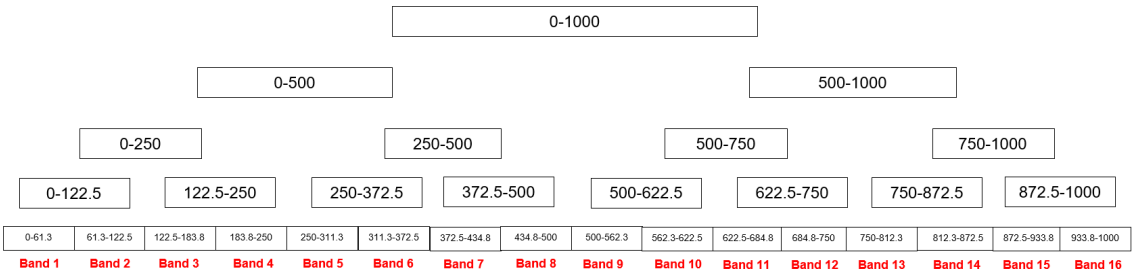


Figure 5.6: Wavelet packets decomposition scheme (16 frequency bands in Hz). Each frequency band has N coefficient values which characterize the behavior of the input of the signal at that specific frequency range over the entire time.

This type of wavelet was chosen, since it was reported to be the best type for vibroarthrography analysis due to its similarity with VAG signals (Wu, 2015). Also, the tree depth was chosen to be 4, due to the limited number of samples for each signal, which does not allow a more detailed analysis of the frequency content of each VAG signal. Thus, 16 equally spaced frequency intervals are analysed for each signal.

All the subsequent parameters are computed for each frequency band distribution over the entire time, so each parameter gives rise to 16 features for each signal.

- Mean wavelet coefficient power (TF_{mwc}): computes the mean of the wavelet coefficient values.
- Standard deviation of wavelet coefficients (TF_{swc}): measures the spreadability of the wavelet coefficient values.
- Variance of wavelet coefficients (TF_{vwc}): measures also the spreadability of the wavelet coefficient values.

- Skewness of wavelet coefficients (TF_{skew}): estimates the asymmetry of the wavelet coefficient distribution.
- Kurtosis of wavelet coefficients (TF_{kur}): assesses the tendency of the wavelet coefficient distribution to have peaks over time.
- Number of significant differences (above a given threshold) observed between successive wavelet coefficients (absolute value). Two types of threshold were used, namely the mean (TF_{mth}) and the standard deviation (TF_{sth}) of that wavelet coefficient difference distribution over time.

From Figure 5.7, it is possible to conclude that the most differentiating band between normal and abnormal signals is 16 for two features (TF_{swc}, TF_{vwc}) and 11 for other two features (TF_{kur}, TF_{sth}). The remaining features ($TF_{mwc}, TF_{skew}, TF_{mth}$) most differentiating bands are 15, 7 and 1, respectively.

The next features refer to the frequency band with most power over time. Each item represents one feature.

- Mean frequency band with most power (TF_{mp}): computes the frequency band which displayed greater power, by averaging the frequency band with most power at all time instants.
- Standard deviation of the band with most power (TF_{sp}): measures the spreadibility of the frequency band with most power over time.
- Number of times that the frequency band with most power is greater than a given threshold. Three types of threshold were used, namely the mean distribution of this band for each time instant (TF_{mpth}), the standard deviation of this distribution (TF_{spth}) and a combination of the previous two, i.e. mean+standard deviation (TF_{mspth}).

As expected, the standard deviation of the frequency band with most power is bigger for the abnormal signals, due to the wider spread of these signals in comparison to the normal (Table 5.4).

The remaining features (especially the last two: TF_{spth} and TF_{mspth}) have slightly bigger values for the abnormal signals, due to the fact that the threshold depends on the standard deviation value of the signals. As stated before, the abnormal signals have bigger standard deviation values, due to their spread profile.

5.2.5 Feature selection

Due to the large number of features extracted (882 features), a feature selection algorithm is required to both reduce computation time and improve classifier's performance, since only the most discriminant features will be used to train the model. To achieve this purpose, two approaches were used: the SFS (wrapper selection) preceded by a filter selection method and the PCA (dimensionality reduction) algorithm.

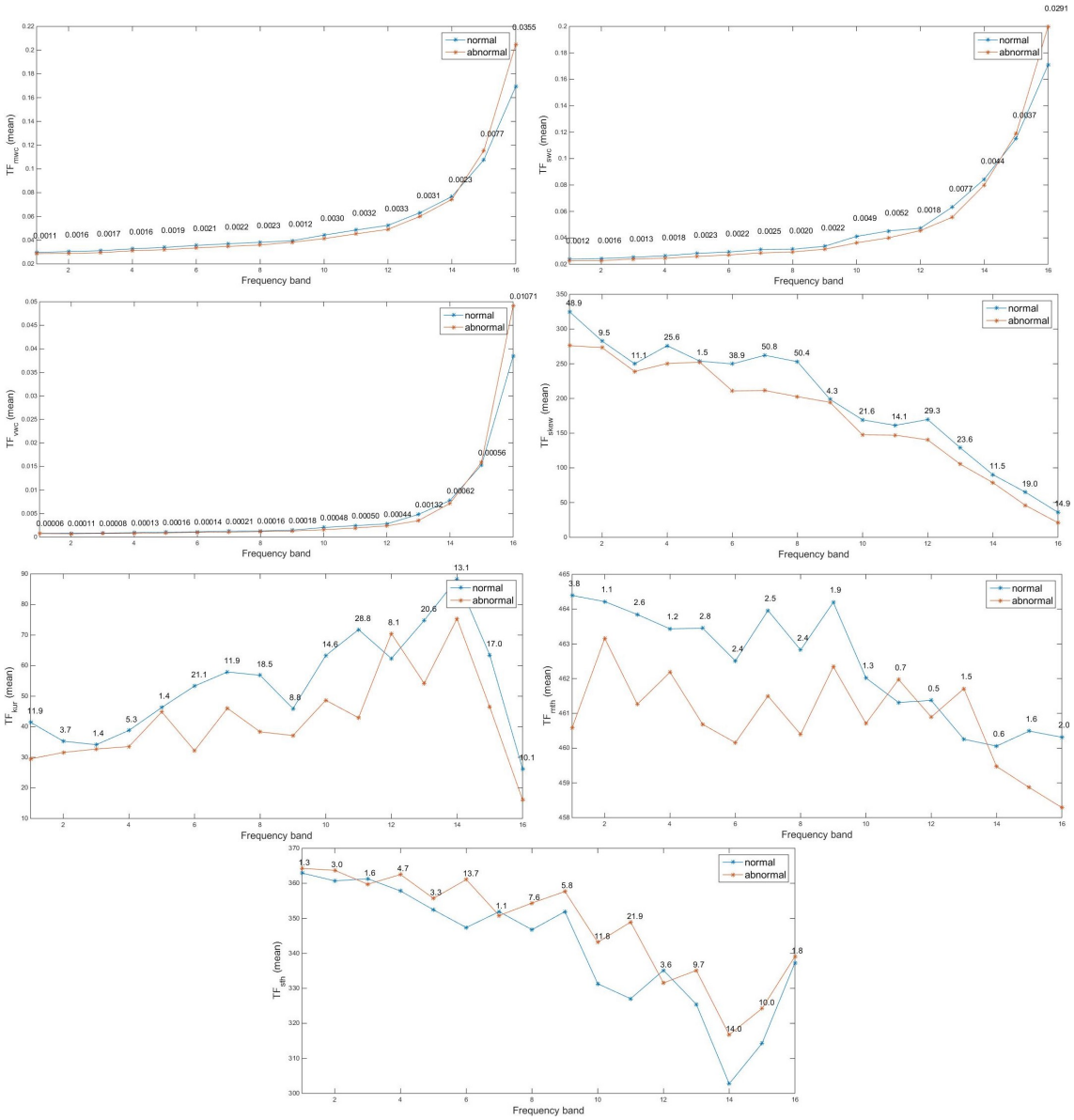


Figure 5.7: Mean (for normal and abnormal signals) of the time-frequency features TF_{mwc}, TF_{swc}, TF_{vwc}, TF_{skew}, TF_{kur}, TF_{mth}, TF_{sth} for the 16 frequency bands over time. The difference between the features' values are shown on the graph.

If PCA is the chosen method, the selected principal components do not depend on the classifier. The singular value decomposition (SVD) algorithm selected two principal components as the optimum number in order to keep 99% of variance in the binomial class data.

If a selection algorithm is preferred, a filter method is indispensable before a wrapper method, since the number of features is very large. Filters are computationally less expensive than wrappers, being applied at a first stage so that the more expensive wrapper methods only have to deal with the features that provide the shortest distance or statistical independence between classes. This is the basis of the most recent hybrid algorithms for feature selection, which try to take the strong points of both previous approaches (Sánchez-Maroño et al., 2007).

Table 5.4: Mean and standard deviation (except TF_{mp}, whose mode is represented) of the features related to the frequency band with most power for the normal (51) and abnormal (38) VAG signals. The first and second row represent, respectively, the values for the $VAG_{norm-hp-cma}$ and the VAG_{norm} signals.

	Normal	**Abnormal**
$\mathbf{TF_{mp}}$	14 16	15 16
$\mathbf{TF_{sp}}$	8.46E-2±6.89E-2 0.504±0.609	0.120±0.181 0.995±0.972
$\mathbf{TF_{mpth}}$	2.86E3±560 3.40E3±493	2.84E3±513 3.48E3±391
$\mathbf{TF_{spth}}$	2.51E3±1.42E3 4.89E3±2.27E3	2.74E3±1.28E3 5.18E3±1.71E3
$\mathbf{TF_{mspth}}$	697±383 953±334	806±329 1.05E3±251

In this case, the SFS selection method was preceded by a two-sample Student's *t*-test, which selected the most significantly different features for the normal and abnormal signal groups (Liu et al., 2014). Then, SFS selects a subset of features (from the remaining ones) that best predict the data (using a specific model) by sequentially adding features until there is no improvement in prediction (MATLAB, 2018c).

However, due to large computation times derived from the great amount of features, SFS was performed having k-NN (k=3) and SVM (linear function) as prediction models. The method was applied after having selected the 70%, 50% and 30% (using *t*-test) of the features whose contribution is more significantly different.

Therefore, for each classifier we can have five different performance results, depending on the features they are trained with:

- No feature selection (882 features).
- *t*-test 70% followed by SFS (6 features for the k-NN as predictive model and 3 features for the SVM).
  - k-NN $TF_{swc}1$: standard deviation value of the frequency band number 1 coefficients over the entire time, extracted from the VAG_{hp-cma}.
  - k-NN $TF_{mwc}5$: average of the frequency band number 5 coefficients over the entire time, extracted from the VAG_{hp}.
  - k-NN $TF_{sth}1$: number of significant differences above the standard deviation value observed between wavelet coefficients of the frequency band number 1, extracted from the $VAG_{norm-hp-cma}$.
  - k-NN $TF_{mwc}10$: average of the frequency band number 10 coefficients over the entire time, extracted from the VAG_{hp-cma}.

  - k-NN $TF_{mwc}15$: average of the frequency band number 15 coefficients over the entire time, extracted from the $VAG_{norm-hp-cma}$.
  - k-NN $TF_{mwc}15$: average of the frequency band number 15 coefficients over the entire time, extracted from the $VAG_{norm-cma}$.
  - SVM $TF_{mwc}1$: average of the frequency band number 1 coefficients over the entire time, extracted from the VAG_{hp}.
  - SVM $TF_{kur}14$: tendency of the frequency band number 14 coefficients to have peaks over time, extracted from the $VAG_{norm-hp-cma}$.
  - SVM $TF_{kur}14$: tendency of the frequency band number 14 coefficients to have peaks over time, extracted from the VAG_{hp-cma}.

- *t*-test 50% followed by SFS (3 features for the k-NN as predictive model and 2 features for the SVM)
  - k-NN TF_{mp}: frequency band which displayed greater power over the entire time for the VAG_{hp-cma}.
  - k-NN $TF_{swc}1$: standard deviation value of the frequency band number 1 coefficients over the entire time, extracted from the VAG_{hp-cma}.
  - k-NN $TF_{swc}1$: standard deviation value of the frequency band number 1 coefficients over the entire time, extracted from the VAG_{cma}.
  - SVM $TF_{mwc}1$: average of the frequency band number 1 coefficients over the entire time, extracted from the VAG_{hp}.
  - SVM $TF_{kur}14$: tendency of the frequency band number 14 coefficients to have peaks over time, extracted from the VAG_{hp-cma}.

- *t*-test 30% followed by SFS (3 features for the k-NN as predictive model and 3 features for the SVM)
  - k-NN TF_{mp}: frequency band which displayed greater power over the entire time for the VAG_{hp-cma}.
  - k-NN $TF_{mwc}1$: average of the frequency band number 1 coefficients over the entire time, extracted from the VAG_{hp}.
  - k-NN $TF_{swc}1$: standard deviation value of the frequency band number 1 coefficients over the entire time, extracted from the VAG_{hp-cma}.
  - SVM $TF_{mwc}2$: average of the frequency band number 2 coefficients over the entire time, extracted from the VAG_{hp}.
  - SVM $TF_{mth}16$: number of significant differences above the mean value observed between wavelet coefficients of the frequency band number 16, extracted from the $VAG_{norm-hp-cma}$.

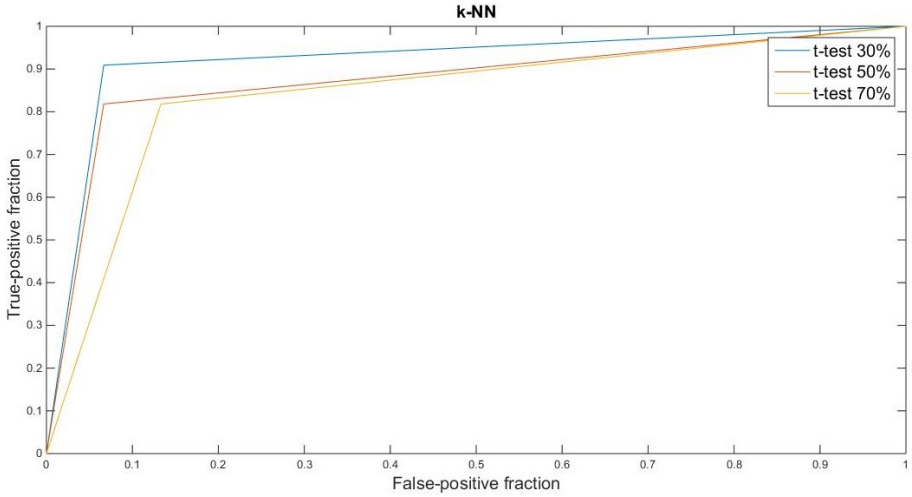


Figure 5.8: ROC plots for the selected features after t-test 30% (TF_{mp}, $TF_{mwc}1$ and $TF_{swc}1$), t-test 50% (TF_{mp}, $TF_{swc}1$ and $TF_{swc}1$) and t-test 70% ($TF_{swc}1$, $TF_{mwc}5$, $TF_{spth}1$, $TF_{mwc}10$, $TF_{mwc}15$ and $TF_{swc}15$) when trained with k-NN (k=3). The corresponding AUC values are, respectively, 0.92, 0.88 and 0.84.

  - SVM $TF_{skew}9$: asymmetry of the frequency band number 9 coefficients, extracted from the VAG_{hp}.

- PCA to keep 99% of variance between the classes (2 principal components).

In order to evaluate which is the preferred t-test filter method for each of the classifiers, the AUC value was computed for the selected features. Figures 5.8 and 5.9 show the ROC curve for three different sets of features (t-test 30%, t-test 50% and t-test 70%) for the k-NN and SVM models, respectively. It can be easily concluded that the discriminant power of the selected features arising from k-NN is much higher.

Namely, t-test 30% followed by the SFS using the k-NN as predictive model obtains the best performance, with an AUC value of 0.92. Thus, these selected features will be used with other classifiers in the next Section, except when SVM is the predictive model.

5.3 Classification

Several classifiers were tested in order to find the one that predicts data with the best performance. In order to do this, the original dataset was split into training and test sets using the 10-fold cross validation method, in order to find the best parameter(s) for one specific model. The best parameter(s) were chosen as the one(s) that returned the best classification accuracy for the majority of iterations. The training and evaluation of each classifier was then done using the 10-fold cross validation approach already with the best chosen parameter(s). Some performance measurements, namely accuracy, sensitivity, precision, specificity and F-measure, were calculated for each iteration. In the case of the ensemble classifiers, the prediction is computed by taking a weighted vote

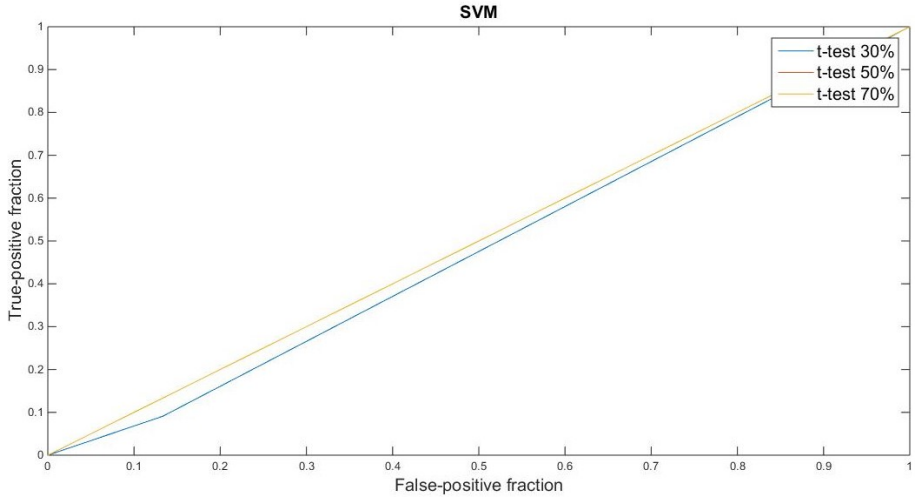


Figure 5.9: ROC plots for the selected features after t-test 30% ($TF_{mwc}2$, $TF_{mth}16$ and $TF_{skew}9$), t-test 50% ($TF_{swc}1$ and $TF_{kur}14$) and t-test 70% ($TF_{mwc}1$, $TF_{kur}14$ and $TF_{kur}14$) when trained with SVM (linear function). The corresponding AUC values are, respectively, 0.48, 0.5 and 0.5.

of the predictions of each classifier, the weights being varied between 0 and 1, so that the best accuracy is obtained.

- k-NN: this model classifies an object by the majority vote of its neighbors, with the object being assigned to the most common class among its k nearest neighbors. Different values of k were tested: $1,3,5,7,9$.

- SVM: this classifier outlines an hyperplane that is able to separate both classes. This hyperplane is determined by a kernel function, which can be of different types: linear, gaussian, radial basis, polynomial.

- FLDA: there is no specific parameter to optimize in this model, since it computes a linear combination of features which leads to the maximization of a discriminant function.

- BPNN: the classification accuracy of this neural network model can depend on the learning rate (α) and the number of nodes in the hidden layer (p). Thus, according to Kim et al. (2009b,c), different values of $\alpha = [0.1, 0.2, ..., 1]$ and $p = [5, 10, ..., 35]$ were tested.

- RBFN: the parameters that most influence this type of neural network are the spread parameter (σ) which was varied between [1,6] and, once again, the number of neurons in the hidden layer, to which values between [1,30] were tested, according to Rangayyan and Wu (2008).

- LS-SVM: this model was proposed by Suykens et al. (2002) as a reformulation to the standard SVM, with an improvement of the moderate complexity. The learning of the LS-SVM

is implemented by minimizing a regularized least-squares cost function with equality constraints. The kernel function used was the radial basis function, according to Cai et al. (2013).

- Naive bayes: although often less accurate than other classifiers, this model was tested, with no parameter to be varied, due to its statistical characteristic as a model.
- Random forest: it is an ensemble classifier which combines the results from different decision tree models. Therefore, the number of decision trees is varied between [1,30] according to Nalband et al. (2016).
- AdaBoost: this model works by repeatedly training a given type of weak-learning machine from different distributed training data sets and then combining their outputs. A combination of a SVM model (rbf), a k-NN (k=3) and Naive Bayes was done, according to Wu (2015).
- Ensemble k-NN/SVM: an ensemble of these two classifiers was not tried in one of the most relevant studies in VAG analysis. For this, the number of neighbors k and the kernel function were varied (in the ranges mentioned before) in order to obtain the best performance possible.
- Ensemble LS-SVM/Random forest: according to Nalband et al. (2016), this ensemble model provided good performance results. The number of decision tree models was varied between [1,30] as before.
- Ensemble BPNN/Random forest: this combination was also not tested before. Due to the good performance of both individual classifiers and the previous use of a random forest in an ensemble classifier (Nalband et al., 2016), this ensemble model was generated. The parameters learning rate, number of nodes in the hidden layer and number of decision tree models were varied in the previously mentioned ranges, in order to obtain the best performance results.
- Ensemble BPNN/k-NN: another innovative ensemble in this field, which was decided due to the good performance of both individual classifiers. The parameters were varied in the previously mentioned ranges.

In Table 5.5, the best performance results for each tested classifier are presented. Specifically for the SFS, the predictive model used was always the k-NN, except when the classifier was the SVM.

As it can be seen in Table 5.5, most of the tested classifiers produced relatively good results in terms of classifying a knee joint as healthy or pathological. Overall, better performance results were obtained when the SFS procedure was applied to the dataset, highlighting the need for the feature selection step for removal of non-discriminant or redundant features.

However, not so good results were obtained when PCA was the used dimensionality reduction technique. Therefore, the ensemble classifiers were only tested applying the SFS technique, due

Table 5.5: Overall performance results (%) for each tested classifier with optimized parameter(s) and a specific feature selection method used.

Classifier	Parameter	FS	Acc	Sen	Pre	Spe	F-meas
k-NN	k=3	*t*-test 30 + SFS	84.62	93.33	82.35	72.73	87.50
k-NN	k=7	PCA	73.08	80	75	63.64	77.42
k-NN	k=3	-	80.77	86.67	81.25	72.73	83.87
SVM	linear	*t*-test 70 + SFS	84.62	93.33	82.35	72.73	87.50
SVM	gaussian	PCA	61.54	100	60	9.09	75
SVM	gaussian	-	57.69	100	57.69	0	73.17
FLDA	-	*t*-test 30 + SFS	53.85	27.27	42.86	73.33	33.33
FLDA	-	PCA	57.69	0	-	1	-
FLDA	-	-	57.69	0	-	1	-
BPNN	p=20, α=0.2	*t*-test 30 + SFS	96.15	93.33	100	100	96.55
BPNN	p=5, α=0.9	PCA	57.69	100	57.69	0	73.17
BPNN	p=15, α=0.9	-	69.23	55.56	76.92	72.73	64.52
RBFN	p=26, σ=2	*t*-test 30 + SFS	80.77	61.90	81.25	72.73	70.27
LS-SVM	-	*t*-test 30 + SFS	92.31	93.33	93.33	90.91	93.33
Naive Bayes	-	*t*-test 30 + SFS	57.69	93.33	58.33	9.09	71.79
Naive Bayes	-	PCA	46.15	80	52.17	0	63.16
Naive Bayes	-	-	61.54	60	69.23	63.64	64.29
Random forest	13 trees	*t*-test 30 + SFS	88.46	93.33	87.50	81.82	90.32
Random forest	4 trees	PCA	61.54	66.67	66.67	54.55	66.67
Random forest	15 trees	-	88.46	86.67	92.86	90.91	89.66
AdaBoost	-	*t*-test 30 + SFS	92.31	86.67	100	100	92.86
0.45SVM + 0.55k-NN	linear; k=3	*t*-test 30 + SFS	96.15	93.33	100	100	96.55
0.5LS-SVM + 0.5RF	17 trees	*t*-test 30 + SFS	88.46	93.33	87.50	81.82	96.55
0.5BPNN + 0.5RF	p=35, α=0.3; 16 trees	*t*-test 30 + SFS	96.15	93.33	100	100	96.55
0.5BPNN + 0.5k-NN	p=5, α=1; k=3	*t*-test 30 + SFS	96.15	100	100	96.55	100

FS: Feature selection; Acc: Accuracy; Sen: Sensitivity; Pre: Precision; Spe: Specificity; F-meas: F-measure.

to the worst results obtained by the classifiers when the PCA or no feature selection algorithm was used.

In terms of classification performance, four models revealed to be the best among all tested: BPNN; ensemble of SVM and k-NN (weight SVM=0.45); ensemble of BPNN and Random forest (weight BPNN=0.5); ensemble of BPNN and k-NN (weight BPNN=0.5). These classifiers were able to correctly discriminate 96.15% of the signals with sensitivity of 93.33%, which means that 93.33% of the pathological signals are correctly identified as pathological. The correct identification of the pathological class is very important in a classification screening system. LS-SVM and AdaBoost should also be highlighted for their good performance, namely an accuracy of 92.31%.

The obtained performances are superior to the ones obtained in other studies, as shown in Section 4.7, despite the lack of information in this study to segment signals in the ascending/descending movement.

Specifically, when the BPNN model is used by Kim et al. (2008, 2009b,c,a), an accuracy of 91.4% was obtained, whereas in this study this value was 96.15%. Kim, in his studies, used a distinct sensor to collect the VAG signals (electrostethoscope) and different time-frequency features were extracted, namely EP, ESP, FP and FSP. Also, the dataset was constituted by a smaller number of samples (20 normal and 11 abnormal signals).

The classifier LS-SVM was used by two authors (Nalband et al., 2016; Cai et al., 2013), with accuracies of 87.64% and 94.31%. A value of 92.31% was obtained in this study using this classifier. The same type of sensor was used (accelerometer), but different types of features were extracted. While in our study, TF features were estimated, Nalband et al. (2016) computed RQA features, as well as ApEn and SampEn, and Cai et al. (2013) measured only statistical parameters. It can be concluded that this type of classifier provides similar performance results independently of the type of features extracted.

Rangayyan et al. (2013); Liu et al. (2014) had the FLDA as classification model and values of 75.10% and 77% were obtained, when an accelerometer was used to collect the signals. Fractal analysis features and statistical parameters were, respectively, computed. The lower accuracy of 57.69% reached in the present study can be explained by the different nature of the features extracted (time-frequency domain), which may not be appropriate for the FLDA classifier.

Better accuracy values were obtained for both the k-NN and SVM. Namely, 84.62% for the k-NN against the 80% obtained by Liu et al. (2014), where statistical parameters were computed. In the case of SVM, a value of 84.62% was also obtained, against the 81.33% by Wu et al. (2013), where once more, only statistical parameters were computed.

Concerning the RF model, similiar accuracies were obtained in this study and in Nalband et al. (2016); Kręcisz and Bączkowicz (2018) studies. Namely, 88.46% against 91.01% and 87.2%. While Nalband et al. (2016) extracted RQA features as well as ApEn and SampEn, Kręcisz and Bączkowicz (2018) computed statistical parameters, but also RQA features and fractal analysis. Similarly to LS-SVM, the RF classifier obtains identical performance results, independently of the nature of the extracted features.

Chapter 6

Analysis using signals collected with OASense

6.1 Study overview

The main goal of this experiment is to evaluate the functionality of OASense, namely its capability of screening patients affected by Osteoarthritis through crepitus detection. The device was used to acquire signals from patients with OA affected knees and healthy participants, in order to validate its working principle. Two different signals (VAG and Phono-A) were acquired and recorded by the device, as explained before. VAG signals were subjected to some processing and classification algorithms. According to those, it is possible to label them as belonging to an OA affected knee or an unaffected knee.

The OA affected patients were selected by a specialist from Reinier de Graaf Hospital. The participants presenting unaffected knees were selected by the researcher and coordinators. An information letter was signed by all participants, which explains the purpose of the study as well as the tasks that should be done. It is also stated that all data will be kept confidential, an identification random number being given to each participant.

In the case of patients presenting OA affected knees, the experiment was conducted immediately after their appointment with the orthopedist at the hospital. Regarding subjects with unaffected knees, the measurements were done at one of the TUDelft laboratories.

All participants should complete a questionnaire (see Appendix C.1) before enrolling the experiment, which contains sociodemographic and disease data as well as questions related to symptoms and effect of OA in their daily life. These questions were adapted from the Oxford Knee Score (OKS) (of Oxford, 2018), which is a 12-item questionnaire specifically designed and developed to assess function and pain after TKR surgery. However, it can also be used to evaluate OA related symptoms and possible implications in daily life. Each of the 12 questions is scored from 0 to 4 with the score decreasing as the reported symptoms increase (i.e. become worse). This method, when summed, produces overall scores from 0 to 48, the best outcome being. The

questionnaire should be analysed by the researcher prior to the experiment itself, to evaluate if the participant respects the inclusion and exclusion criteria, which can be found in Appendix C.2.

After that, the measurement requires the subject to be sat on a rigid table in a relaxed position with the leg under test freely suspended in air. The signal is recorded by the researcher as the subject executes one movement cycle, which consists on swinging the leg over an approximate angle range of 90º (full flexion) to 0º (full extension) and back to 90º in 3s. This velocity should be maintained constant through the use of beeping sounds controlled by the researcher.

The accelerometer should be held at the medial compartment slightly below the midline of the patella (medial condyle on the patella) by the researcher while the participant executes the movement. The experiment comprises a training session and an active movement phase. The training is required so that the subject is warmed up passively in order to keep the speed of each cycle constant. Both phases consist on repetitions of blocks of 10 cycles (3s each) with pauses of 1 minute between them, in order to avoid muscle fatigue. The training is composed of 1 block while the active movement phase has 3 blocks.

6.2 Data acquisition

Due to the different nature and acquisition conditions of the signals, a different approach for signal analysis is used (Figure 6.1). This new method adds data preparation and segmentation. The selected extracted features and the classifier which returns the best performance results were already obtained in Chapter 5.

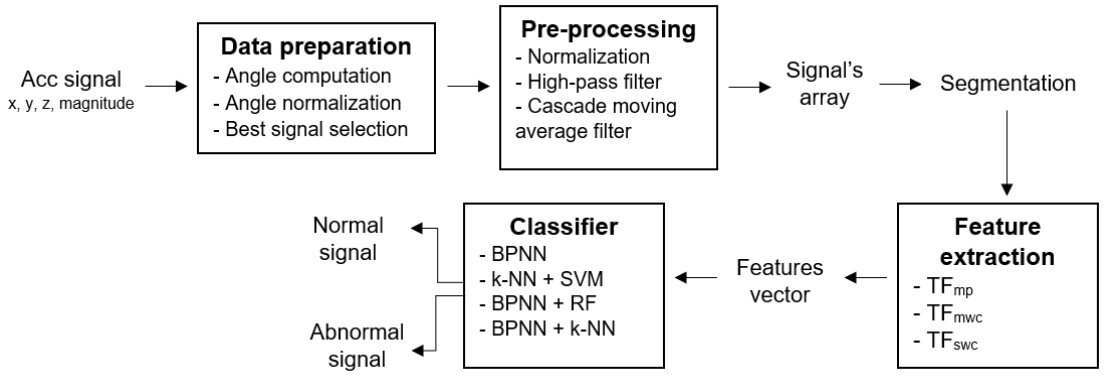


Figure 6.1: Approach for the analysis of the collected signals with OASense.

6.2.1 Dataset

The dataset (see Table 6.1) consists of 48 signals, with 30 from unaffected participants (8 male and 22 female, 17 right and 13 left knees) and 18 from subjects with diagnosed knee OA (6 male and 12 female, 7 right and 11 left knees). This dataset does not permit classification of the signals

into the several severity stages of OA. Therefore, it can only be used for screening purposes, i.e., classification of the signals and the corresponding knee joints as affected or unaffected by OA.

For analysis, only 16 out of the 18 abormal signals and 26 out of the 30 normal signals could be considered. The 2 abnormal signals had to be excluded since the patients were not able to execute the extension-flexion movement the required amount of times and with the constant speed, due to pain. 4 normal signals were excluded, since the participants didn't meet the exclusion/inclusion criteria to be considered eligible for the study (see Appendix C.2).

Table 6.1: Mean and standard deviation of demographic data from both the OA affected and unaffacted participants.

	Total	**BMI**	**Age**	**OKS**
Normal	30	22.7±2.9	27±6.4	47.7±0.7
OA	18	29.2±6.6	70±9.1	21.3±7.6

6.2.2 Acceleration computation

Due to hardware and firmware limitation, the maximum available sampling rate for acceleration recordings was set to 200 Hz which should be enhanced in a near future. For each trial, the three aceleration components (in m/s^2), under the form of x, y and z axis signals, were computed for all time instants, after collection of the digital values from the sensor. Additionally, the magnitude signal which is a combination of the three components of the acceleration signal was computed according to Equation 6.1.

$$A(t) = \sqrt{x(t)^2 + y(t)^2 + z(t)^2}, \tag{6.1}$$

where $A(t)$ is the resulting magnitude signal and $x(t)$, $y(t)$ and $z(t)$ are each one of the computed acceleration from the x, y and z axis, respectively. These 4 signals can be depicted in Figure 6.2.

6.3 Pre-processing

6.3.1 Signal selection

As already mentioned, four different recordings were collected from each participant (one for training and three for the active movement phase). Therefore, in order to select the best signal, the angular data was firstly computed for each trial. After normalization, some parameters were computed to choose the best signal.

6.3.1.1 Angle computation

Instead of using the normal apparatus for angle measurement, the electrogoniometer, an innovative approach was used: the calculation of the angles using the three-axis accelerometer values. This

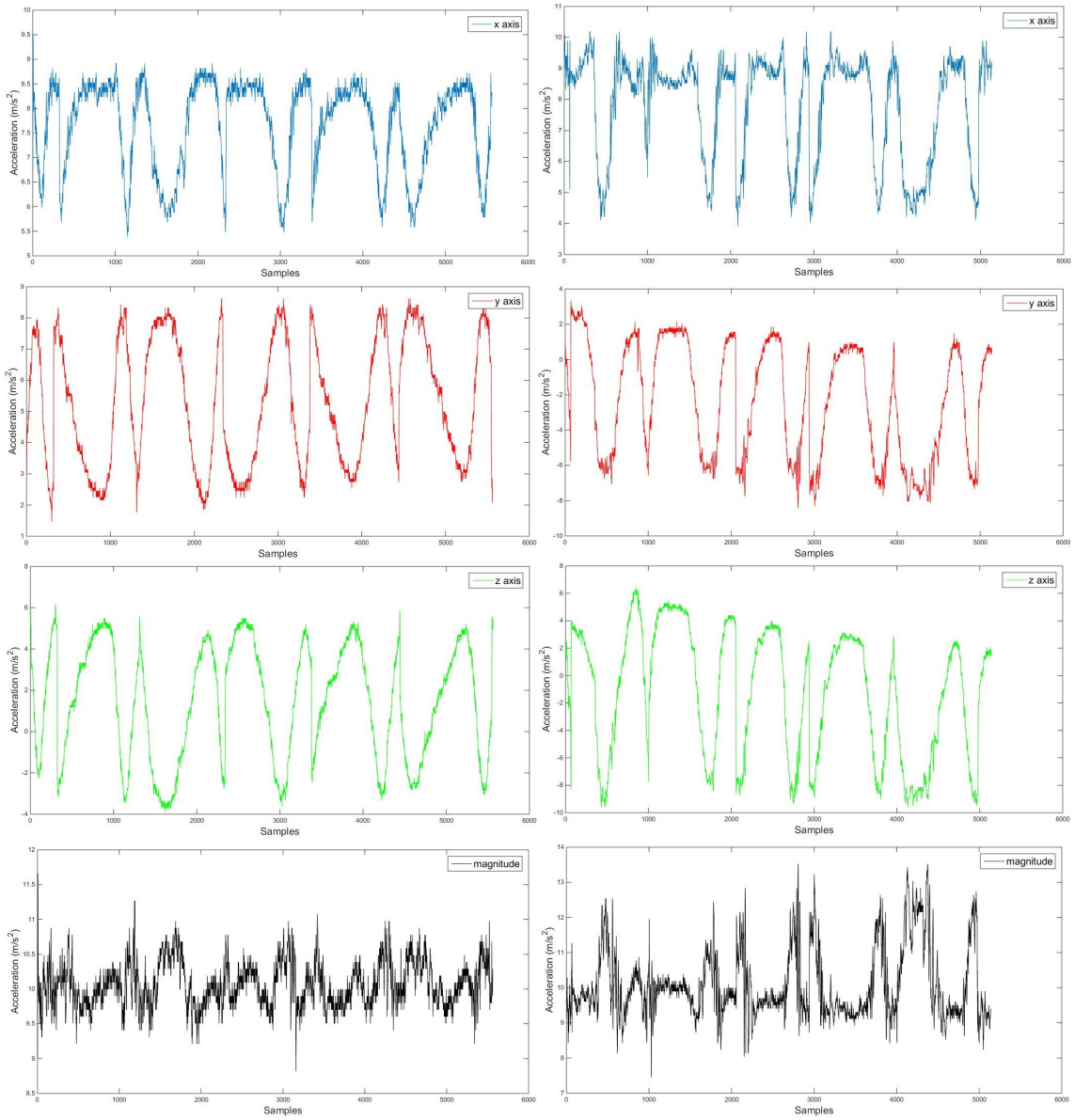


Figure 6.2: Three acceleration components ($x(t)$, $y(t)$ and $z(t)$) and the magnitude signal (top to bottom) for an unaffected knee (left) and an OA affected knee (right).

approach was based on the dot product between two vectors (A and B), which involves the angle formed by them (θ), as shown in Equation 6.2.

$$A.B = |A|\,|B|\,cos(\theta) \tag{6.2}$$

The angles that are expected to be obtained should be calculated between the acceleration vector and a reference frame, which is composed of three vectors (Ref_x, Ref_y and Ref_z), one for each of the axis of the accelerometer. The coordinates for these vectors are (1,0,0), (0,1,0) and (0,0,1), respectively.

Thus, for each x, y and z acceleration values, a vector with the coordinates (x,y,z) was created

and the dot product was applied to calculate the angle between each one of the reference vectors (Ref_x, Ref_y and Ref_z) and the acceleration-derived vector for all time instants, resulting into 3 different angular values (x_{ang}, y_{ang} and z_{ang}), as represented in Figure 6.3.

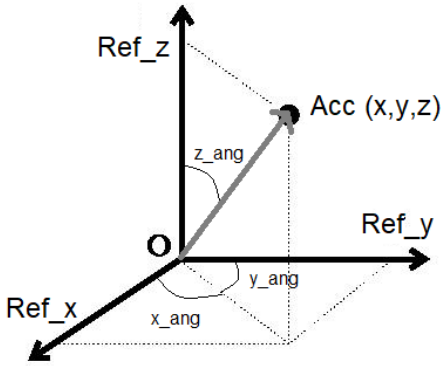


Figure 6.3: Reference frame composed of three directional vectors (Ref_x, Ref_y and Ref_z) and the accelerometer's three dimensional position ($Acc(x,y,z)$) for each time instant. x_{ang}, y_{ang} and z_{ang} are the resulting angles between the sensor and Ref_x, Ref_y and Ref_z, respectively.

Analyzing Figure 6.3, it is easy to conclude that the relevant angle which corresponds to the flexion-extension movement is the z_{ang}.

In Figure 6.4, the z_{ang} measurements are shown for a signal corresponding to an unaffected knee and an OA affected knee. It is possible to see that it becomes more difficult for the OA participant to reach the extension position throughout the trial.

6.3.1.2 Angle normalization

While doing the measurements, it was observed that some participants, mostly the knee OA affected ones, were not able to complete the full range motion (see Figure 6.4). Thus, the computed z_{ang} did not vary between 0º (full extension) and 90º (full flexion).

In order to increase algorithm robustness by taking into account the observed differences in terms of the angle values across subjects and trials, a normalization approach was applied to all the computed angles. This allowed us to have a new range of z_{ang} values bounded to the interval [0º 90º] (see Figure 6.5). The applied method is based on the minimum and maximum amplitude values (min-max) of each signal.

6.3.1.3 Selection

The first computed parameters were the maximum (max) and minimum (min) values for the angle. Since each trial consists on 10 cycles, a maximum of 20 extremes (10 max and 10 min) can be found. In addition, to be considered a max, the value has to be over 65º and a min below 35º. Also, each max and min has to be separated from the next same extreme at least 2s, knowing that the required time for each trial was defined as 3s. For instance, in Figure 6.6, it is possible to see

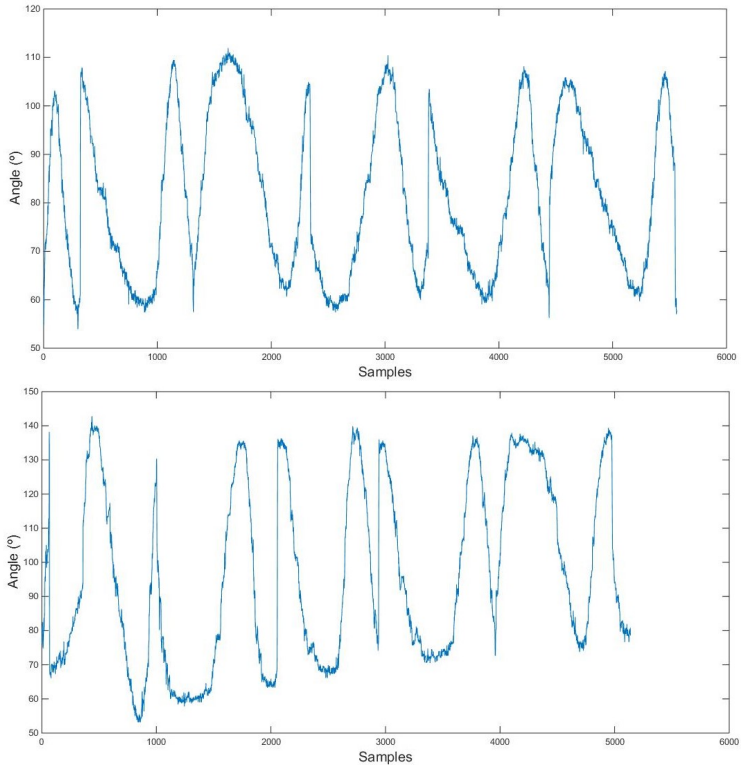


Figure 6.4: z_{ang} measurements for an unaffected knee (top) and an OA affected knee (bottom).

that the first max value was not considered, because it was not distant enough from the next max. However, even with all these criteria to find the extremes, there are still some signals that have two adjacent max or min without the other extreme in the middle. These signals are detected and eliminated.

Secondly, the distance between every max and min is computed for each angle measurement and the average value is obtained. Also, the average is calculated for the maximum and for the minimum values. Each of these 4 parameters is given a score from 1 to 4 for each of the 4 signals collected from the same participant, according to the proximity of the computed distance to 3s, and the average max and min values to 90º and 0º, respectively. After this, each of the 4 signals has a score, according to the performance of each of the parameters. The chosen signal for further analysis is the one with the highest score.

6.3.2 Signal's array creation

Similiarly to the approach in Chapter 5, an array of signals was created in order to increase the number of features that may help in discriminating between the normal and abnormal signals.

This array is constituted by seven different signals, which differ in the pre-processing techniques used. The same methods were used as in Chapter 5: normalization followed by an high-pass filtering and a cascade moving average filtering. Figure 6.7 shows the effect of the pre-processing in an unaffected signal.

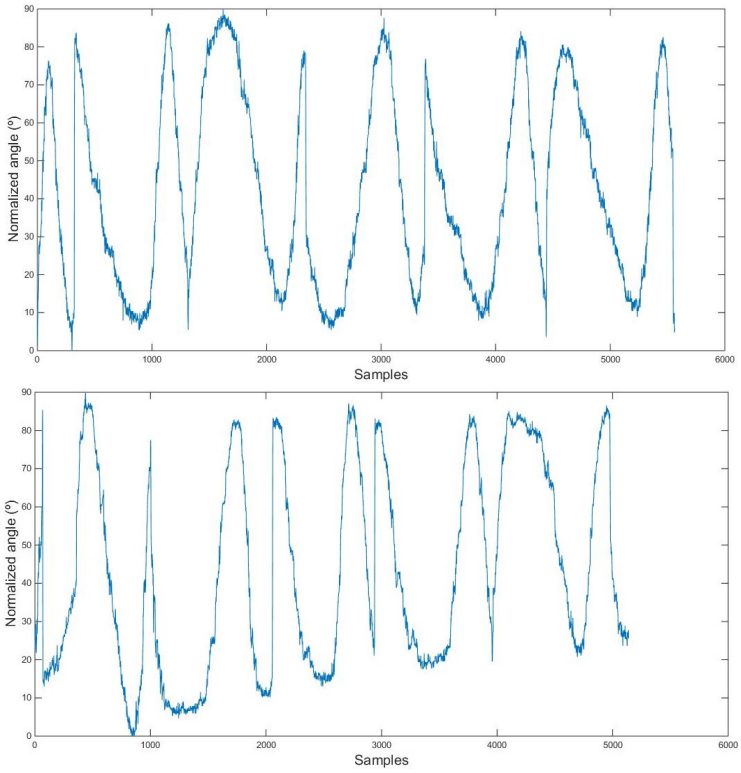


Figure 6.5: z_{ang} normalized measurements for an unaffected knee (top) and an OA affected knee (bottom).

Therefore, the array contains the normalized signals (A_{norm}), the high-pass filtered signals (A_{hp}), the CMA filtered signals (A_{cma}), the normalized high-pass filtered signals ($A_{norm-hp}$), the normalized CMA filtered signals ($A_{norm-cma}$), the high-pass and CMA filtered signals (A_{hp-cma}) and the CMA high-pass filtered normalized signals ($A_{norm-hp-cma}$).

For each signal, which is composed of 3 components (x, y, z) and also the magnitude signal, a total of 126 features were obtained. Thus, a total of 882 features were extracted (for each of the 4 components), taking into account the 7 element signal's array.

6.3.3 Segmentation

In order to select the most meaningful data and desired segments of the signals, as described in Chapter 5, a segmentation approach was applied to all of the available signals. The signals were segmented according to the normalized angular data (z_{ang}) into segments corresponding to a specific phase of the knee flexion/extension movement, i.e., the middle phase of the movement. This phase was defined as being 2/4 of the total angular displacement obtained during one leg swing.

Therefore, the segmentation was performed (see Figure 6.8), taking into account the initial angle ($iangle$), the final angle ($fangle$) and the angular displacement (δ_{ang}), which corresponds to the difference between the first two, for each leg swing movement/cycle of motion. In the ascending movement $iangle$ is close to 90º, which represents the maximum flexion of the knee in

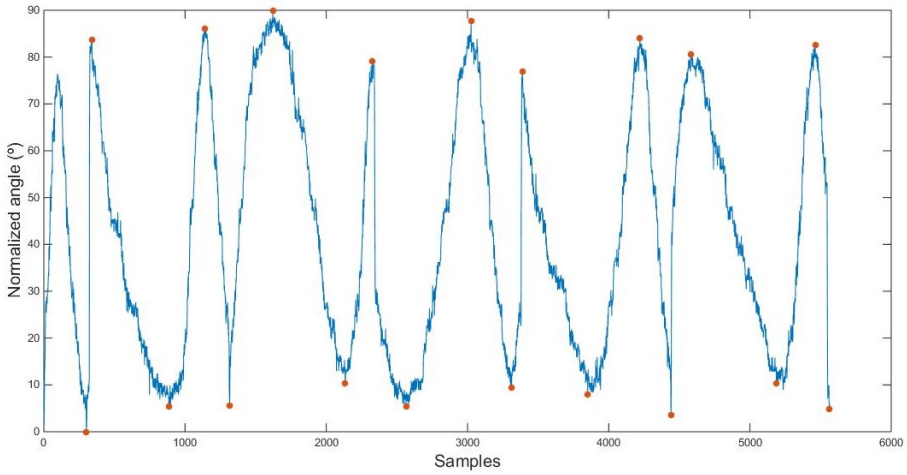


Figure 6.6: z_{ang} normalized measurements for an unaffected knee with the computed extremes in orange (min and max).

that cycle and $fangle =$ is close to 0º, which corresponds to the maximum knee extension in that cycle, whereas in the descending phase, it is the opposite scenario.

In order to extract only the middle phase (2/4 of the total angular displacement) for each leg swing movement, two thresholds are created, one for the lower cut-off angle (Th_1) and another for the higher cut-off angle (Th_2). Equations 6.3 and 6.4 represent the calculation of these thresholds for the ascending movement and Equations 6.5 and 6.6 for the descending movement.

$$Th_1 = iangle + \frac{1}{4} * \delta_{ang} \tag{6.3}$$

$$Th_2 = fangle - \frac{1}{4} * \delta_{ang} \tag{6.4}$$

$$Th_1 = fangle + \frac{1}{4} * \delta_{ang} \tag{6.5}$$

$$Th_2 = iangle - \frac{1}{4} * \delta_{ang} \tag{6.6}$$

In Figure 6.8, it is clear that segments have different number of samples, due to the fact that the motion speed was not kept constant at 3s/cycle. In general, the ascending movement (from 90º to 0º) is executed slower than the descending one.

6.4 Feature extraction and selection

As the sampling rate was set to 200Hz, the wavelet packet transform (with 4 levels of decomposition and the DB2 wavelet) will give rise to 16 different frequency band values. Thus, the first

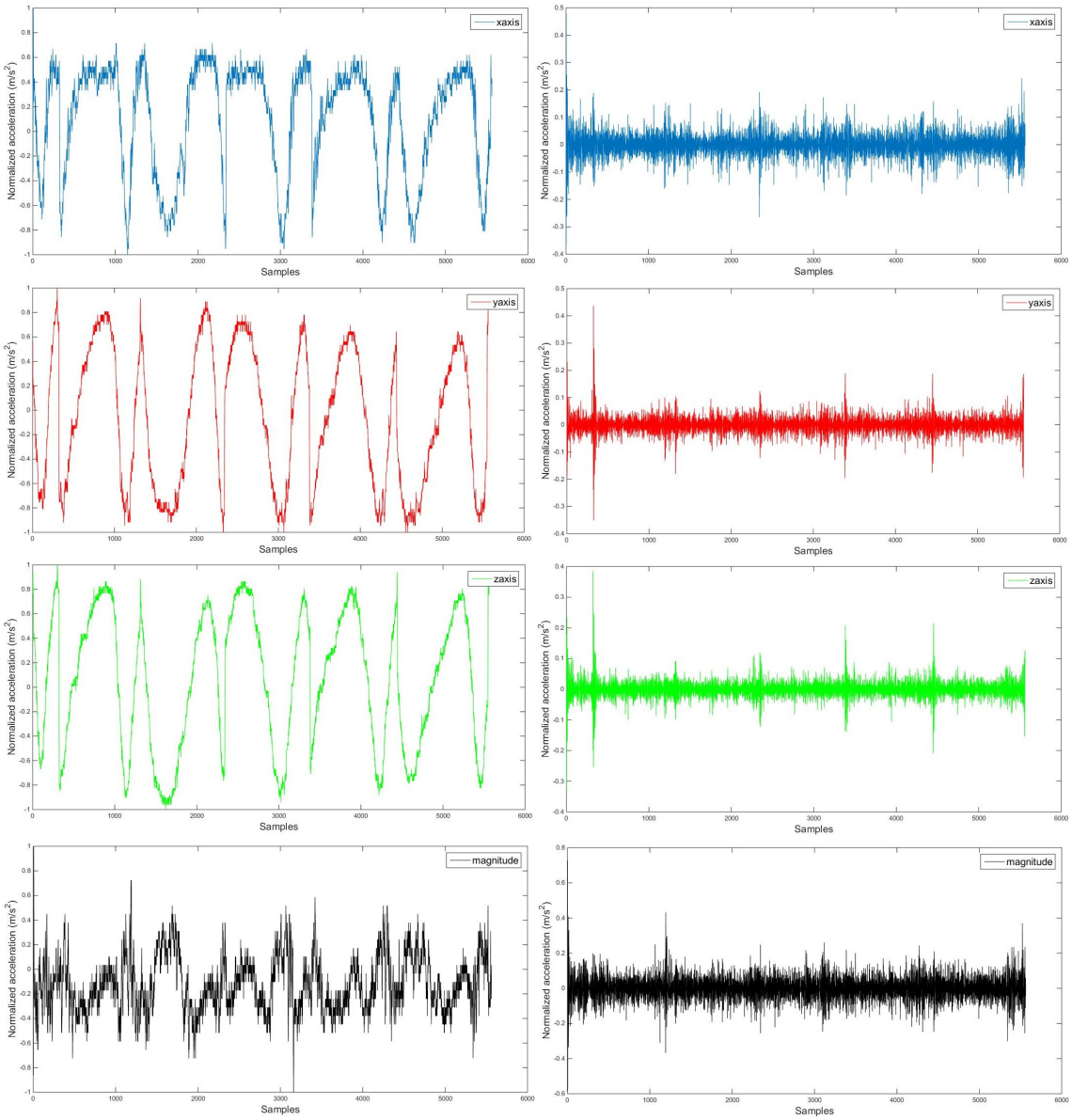


Figure 6.7: Normalized $x(t)$, $y(t)$, $z(t)$ and magnitude signal (left) and their high-pass and CMA filtered versions (right).

level of the WPT is 0-100Hz and then the ranges for each band are computed similarly to what was done in Chapter 5

Having the WPT, the features which provided the best results for the open-source VAG dataset were extracted, which were obtained via *t*-test 30% followed by SFS with k-NN (k=3) as predictive model.

- TF_{mp}: frequency band which displayed greater power over the entire time for the A_{hp-cma}.
- $TF_{mwc}1$: average of the frequency band number 1 coefficients over the entire time, extracted from the A_{hp}.

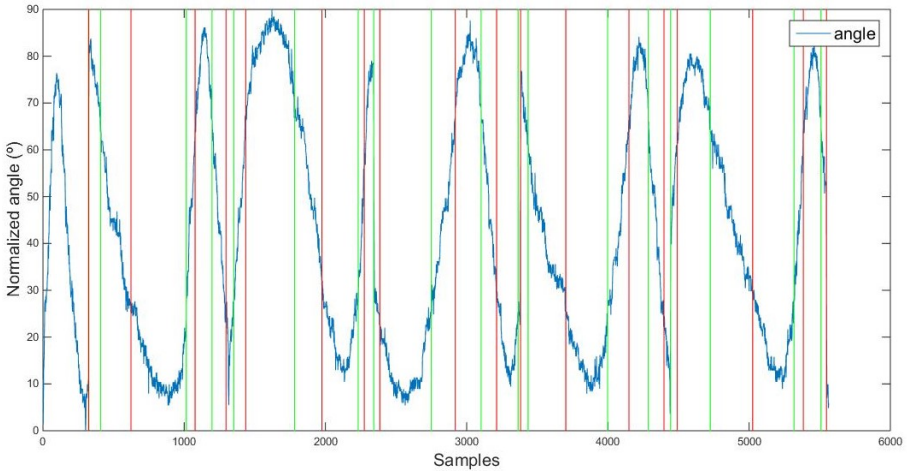


Figure 6.8: Example of the angular segmentation algorithm performed on the z_{ang} of a normal signal. Samples delimited by the green (at the left) and the red line (at the right) represent the chosen acceleration segments correspondent to the middle phase of the leg swing movement.

- $TF_{swc}1$: standard deviation value of the frequency band number 1 coefficients over the entire time, extracted from the A_{hp-cma}.

With this dataset, there are four different signals from which these three features are computed: the three acceleration components (x, y and z) and the resultant magnitude signal. Therefore, features were extracted from different signals in turns, having 64 possible combinations of features (4^3). The best combination was selected taking into account the accuracy obtained with the trained classifier.

Table 6.2: Mode of TF_{mp} feature values for the A_{hp-cma} signals, mean of $TF_{mwc}1$ for the A_{hp} signals and mean of $TF_{swc}1$ for the A_{hp-cma} signals (taking into account the 4 components).

	x axis		**y axis**		**z axis**		**magnitude**	
	N	**OA**	**N**	**OA**	**N**	**OA**	**N**	**OA**
$\mathbf{TF_{mp}}$	10	11	11	10	11	11	11	10
$\mathbf{TF_{mwc}1}$	6.80E-3	6.70E-3	1.11E-2	1.09E-2	1.10E-2	1.13E-2	9.70E-3	9.40E-3
$\mathbf{TF_{swc}1}$	6.40E-3	7.10E-3	1.09E-2	8.20E-3	1.13E-2	8.80E-3	9.40E-3	6.90E-3

N: normal signals, OA: OA affected signals

From Table 6.2, it is possible to conclude that the frequency band with most power is, mostly 11 for the normal signals and 10 for the abnormal ones. The other two feature values match the results obtained in Figure 5.7 (although extracted from different elements of the signal's array), with $TF_{mwc}1$ and $TF_{swc}1$ being bigger for the normal signals.

6.5 Classification

The models which provided the best performance results for the open-source VAG dataset were tested with the extracted features from the collected signals, namely BPNN (p=20,α=0.2), ensemble k-NN/SVM (weight k-NN=0.55, k=3, linear kernel function), ensemble BPNN/RF (weight BPNN=0.5, p=35, α=0.3, 16 trees) and ensemble BPNN/k-NN (weight BPNN=0.5, p=5, α=1, k=3). The results are displayed in Table 5.5.

It is possible to conclude that BPNN is the classifier that provides the best performance results for the collected dataset. It is able to correctly discriminate 93% of the signals, with a sensitivity of 96,67%, which means that 96,67% of the signals collected from OA patients are correctly identified as unhealthy. Although the specificity value (ability to correctly identify signals collected from unaffected people), 85%, is a bit lower, the correct identification of the pathological cases is much more important in a classification screening system. Both the ensemble BPNN/k-NN and BPNN/RF provided also good results. However, lower performance values were obtained with the ensemble k-NN/SVM.

Table 6.3: Performance results (%) for each classifer for the collected signals using OASense.

	Accuracy	**Sensitivity**	**Precision**	**Specificity**	**F-measure**
BPNN	93	96.67	94.17	85	94.57
k-NN/SVM	83.5	85	91.67	85	84.81
BPNN/RF	90.5	93.33	95	80	93.14
BPNN/k-NN	68	85	71.67	40	76.14

Chapter 7

Conclusions and Future work

The development of a classification system that is able to discriminate between signals collected from unaffected and affected knees was successfully obtained. Thus, a powerful algorithm, which includes signal pre-processing, feature extraction and training of the best model was implemented. In comparison with the results reported in previous studies on the analysis of VAG signals with the same dataset (51 unaffected and 38 affected subjects), the present study has provided comparable or better screening accuracies. Having obtained an accuracy of 96.15%, the adoption of this same classification system with the collected data from OASense, has allowed the development of a screening test, with an accuracy of 93%, for knee OA based on the detection of VAG signals. The validation of this device was accomplished through the comparison of the results with X-ray images acquired from OA affected patients.

OASense may be used as a reliable, accurate, cheap and non-invasive screening diagnostic tool at primary healthcare level. Such system could also be used as a monitoring tool in combination with physiotherapy which would improve the overall rehabilitation process. At secondary healthcare level, orthopedists may use the OASense allied to X-ray images to monitor patient's condition and to check the need for a revision surgery. The device would provide detailed insight, at cartilage level, about the knee joint status and affected structures that may not be detected with other current diagnostic tools (only gross and symptomatic changes are detected with the current image-based techniques), possibly enabling the early detection of knee OA.

Despite the promising results, future work can be performed in both data collection and analysis process to increase the overall robustness and broadness of the screening test. It is suggested to acquire signals under different types of friction and/or loading conditions, increase the sensor's sampling frequency and optimize the experimental protocol (uniformize the leg swing velocity is very important, which can be improved through the use of beeps for each ascending/descending phase instead of the complete cycle). Also, data from more participants should be collected to enhance the power of the classification system. Signal pre-processing can also be optimized, namely a better reduction of the random noise is advisable. This can be done through the implementation of the Singular Value Decomposition (SVD) algorithm, described by Kim et al. (2008), for example. Some new features could be extracted, such as the ones resulting from fractal analysis, which

gave good results in Rangayyan et al. (2013). Different filter feature selection techniques could be tested as well as other neural networks or ensemble classifiers. Regarding the OASense device itself, it would be useful to install a Bluetooth unit so it can communicate directly with an external processing unit that would analyse the data automatically.

The device was initially conceptualized to be used as a knee OA screening tool. However, VAG signals have been to shown to be associated to other knee-joint disorders, such as Rheumathoid Arthritis. Thus, it is suggested the VAG signals be collected from patients affected by other diseases. It would be also helpful to discriminate between different severity stages, including the earliest stage, in which patients are affected by knee pain but the cartilage defect is barely noticeable. Therefore, data collection from patients at different disease stages is suggested, which can be evaluated with the KL grade obtained from X-ray images, for example.

Bibliography

J. Abulhasan and M. Grey. Anatomy and Physiology of Knee Stability. *Journal of Functional Morphology and Kinesiology*, 2(4):34, 2017.

P. A. Aguilera, A. Fernández, R. Fernández, R. Rumí, and A. Salmerón. Bayesian networks in environmental modelling. *Environmental Modelling & Software*, 26(12):1376–1388, 2011.

U. Ahmed, A. Anwar, R. S. Savage, M. L. Costa, N. Mackay, A. Filer, K. Raza, R. A. Watts, P. G. Winyard, J. Tarr, R. C. Haigh, P. J. Thornalley, and N. Rabbani. Biomarkers of early stage osteoarthritis, rheumatoid arthritis and musculoskeletal health. *Nature scientific reports*, 5:1–7, 2015.

L.I. Alekseeva and E. L. Nasonov. Osteoarthritis and comorbidities. *Medicographia*, 35(2):152–157, 2013.

AnalogDevices. Small, low power, 3-axis ±3g accelerometer adxl335, 2018. available in `http://www.analog.com/media/en/technical-documentation/data-sheets/ADXL335.pdf`, accessed 16-05-2018.

Arduino. Arduino™, 2018. available in `https://www.arduino.cc/`, accessed 26-01-2018.

Autodesk. Eagle, 2018. available in `https://www.autodesk.com/products/eagle/overview`, accessed 16-05-2018.

G. Balint and B. Szebenyi. Diagnosis of osteoarthritis. *Drugs*, 52(3):1–13, 1996.

H. M. Bassiouni. Phonoarthrography: A New Technique for Recording Joint Sounds. *Osteoarthritis - Diagnosis, Treatment and Surgery*, 16:275,288, 2012.

H. M. Bassiouni, M. El-Deeb, N. Kenawy, E. Abdul-Azim, and M. Khairy. Phonoarthrography, musculoskeletal ultrasonography, and biochemical biomarkers for the evaluation of knee cartilage in osteoarthritis. *Modern Rheumatology*, 21(5):500–508, 2011.

M. Bassiouni, H. Bassiouni, and M. El-Feki. Sensitivity versus specificity of phonoarthrography as an indicator for cartilage degeneration. *Clinical Rheumatology*, 14(2):135–142, 1995.

M. H. Bassiouni. Phonoarthrography of the knee in health and diseases. *MSc thesis. Department of Rheumatology and Rehabilitation, Al-Azhar University*, pages 97–102, 1986.

N. Befrui, J. Elsner, A. Flesser, J. Huvanandana, O. Jarrousse, T. Le, W. H. W. Schulze, S. Taing, and S. Weidert. Vibroarthrography for early detection of knee osteoarthritis using normalized frequency features. *Medical and Biological Engineering and Computing*, 2018.

A. Berger. How does it work?: Positron emission tomography. *BMJ: British Medical Journal*, 326(7404):1449, 2003.

W. E. Blodgett. Auscultation of the knee joint. *The Boston Medical and Surgical Journal*, 146(3): 63–66, 1902.

P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the institute of phonetic sciences*, volume 17, pages 97–110. Amsterdam, 1993.

V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos. A review of feature selection methods on synthetic data. *Knowledge and information systems*, 34(3):483–519, 2013.

A. P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.

H. J. Braun and G. E. Gold. Diagnosis of osteoarthritis: imaging. *Bone*, 51(2):278–288, 2012.

S. Cai, Y. Wu, N. Xiang, Z. Zhong, J. He, L. Shi, and F. Xu. Detrending knee joint vibration signals with a cascade moving average filter. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 4357–4360, 2012.

S. Cai, S. Yang, F. Zheng, M. Lu, Y. Wu, and S. Krishnan. Knee joint vibration signal analysis with matching pursuit decomposition and dynamic weighted classifier fusion. *Computational and Mathematical Methods in Medicine*, 2013, 2013.

P. Castiglioni. Wigner–ville distribution. *Encyclopedia of Biostatistics, 1*, 2015.

CDC. Risk factors, 2018. available in `https://www.cdc.gov/arthritis/basics/risk-factors.htm`, accessed 16-05-2018.

D. Chaudhuri and R. Singh. Applications of accelerometer as a vibration detector. *International Journal of Emerging Trends in Science and Technology*, 2(03), 2015.

G. Cherryman. Imaging in primary care. *The British Journal of General Practice*, 56(529):563–564, 2006.

E. M. Christensen, R.and Bartels, A. Astrup, and H. Bliddal. Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. *Annals of the rheumatic diseases*, 66(4):433–439, 2007.

C. R. Chu, A. A. Williams, C. H. Coyle, and M. E. Bowers. Early diagnosis to enable early treatment of pre-osteoarthritis. *Arthritis Res Ther.*, 14(3), 2012.

C. Cooper. Epidemiology of osteoarthitis. *Medicographia*, 35(2):145–151, 2013.

M. D. Crema, A. Guermazi, E. C. Sayre, F. W. Roemer, H. Wong, A. Thorne, J. Singer, J. M. Esdaile, M. D. Marra, J. A. Kopec, et al. The association of magnetic resonance imaging (mri)-detected structural pathology of the knee with crepitus in a population-based cohort with knee pain: the modeko study. *Osteoarthritis and cartilage*, 19(12):1429–1432, 2011.

M. Dash and H. Liu. Feature selection for classification. *Intelligent data analysis*, 1(3):131–156, 1997.

S. C. Douglas. Introduction to adaptive filters. *Digital Signal Processing Handbook. Ed. Vijay K. Madisetti and Douglas B. Williams. Boca Raton: CRC Press LLC*, 1999.

C. Drugă, I. Șerban, and D. Cotoros. Vibration measurement system with mems accelerometers based on arduino. *Computational Mechanics and Virtual Engineering*, 2017.

Dytran. Accelerometers, 2018. available in `https://www.dytran.com/Accelerometers/`, accessed 16-05-2018.

K. H. Erb. Auscultation and recording of knee joint noises (english translation). *Deutsche Ztschr Chir*, 241:237–245, 1933.

D. T. Felson, R. C. Lawrence, P. A. Dieppe, R. Hirsch, C. G. Helmick, J. M. Jordan, R. S. Kington, N. E. Lane, M. C. Nevitt, Y. Zhang, et al. Osteoarthritis: new insights. part 1: the disease and its risk factors. *Annals of internal medicine*, 133(8):635–646, 2000.

C. Ferri, J. Hernández-Orallo, and R. Modroiu. An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, 30(1):27–38, 2009.

Y. Freund and R. E. Schapire. Large margin classification using the perceptron algorithm. *Machine learning*, 37(3):277–296, 1999.

N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine learning*, 29 (2-3):131–163, 1997.

S. Glyn-Jones, A. J. R. Palmer, R. Agricola, A. J. Price, T. L. Vincent, H. Weinans, and A. J. Carr. Osteoarthritis. *The Lancet*, 386(9991):376–387, 2015.

M. Y. Gokhale and D. K. Khanduja. Time domain signal analysis using wavelet packet decomposition approach. *International Journal of Communications, Network and System Sciences*, 3 (03):321, 2010.

J. P. Goldblatt and J. C. Richmond. Anatomy and biomechanics of the knee. *Operative Techniques in Sports Medicine*, 11(3):172–186, 2003.

I. M. Grais. Proper use of the stethoscope: Three heads and one tale. *Texas Heart Institute Journal*, 40(2):120, 2013.

A. Guermazi, F.W. Roemer, and H.K. Genant. Role of imaging in osteoarthritis: diagnosis, prognosis and follow-up. *Medicographia*, 35(2):164–171, 2013.

M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

R. Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier, 1992.

C. Heuter. *Hueter-Lossen's Grundriss der Chirurgie*. Leipzig : F.C.W. Vogel, 1889.

V. D. Hoang. Wavelet-based spectral analysis. *TrAC Trends in Analytical Chemistry*, 62:144–153, 2014.

M. C. Hochberg. Osteoarthritis: New approaches. *Medicographia*, 35(2):139–141, 2013.

J. J. Hopfield. Artificial neural networks. *IEEE Circuits and Devices Magazine*, 4(5):3–10, 1988.

K. K. Hovis, C. Stehling, R. B. Souza, B. D. Haughom, T. Baum, M. Nevitt, C. McCulloch, J. A. Lynch, and T. M. Link. Physical activity is associated with magnetic resonance imaging–based knee cartilage t2 measurements in asymptomatic subjects with and those without osteoarthritis risk factors. *Arthritis & Rheumatology*, 63(8):2248–2256, 2011.

HP. Hp, 2018. available in `https://www.hp.com`, accessed 16-05-2018.

V. L. Johnson and D. J. Hunter. The epidemiology of osteoarthritis. *Best practice & research Clinical rheumatology*, 28(1):5–15, 2014.

A. M. Kandahari, X. Yang, A. S. Dighe, D. Pan, and Q. Cui. Recognition of Immune Response for the Early Diagnosis and Treatment of Osteoarthritis. *Journal of Imunology Research*, 2015, 2015.

M. Kaur, B. Singh, et al. Comparison of different approaches for removal of baseline wander from ecg signal. In *Proceedings of the International Conference & Workshop on Emerging Trends in Technology*, pages 1290–1294. ACM, 2011.

J. Khan, J. S. Wei, M. Ringner, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, and C.and others Peterson. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature medicine*, 7(6):673, 2001.

K. S. Kim, D. Y. Yoon, S. O. Lee, J. H. Seo, and C. G. Song. Classification of arthritic pathology using acoustic signal processing. 2005.

K. S. Kim, C. G. Song, and J. H. Seo. Feature extraction of knee joint sound for non-invasive diagnosis of articular pathology. *2008 IEEE-BIOCAS Biomedical Circuits and Systems Conference, BIOCAS 2008*, pages 349–352, 2008.

K. S. Kim, J. H. Seo, J. U. Kang, and C. G. Song. An enhanced algorithm for knee joint sound classification using feature extraction based on time-frequency analysis. *Computer Methods and Programs in Biomedicine*, 94(2):198–206, 2009a.

K. S. Kim, C. G. Song, and J. H. Seo. Non-invasive monitoring of knee pathology based on automatic knee sound classification. In *Proceedings of the World Congress on Engineering and Computer Science*, volume 1, 2009b.

K. S. Kim, C. G. Song, J. H. Seo, K. H. Park, J. Y. La, and J. C. Kim. An efficient algorithm to improve feature extraction and classification of knee joint sound. *2009 ICME International Conference on Complex Medical Engineering*, pages 1–6, 2009c.

K. S. Kim, J. H. Seo, and C. G. Song. An Acoustical Evaluation of Knee Sound for Non-invasive Screening and Early Detection of Articular Pathology. *J Med Syst*, 36:715–722, 2012.

S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3):159–190, 2006.

K. Kręcisz and D. Bączkowicz. Analysis and multiclass classification of pathological knee joints using vibroarthrographic signals. *Computer Methods and Programs in Biomedicine*, 154:37–44, 2018.

S. Krishnan. Adaptive Signal Processing Techniques for Analysis of Knee Joint Vibroarthrographic Signals, 1999.

S. Krishnan, R. M. Rangayyan, G. D. Bell, and C. B. Frank. Auditory display of knee-joint vibration signals. *The Journal of the Acoustical Society of America*, 110(6):3292, 2001.

D. Kumar, A.F.M. Hani, A.S. Malik, R. Kamil, R. Razak, and A. Kiflie. Development of a non-invasive diagnostic tool for early detection of knee osteoarhritis. *2011 National Postgraduate Conference - Energy and Sustainability: Exploring the Innovative Minds, NPC 2011*, 2011.

G. Kumar and P. K. Bhatia. A detailed review of feature extraction in image processing systems. In *Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on*, pages 5–12. IEEE, 2014.

L. I. Kuncheva and J. J. Rodríguez. A weighted voting framework for classifiers ensembles. *Knowledge and Information Systems*, 38(2):259–275, 2014.

D. T. L. Lee and A. Yamamoto. Wavelet analysis: theory and applications. *Hewlett Packard journal*, 45:44–44, 1994.

T. F. Lee, W. C. Lin, L. F. Wu, and H. Y. Wang. Analysis of vibroarthrographic signals for knee osteoarthritis diagnosis. *Proceedings - 2012 6th International Conference on Genetic and Evolutionary Computing, ICGEC 2012*, pages 223–228, 2012.

A. Legare, M. Garon, R. Guardo, P. Savard, A. R. Poole, and M. D. Buschmann. Detection and analysis of cartilage degeneration by spatially resolved streaming potentials. *Journal of orthopaedic research*, 20(4):819–826, 2002.

X. Li, C. B. Ma, T. M. Link, D. Castillo, G. Blumenkrantz, J. Lozano, J. Carballido-Gamio, M. Ries, and S. Majumdar. In vivo t 1ρ and t 2 mapping of articular cartilage in osteoarthritis of the knee using 3t mri. *Osteoarthritis and cartilage*, 15(7):789–797, 2007.

W. . Lin, T. F. Lee, S. Y. Lin, L. F. Wu, H. Y. Wang, L. Y. Chang, J. M. Wu, J. C. Jiang, C. C. Tuan, M. F. Horng, C. S. Shieh, and P. J. Chao. Non-invasive knee osteoarthritis diagnosis via vibroarthrographic signal analysis. *Journal of Information Hiding and Multimedia Signal Processing*, 5(3):497–507, 2014.

A. Litwic, M. H. Edwards, E. M. Dennison, and C. Cooper. Epidemiology and burden of osteoarthritis. *British medical bulletin*, 105(1):185–199, 2013.

H. Liu and H. Motoda. *Computational methods of feature selection*. CRC Press, 2007.

K. Liu, X. Luo, S. Yang, S. Cai, F. Zheng, and Y. Wu. Classification of knee joint vibroarthrographic signals using k-nearest neighbor algorithm. *Canadian Conference on Electrical and Computer Engineering*, pages 1–4, 2014.

L. S. Lohmander. Knee replacement for osteoarthritis: facts, hopes and fears. *Medicographia*, 35 (2):181–188, 2013.

L. S. Lohmander, M. W. Lark, L. Dahlberg, L. A. Walakovits, and H. Roos. Cartilage matrix metabolism in osteoarthritis: markers in synovial fluid, serum, and urine. *Clinical biochemistry*, 25(3):167–174, 1992.

K. Ludloff. Die auscultation der wirbelsacule des kruezbeines und des beckheus. *The Boston Medical and Surgical Journal*, 53(1197), 1906.

H. Madry, E. Kon, V. Condello, G. M. Peretti, M. Steinwachs, R. Seil, M. Berruto, and L. Engebretsen. Early osteoarthritis of the knee. *Knee Surgery, Sports Traumatology, Arthroscopy*, 2016.

S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, 41(12):3397–3415, 1993.

J. Martel-Pelletier, A. J. Barr, F. M. Cicuttini, P. G. Conaghan, C. Cooper, M. B. Goldring, S. R. Goldring, G. Jones, A. J. Teichtahl, and J. Pelletier. Osteoarthritis. *Nature Reviews Disease Primers*, 2:1–18, 2016.

B. Mascaro, J. Prior, L. K. Shark, J. Selfe, P. Cole, and J. Goodacre. Exploratory study of a non-invasive method based on acoustic emission for assessing the dynamic integrity of knee joints. *Medical Engineering and Physics*, 31(8):1013–1022, 2009.

MATLAB. Supervised machine learning, dimensional reduction and principal component analysis, 2015. available in https://hackernoon.com/supervised-machine-learning-dimensional-reduction-and-principal-component-analysis, accessed 16-05-2018.

MATLAB. Boxplot, 2018a. https://www.mathworks.com/help/stats/boxplot.html, accessed 26-01-2018.

MATLAB. Matlab®, 2018b. https://www.mathworks.com/products/matlab.html, accessed 26-01-2018.

MATLAB. sequentialfs, 2018c. available in https://nl.mathworks.com/help/stats/sequentialfs.html, accessed 16-05-2018.

MATLAB. Supervised learning workflow and algorithms, 2018d. available in https://www.mathworks.com/help/stats/supervised-learning-machine-learning-workflow-and-algorithms.html, accessed 16-05-2018.

MATLAB. Treebagger, 2018e. https://www.mathworks.com/help/stats/treebagger.html, accessed 26-01-2018.

MayfieldClinic. Single photon emission computed tomography (spect), 2016. https://d3djccaurgtij4.cloudfront.net/PE-SPECT.pdf, accessed 15-05-2018.

S. McCarthy, P. O'Raghallaigh, S. Woodworth, Y. L. Lim, L. C. Kenny, and F. Adam. An integrated patient journey mapping tool for embedding quality in healthcare service reform. *Journal of Decision Systems*, 25(sup1):354–368, 2016.

N. M. Menezes, M. L. Gray, J. R. Hartke, and D. Burstein. T2 and t1ρ mri in articular cartilage systems. *Magnetic resonance in medicine*, 51(3):503–509, 2004.

J. Mitchell. An introduction to patient journey mapping, 2017. https://www.linkedin.com/pulse/introduction-patient-journey-mapping-john-mitchell/, accessed 10-05-2018.

D. B. F. Moreira. Classification of knee arthropathy with accelerometer-based vibroarthrography. Master's thesis, Faculty of Engineering of the University of Porto, 2015.

H. Nakamura, K. Masuko, K. Yudoh, T. Kato, K. Nishioka, T. Sugihara, and M. Beppu. Positron emission tomography with 18 f-fdg in osteoarthritic knee. *Osteoarthritis and cartilage*, 15(6):673–681, 2007.

S. Nalband, A. Sundar, A. A. Prince, and A. Agarwal. Feature selection and classification methodology for the detection of knee-joint disorders. *Computer methods and programs in biomedicine*, 127:94–104, 2016.

NationalInstruments. Understanding basic statistical values (advanced signal processing toolkit), 2018. available in http://zone.ni.com/reference/en-XX/help/371419D-01/lvasptconcepts/tsa_basic_stat_analysis/, accessed 16-05-2018.

E. Nüesch, P. Dieppe, S. Reichenbach, S. Williams, S. Iff, and P. Jüni. All cause and disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. *Bmj*, 342, 2011.

University of Oxford. The oxford knee score (oks), 2018. available in https://innovation.ox.ac.uk/outcome-measures/oxford-knee-score-oks/, accessed 16-05-2018.

University of Strasbourg. Bagging ensemble c4 c2 c8 c1 learning algorithm model m1 s1, 2018. available in http://slideplayer.com/slide/3312931/11/images/15/Bagging+ENSEMBLE+C4+C2+C8+C1+Learning+algorithm+Model+M1+S1.jpg, accessed 16-05-2018.

J. Oleinik. It's not always due to sjogren's syndrome, 2013. available in http://www.reasonablywell.net/2013/01/its-not-always-due-to-sjogrens-syndrome.html, accessed 14-05-2018.

World Health Organization. Bmi classification, 2018. available in http://apps.who.int/bmi/index.jsp?introPage=intro_3.html, accessed 16-05-2018.

Orthopod. Knee anatomy, 2016. https://eorthopod.com/knee-anatomy/, accessed 20-01-2018.

C. Palazzo, C. Nguyen, F. Lefevre-Colau, M.and Rannou, and S. Poiraudeau. Risk factors and burden of osteoarthritis. *Annals of physical and rehabilitation medicine*, 59(3):134–138, 2016.

V. K. Patel and M. N. Patel. Development of smart sensing unit for vibration measurement by embedding accelerometer with the arduino microcontroller. *International Journal of Instrumentation Science*, 6(1):1–7, 2017.

V. R. Patel and R. G. Mehta. Impact of outlier removal and normalization approach in modified k-means clustering algorithm. *IJCSI International Journal of Computer Science Issues*, 8(5), 2011.

D. Patra and L. J. Sandell. Recent advances in biomarkers in osteoarthritis. *Current opinion in rheumatology*, 23(5):465–470, 2011.

M. Pérez-Ortiz, S. Jiménez-Fernández, P. A. Gutiérrez, E. Alexandre, C. Hervás-Martínez, and S. Salcedo-Sanz. A review of classification problems and algorithms in renewable energy applications. *Energies*, 9(8):607, 2016.

T. N. Phyu. Survey of classification techniques in data mining. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1, pages 18–20, 2009.

T. C. B. Pollard, S. E. Gwilym, and A. J. Carr. The assessment of early osteoarthritis. *Bone & Joint Journal*, 90(4):411–421, 2008.

A. M. Préville, P. Lavigne, M. D. Buschmann, J. Hardin, Q. Han, L. Djerroud, and P. Savard. Electroarthrography: A novel method to assess articular cartilage and diagnose osteoarthritis by non-invasive measurement of load-induced electrical potentials at the surface of the knee. *Osteoarthritis and Cartilage*, 21(11):1731–1737, 2013.

J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

R. M. Rangayyan. Index of signal data files, 2006. available in `http://people.ucalgary.ca/~ranga/enel563/SIGNAL_DATA_FILES/`, accessed 16-05-2018.

R. M. Rangayyan and Y. Wu. Analysis of vibroarthrographic signals with features related to signal variability and radial-basis functions. *Annals of Biomedical Engineering*, 37(1):156–163, 2009a.

R. M. Rangayyan and Y. Wu. Analysis of vibroarthrographic signals with features related to signal variability and radial-basis functions. *Annals of biomedical engineering*, 37(1):156–163, 2009b.

R. M. Rangayyan and Y. Wu. Screening of knee-joint vibroarthrographic signals using probability density functions estimated with Parzen windows. *Biomedical Signal Processing and Control*, 5(1):53–58, 2010.

R. M. Rangayyan and Y. F. Wu. Screening of knee-joint vibroarthrographic signals using statistical parameters and radial basis functions. *Medical and Biological Engineering and Computing*, 46 (3):223–232, 2008.

R. M. Rangayyan, S. Krishnan, G. D. Bell, C. B. Frank, and K. O. Ladly. Parametric representation and screening of knee joint vibroarthrographic signals. *IEEE Transactions on Biomedical Engineering*, 44(11):1068–1074, 1997.

R. M. Rangayyan, F. Oloumi, Y. Wu, and S. Cai. Fractal analysis of knee-joint vibroarthrographic signals via power spectral analysis. *Biomedical Signal Processing and Control*, 8(1):23–29, 2013.

RaspberryPi. Raspberry pi™, 2018. available in `https://www.raspberrypi.org/`, accessed 26-01-2018.

B. D. Ripley. *Pattern recognition and neural networks*. Cambridge university press, 2007.

D. G. E. Robertson and J. J. Dowling. Design and responses of butterworth and critically damped digital filters. *Journal of Electromyography and Kinesiology*, 13(6):569–573, 2003.

J. F. A. Romero, M. S. D'angelo, O. Saotome, and S. F. Müller. Wavelet packet feature extraction for vibration monitoring in high speed milling. submitted to the 18th International Congress of Mechanical Engineering, Ouro Preto, 2005.

C. Roubille, J. Martel-Pelletier, and J. P. Pelletier. Osteoarthritis treatments: where do we stand at the moment? *Medicographia*, 35(2):172–180, 2013.

N. Sánchez-Maroño, A. Alonso-Betanzos, and M. Tombilla-Sanromán. Filter methods for feature selection–a comparative study. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 178–187. Springer, 2007.

R. J. Schilling, J. J. Carroll, and A. F. Al-Ajlouni. Approximation of nonlinear systems with radial basis function neural networks. *IEEE Transactions on neural networks*, 12(1):1–15, 2001.

E. Schwartz. Mapping the patient journey: A case study, 2017. `https://www.tandemseven.com/blog/mapping-patient-journey-case-study/`, accessed 10-05-2018.

G. H. Seng, T. T. Swee, and H. Y. Chai. Parametric Study of Vibroarthrographic Signal Characteristic for Future Osteoarthritis Analysis. *Recent Advances in Telecommunications, Signals and Systems*, pages 40–45, 2013.

P. Senin. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, 855:1–23, 2008.

L. K. Shark, H. Chen, and J. Goodacre. Knee acoustic emission: A potential biomarker for quantitative assessment of joint ageing and degeneration. *Medical Engineering and Physics*, 33 (5):534–545, 2011.

R. Shenoy, P. S. Pastides, and D. Nathwani. (iii) biomechanics of the knee and tkr. *Orthopaedics and Trauma*, 27(6):364–371, 2013.

K. Singh, P. Abrol, and N. Rathi. Review on digital stethoscope. *International Journal of Soft Computing and Engineering (IJSCE)*, 3:96–97, 2013.

K. Sinusas. Osteoarthritis: Diagnosis and Treatment. *American Family Physician*, 1(86):49–56, 2012.

M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009.

SolidWorks. Solidworks®, 2018. `http://www.solidworks.com/`, accessed 26-01-2018.

Sparkfun. Accelerometer basics, 2018a. available in `https://learn.sparkfun.com/tutorials/accelerometer-basics`, accessed 16-05-2018.

Sparkfun. Teensy™3.1, 2018b. available in `https://www.sparkfun.com/products/retired/12646`, accessed 26-01-2018.

I. Steinwart and A. Christmann. *Support vector machines*. Springer Science & Business Media, 2008.

J. A. K. Suykens, T. Van Gestel, and J. De Brabanter. *Least squares support vector machines*. World Scientific, 2002.

N. Tanaka and M. Hoshiyama. Vibroarthrography in patients with knee arthropathy. *Journal of back and musculoskeletal rehabilitation*, 25(2):117–122, 2012.

TEconnectivity. Contact microphone cm-01b, 2017. available in `http://www.te.com/commerce/DocumentDelivery/DDEController?Action=showdoc&DocId=Data+Sheet%7FContact_Microphone_CM-01B%7FA%7Fpdf%7FEnglish%7FENG_DS_Contact_Microphone_CM-01B_A.pdf%7FCAT-PFS0013`, accessed 16-05-2018.

C. J. Tiderius, L. E. Olsson, P. Leander, O. Ekberg, and L. Dahlberg. Delayed gadolinium-enhanced mri of cartilage (dgemric) in early knee osteoarthritis. *Magnetic resonance in medicine*, 49(3):488–492, 2003.

C. M. Torio, , and B. J. Moore. National inpatient hospital costs: The most expensive conditions by payer, 2013. *Healthcare Cost and Utilization Project*, (204), 2016.

UCDavis. Introduction to digital filters, 2014. available in `http://123.physics.ucdavis.edu/week_5_files/filters/digital_filter.pdf`, accessed 16-05-2018.

K. T. Ulrich. *Product design and development*. Tata McGraw-Hill Education, 2003.

P. Vaish. k-nearest neighbors(knn), 2016. available in `http://adataanalyst.com/machine-learning/knn/`, accessed 16-05-2018.

K. K. Vasan and B. Surendiran. Dimensionality reduction using principal component analysis for network intrusion detection. *Perspectives in Science*, 8:510–512, 2016.

N. K. Visalakshi and K. Thangavel. Impact of normalization in distributed k-means clustering. *international Journal of Soft computing*, 4(4):168–172, 2009.

S. E. Vt and Y. C. Shin. Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems. *IEEE Transactions on Neural Networks*, 5(4):594–603, 1994.

I. J. Wallace, S. Worthington, D. T. Felson, R. D. Jurmain, K. T. Wren, H. Maijanen, R. J. Woods, and D. E. Lieberman. Knee osteoarthritis has doubled in prevalence since the mid-20th century. *Proceedings of the National Academy of Sciences*, 114(35):9332–9336, 2017.

C. F. Walters. The value of joint auscultation. *The Lancet*, 213(5514):920–921, 1929.

WebMD. Why knee hurts, 2018. available in `https://www.webmd.com/pain-management/knee-pain/ss/slideshow-why-knees-hurt`, accessed 14-05-2018.

H. A. Wieland, M. Michaelis, B. J. Kirschbaum, and K. A. Rudolphi. Osteoarthritis—an untreatable disease? *Nature reviews Drug discovery*, 4(4):331, 2005.

A. Williams, Y. Qian, D. Bear, and C. R. Chu. Assessing degeneration of human articular cartilage with ultra-short echo time (ute) t 2* mapping. *Osteoarthritis and cartilage*, 18(4):539–546, 2010.

W. Y. Wong, M. S. Wong, and K. H. Lo. Clinical applications of sensors for human posture and movement analysis: a review. *Prosthetics and orthotics international*, 31(1):62–75, 2007.

A. D. Woolf and B. Pfleger. Burden of major musculoskeletal conditions. *Bulletin of the World Health Organization*, 81(9):646–656, 2003.

Y. Wu. *Knee Joint Vibroarthrographic Signal Processing Analysis*. Springer Briefs in Bioengineering, 2015. ISBN 9783662442838.

Y. Wu, S. Krishnan, and R. M. Rangayyan. Computer-aided diagnosis of knee-joint disorders via vibroarthrographic signal analysis: a review. *Critical Reviews™ in Biomedical Engineering*, 38 (2), 2010.

Y. Wu, S. Cai, S. Yang, F. Zheng, and N. Xiang. Classification of Knee Joint Vibration Signals Using Bivariate Feature Distribution Estimation and Maximal Posterior Probability Decision Criterion. *Entropy*, 15(4):1375–1387, 2013.

Y. Wu, S. Yang, F. Zheng, S. Cai, M. Lu, and M. Wu. Removal of artifacts in knee joint vibroarthrographic signals using ensemble empirical mode decomposition and detrended fluctuation analysis. *Physiological Measurement*, 35(3):429–439, 2014.

Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 1(01):1–41, 2009.

H. Zhang. The optimality of naive bayes. *AA*, 1(2):3, 2004.

J. Zhang, J. Zhang, T. Lok, and M. R. Lyu. A hybrid particle swarm optimization–backpropagation algorithm for feedforward neural network training. *Applied mathematics and computation*, 185(2):1026–1037, 2007.

Y. Zhang, H. Zhang, J. Cai, and B. Yang. A weighted voting classifier based on differential evolution. In *Abstract and Applied Analysis*, volume 2014. Hindawi, 2014.

L. Zhu, E. Garon, M.and Quenneville, M. D. Buschmann, and P. Savard. Electrical potentials measured on the surface of the knee reflect the changes of the contact force in the knee joint produced by postural sway. *Gait and Posture*, 52:159–164, 2017.

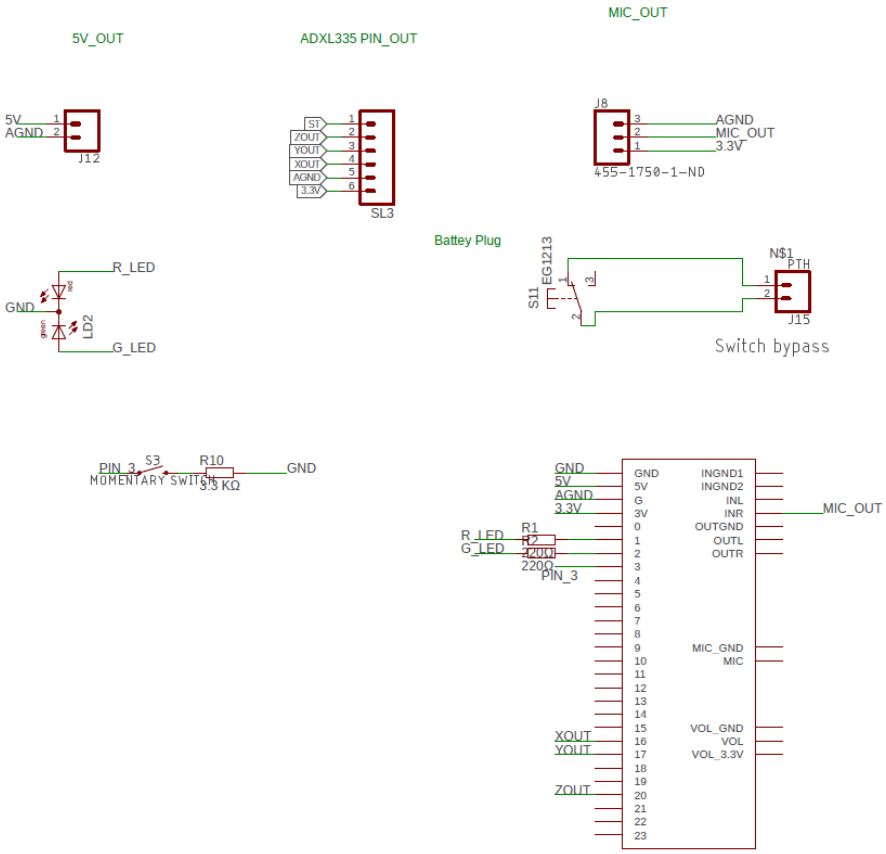
ZimmerBiomet. Zimmer biomet, 2018. available in `https://www.zimmerbiomet.com/`, accessed 16-05-2018.

Appendix A

OASense

A.1 Electronics schematics

This schematics was done using Auto Desk EAGLE (Autodesk, 2018) software. It is possible to see the different components connected to the Teensy, namely the accelerometer (ADXL335), the microphone, the RG-LED, the battery, the charger and the two buttons.



Appendix B

Open-source VAG dataset

B.1 Signal's array

In this figure, it is possible to see that feature $TF_{mwc}5$, when trained with k-NN (k=3), provides different AUC values. Thus, the computation of this array of signals is important, since some features are more discriminant in some signals, when comparing to others.

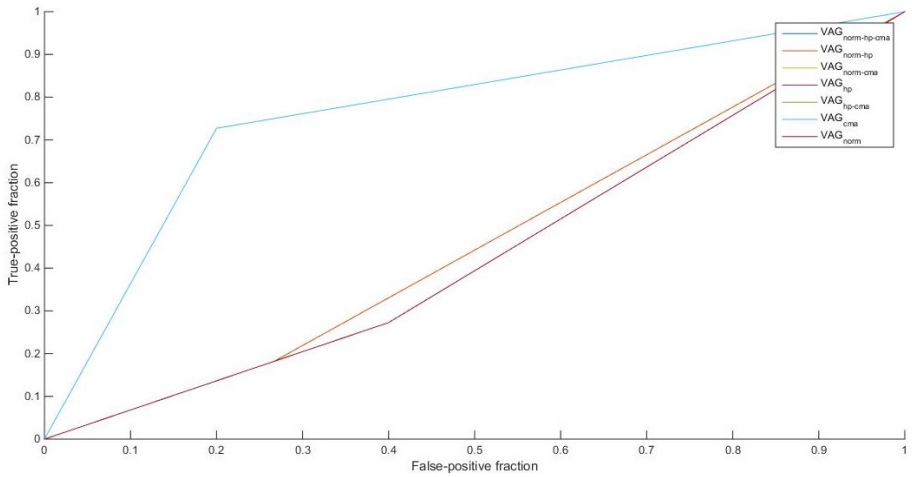


Figure B.1: AUC values of 0.46 for the $VAG_{norm-hp-cma}$ and $VAG_{norm-hp}$ signals, 0.44 for the $VAG_{norm-cma}$ and VAG_{norm} and 0.76 for the VAG_{hp}, VAG_{hp-cma} and VAG_{cma} signals were obtained training a k-NN model (k=3) with the feature $TF_{mwc}5$.

B.2 Skewness

As it can be depicted below, the PDF of the normal signal is concentrated on the right of the mean, which implies a negative value for the skewness. On the other hand, the PDF of the abnormal is mostly on the left of its mean, resulting in a positive skewness.

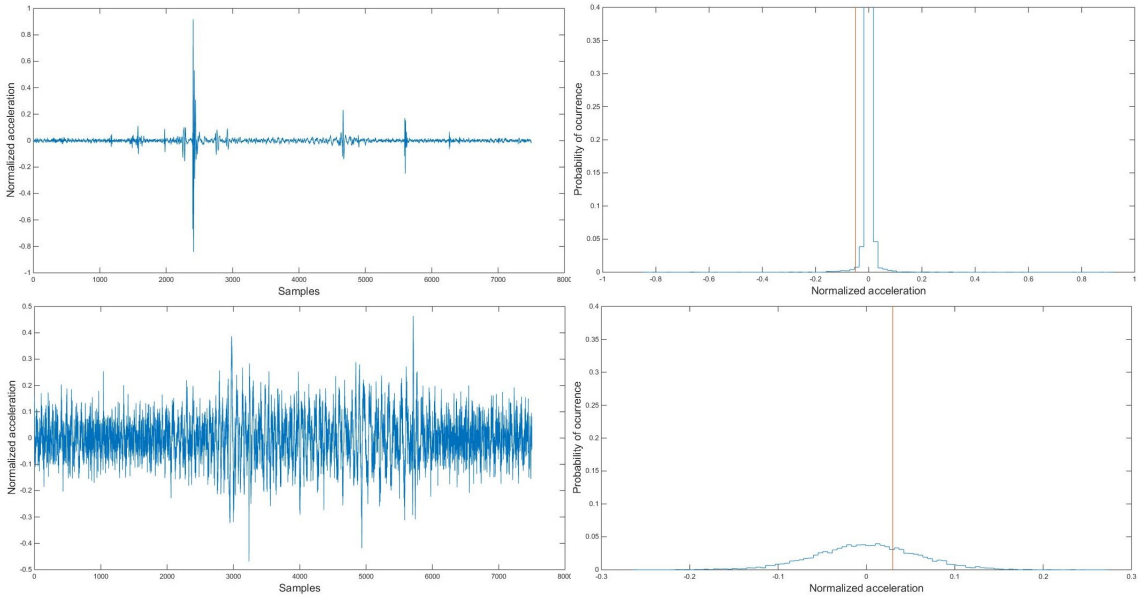


Figure B.2: High-pass filtered and normalized normal (top) and abnormal (bottom) VAG signals, with the mean of the signal in orange. The normalized histogram (estimated with L=100 bins) is also represented at the right of each signal.

Appendix C

OASense collected signals

C.1 Questionnaire for participants

The first page of the questionnaire includes sociodemographic data (age, gender, weight and height), as well as occupation, physical activity and history of knee articular diseases. The OA unaffected participants shouldn't answer to the question regarding the affected joint. The two following pages have the 12 OKS questions, which allow the evaluation of knee related symptoms.

Participant questionnaire

Today's date:
___ - ___ - _____
DD MM YYYY

Age ____
Gender F ☐ M ☐
Weight _____ kg
Height _____ cm

Occupation (previous and current, and for how long):
_____

Physical activity (previous and current, and for how long):
_____

On which side of your body is the affected joint:
Left ☐
Right ☐
Both ☐
If you said 'Both', please complete the first questionnaire thinking about the right side. A second questionnaire, for the left side, will follow (if you are willing to do the experiment on both knees).

Have you haver had one of these problems before? If yes, when?
Intra-articular fracture ☐
Dislocation ☐
Anterior cruciate ligament rupture ☐
Meniscal fracture ☐
Other. Which one? _____

Were you affected by a knee articular disease before (e.g. Rheumathoid arthritis)? Y ☐ N ☐

Were you subjected to a total/partial knee replacement? Y ☐ N ☐

1. During the past 4 weeks…

How would you describe the pain you usually have from your knee?

None	Very mild	Mild	Moderate	Severe
☐	☐	☐	☐	☐

2. During the past 4 weeks…

Have you had problems with your knee while performing your general daily life activities (ADLs) such as eating, drinking, showering, dressing and undressing, etc?

No trouble at all	Very little trouble	Moderate trouble	Extreme difficulty	Impossible to do
☐	☐	☐	☐	☐

3. During the past 4 weeks…

Have you had any trouble getting in and out of a car or using public transport because of your knee? (whichever you would tend to use)

No trouble at all	Very little trouble	Moderate trouble	Extreme difficulty	Impossible to do
☐	☐	☐	☐	☐

4. During the past 4 weeks…

For how long have you been able to walk before pain from your knee becomes severe? (with or without a stick)

No pain/more than 30 minutes	16 to 30 minutes	5 to 15 minutes	Around the house only	Not at all/ severe pain
☐	☐	☐	☐	☐

5. During the past 4 weeks…

Is it painful to get up from a sitting position?

Not at all painful	Slightly painful	Moderately painful	Very painful	Unbearable
☐	☐	☐	☐	☐

6. During the past 4 weeks…

Did you have to adjust your walking position because of pain in your knee?

Rarely/never	Sometimes, or just at first	Often, not just at first	Most of the time	All of the time
☐	☐	☐	☐	☐

7. During the past 4 weeks…

Could you knee down and get up again afterwards?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐	☐	☐	☐	☐

8. During the past 4 weeks…

Have you been troubled by pain from your knee in bed at night?

No nights	1 or 2 nights	Some nights	Most nights	Every night
☐	☐	☐	☐	☐

9. During the past 4 weeks…

How much has pain from your knee interfered with your usual work (including housework)?

Not at all	A little bit	Moderately	Greatly	Totally
☐	☐	☐	☐	☐

10. During the past 4 weeks…

Did you feel that your knee is failing during strenuous activities?

Rarely/never	Sometimes or just at first	Often, not just at first	Most of the time	All of the time
☐	☐	☐	☐	☐

11. During the past 4 weeks…

Could you do the household shopping on your own?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐	☐	☐	☐	☐

12. During the past 4 weeks…

Could you walk down one flight of stairs?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐	☐	☐	☐	☐

Finally, please check back that you have answered each question.

Thank you very much for your collaboration!

C.2 Inclusion and exclusion criteria

Both OA patients and unaffected participants have to be above the age of majority in the Netherlands (18) and have to be willing to participate in the study. OA patients need to have an X-ray image previously analysed by a doctor, which proves they are affected by Osteoarthritis. Also, the OA knee that is being measured should not have been subjected to surgery or be affected by any other type of articular disease previous to the experiment.

The inclusion and exclusion criteria for the unaffected participants include the OA risk factors. Namely, healthy participants should be below 45 years old, since this is the age when the OA risk increases exponentially, according to Palazzo et al. (2016). The OKS score should be above 45, which reflects the absence of OA related symptoms. Also, the Body Mass Index (BMI), which constitutes a simple index of weight-for-height that is commonly used to classify underweight, overweight and obesity in adults, should be below 30 (Organization, 2018). Above this number, the person is considered obese, which increases the risk of having OA, even if the symptoms are not noticeable yet. Participants should also have no surgical or knee articular disease history.

If the participant has/had an occupation which involves repetitive knee bending and squatting, such as construction workers, professional athletes, farmers, among others, he/she should be excluded. Also, if the subject does/did any kind of high-intensity sports for a long period, such as soccer, basketball, jumping sports and running, he/she should not be considered for the study. These participants should be excluded, because although they might not feel the symptoms yet, they could have a knee articular disease in the near future and their VAG signals can already present some alterations.

Table C.1: Inclusion and exclusion criteria

	OA patients	**Unaffected participants**
Inclusion	Age>18 Willingness to participate X-ray data (diagnosed OA)	18<Age<45 Willingness to participate OKS>45 BMI<30
Exclusion	Surgical history (e.g. TKR) History of other knee articular disease (e.g. RA)	History of any knee joint pathologies Risky occupation and/or physical activity