UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

PosturAll: A Posture Assessment Software for Children

Ana Beatriz Góis Ferreira Gaspar Neves

Mestrado Integrado em Engenharia Biomédica e Biofísica

Engenharia Clínica e Instrumentação Médica

Dissertação orientada por:

Professor Doutor Nuno Miguel de Pinto Lobo Matela

Professor Doutor Tiago João Viegas Atalaia

2022

Agradecimentos

Em primeiro lugar quero agradecer aos meus orientadores Professor Tiago Atalaia e Professor Nuno Matela. Pela orientação, apoio e contributos decisivos que me deram ao longo do desenvolvimento deste trabalho, por toda a ajuda na solução dos vários problemas e respostas às minhas dúvidas.

Gostaria de agradecer também à Escola Superior de Saúde da Cruz Vermelha Portuguesa de Lisboa por me permitirem utilizar as suas infraestruturas e equipamento para a realização deste estudo. Bem como, agradeço a todos os alunos desta instituição que se voluntariaram para participar no estudo que permitiu elaborar esta dissertação.

Quero agradecer também a todos os meus amigos que me apoiaram ao longo da realização deste projeto. Em especial, à Ema, ao João, à Sara e à Soraia por se terem disponibilizado a participarem neste trabalho.

Finalmente, quero agradecer a toda a minha família que me deram sempre apoio e carinho. Em especial aos meus pais, irmãos e avós que me sempre incentivaram durante este percurso, e que acreditaram em mim tornando possível os meus estudos.

Resumo

Os problemas e distúrbios músculo-esqueléticos como a dor lombar, a dor cervical e a escoliose, são um frequente fator de perturbação da qualidade de vida e da atividade laboral com elevados custos diretos e indiretos. Em 2019, 37.3 % da população portuguesa reportou sofrerem de dores lombares ou outros problemas crónicos associados à dor nas costas, e 27.1% da população portuguesa afirmou ter dor na cervical ou outros problemas crónicos associados à dor no pescoço. Contudo, com uma deteção precoce durante a infância e adolescência, pode ser possível reduzir a sua incidência e impacto.

Desde cedo, as crianças e adolescentes são expostas a estímulos e fatores de risco que podem levar ao desenvolvimento destes distúrbios músculo-esqueléticos. Alguns destes fatores de risco incluem ser-se do género feminino, o aumento da idade, o sedentarismo ou a prática intensiva de exercício físico, o peso e modelo de mochila utilizada na escola, e os comportamentos posturais, como a postura adotada enquanto se está sentado na sala de aula ou sentado a ver televisão ou a jogar videojogos, ou a postura de flexão da cervical quando se utiliza o telemóvel.

Para minimizar estes fatores de risco é importante a sensibilização da população para os problemas músculo-esqueléticos, por exemplo através de campanhas que promovam comportamentos saudáveis que as pessoas devem adotar desde jovens, bem como, implementar avaliações médicas regulares nas escolas para a deteção precoce dos distúrbios músculo-esqueléticos.

Atualmente, existem diversos métodos de avaliação postural e de programas multidisciplinares de tratamento da dor, no entanto a maioria está acessível apenas a profissionais de saúde em meio clínico com custos elevados. Recentemente, têm começado a aparecer no mercado aplicações para telemóvel que incluem tratamento multidisciplinar da dor e que analisam de forma simplificada a postura da pessoa por um preço acessível.

O objetivo desta dissertação é melhorar um software anteriormente desenvolvido por estudantes do mestrado integrado em engenharia biomédica e biofísica, e posteriormente testar a sua performance e validar os resultados obtidos com esse software usando uma população não-patológica. Com os dados adquiridos iremos num último passo analisar a performance de classificadores. Este software processa imagens de 4 vistas diferentes do paciente (vista anterior, posterior, lateral esquerda e direita), enquanto este utiliza diversos marcadores em locais anatómicos específicos. Com os marcadores, o software calcula diversos ângulos e comprimentos de segmentos corporais, que usa para obter um valor de indicador de risco para possíveis problemas e distúrbios da zona da lombar, cervical e escoliose.

A realização deste projeto foi dividida em quatro fases. Primeiro, começámos por selecionar os marcadores e as métricas corporais a usar na avaliação postural, em conjunto com especialistas de fisioterapia e com base em literatura científica, do qual resultaram 29 marcadores e 38 parâmetros. Após esta seleção, modificámos o software para incluir o cálculo destes parâmetros. De seguida, acrescentámos a utilização de um retângulo de comprimento conhecido para se aferir a escala da imagem e um fio com um peso na ponta para se obter a linha vertical ao chão. Também modificámos a forma de deteção dos marcadores pelo software, tornando possível a deteção em imagens provenientes de diferentes tipos de câmaras, através do método *blob detection.*

Na segunda fase, aferimos a performance do software e efetuámos testes de calibração. Nestes passos os marcadores foram colocados num manequim. Começámos por analisar o uso de marcadores de diferentes cores e tamanhos para percebermos quais os mais adequados. Os diferentes marcadores foram colocados em 4 sujeitos de alturas e volumes diferentes e com roupas de cores diferentes. Desta análise ficou definido a utilização de marcadores verdes de 20 mm de diâmetro, bem como a cor azul

para o retângulo que permite obter a escala da imagem, e a cor vermelha para o fio usado para determinar a linha vertical ao chão.

Após este passo, avaliámos a capacidade de o software detetar os marcadores verdes, o retângulo azul e o fio vermelho utilizando diferentes cores de fundo e com dois tipos de câmaras diferentes. Dos fundos usados foi possível detetar os marcadores e o retângulo em todos, no entanto, o fio vermelho não foi possível detetar quando o fundo utilizado era vermelho.

Nos testes de calibração, verificámos a influência da utilização ou não de flash, da variação da distância da câmara aos marcadores, da variação da altura da câmara e da variação do tipo de câmara utilizada. Concluímos que a utilização ou não de flash é a variável que menos influenciou os resultados, que no geral houve maior variação dos valores obtidos quando utilizada uma câmara fotográfica face a uma câmara de telemóvel, e que os parâmetros anatómicos que utilizam o fio vermelho no seu cálculo tiveram maiores variações do que os parâmetros que usam o retângulo azul. O fio vermelho foi colocado num plano diferente dos marcadores levando a erros de paralaxe mais acentuados. No entanto, mesmo colocando o retângulo ao lado dos marcadores os erros de paralaxe não foram totalmente corrigidos, pois o volume do manequim faz com que os marcadores não fiquem todos no mesmo plano.

Numa terceira fase, adquirimos dados de uma população de estudo composta por estudantes da Escola Superior de Saúde da Cruz Vermelha Portuguesa - Lisboa. Para comparação e validação dos resultados obtidos, as imagens capturadas de cada participante foram analisadas com o nosso software e com um software clínico comercial de análise postural, o software Templo® no sistema da *Contemplas*, GMHB. Os dados de ambos os softwares foram comparados através de testes de hipóteses de comparação de médias e testes de equivalência de médias. Para a maioria dos parâmetros anatómicos, observámos nos testes estatísticos de comparação que as diferenças entre médias não eram significativas, e nos testes de equivalência obtivemos um valor das diferenças das médias entre -0.05 e 0.05, ambos para uma confiança de 95%.

Seguidamente, com base na avaliação postural feita pelo software clínico e por especialistas de fisioterapia, considerámos uma população composta por 16 sujeitos sem evidências de problemas posturais, 25 sujeitos que apresentavam evidências ligeiras e 16 sujeitos com evidências moderadas a acentuadas. Com os dados desta população avaliámos o desempenho de classificadores, em que implementámos os modelos Linear Discriminant Analysis e k-Nearest Neighbors, cada um com o método de validação divisão treino/teste e com o método de validação cruzada 10-fold.

Inicialmente considerámos uma classificação multiclasse, e posteriormente uma classificação binária de dois níveis. O primeiro nível classifica os sujeitos entre sem evidências e com evidências de problemas posturais, e o segundo nível divide este segundo grupo em sujeitos com evidências ligeiras e sujeitos com evidências moderadas a acentuadas. No geral, a classificação binária apresentou melhores valores que a classificação multiclasse. Em ambos os níveis da classificação binária, os valores de f1-score foram mais altos para as classes compostas por um maior número de sujeitos, o que pode ser justificado por serem classes com mais dados para treinar o classificador. Adicionalmente, verificámos que obtivemos valores mais altos com o uso do método de validação divisão treino/teste do que com o método de validação cruzada 10-fold. Isto pode ser explicado pela base de dados ser composta por poucos participantes, em que cada fold tem no máximo dados de 6 participantes, podendo levar a uma má distribuição dos dados.

Em projetos futuros, é necessário melhorar a forma de corrigir os erros de paralaxe inerentes à captura de imagens, possivelmente com a identificação do ponto de fuga da imagem, e, sabendo o campo de visão ou o tamanho do sensor, tornar menor a influência da utilização de diferentes tipos de câmara nos resultados. Além disso, no futuro, pode testar-se a utilização de um setup mais simplificado,

juntando a função do retângulo azul e do fio vermelho num só objeto, por exemplo, um esquadro com um peso na ponta. Outro passo a realizar é tornar este software, que ainda está num modo offline, numa aplicação para telemóvel, que apenas armazene e processe os resultados da análise do paciente num servidor seguro, e que elimine automaticamente as suas imagens.

Nesta dissertação, os participantes da população de estudo eram maiores de 18 anos, no entanto, é necessário testar o software numa população constituída por crianças e adolescentes para se começar a criar uma base de dados com valores de referência adaptados a essas idades, que são o público-alvo da aplicação.

Em conjunto com este software, no futuro pretende-se desenvolver um kit de avaliação postural que é composto por uma mochila que contenha no seu interior um arnês ajustável de corpo inteiro com os marcadores costurados, o objeto necessário para a obtenção da escala da imagem e da linha vertical ao chão, e marcas para serem colocadas no chão para marcar a posição em que o sujeito e a câmara devem estar. O objetivo final é implementar a utilização deste kit em escolas e associações desportivas juvenis, incentivando a um rastreio desde cedo na vida das pessoas, permitindo uma diminuição da probabilidade do aparecimento de distúrbios musculoesqueléticos ou a progressão destes para um estado crónico na vida adulta.

Palavras-chaves: Avaliação Postural Infantil, Distúrbios Músculo-esqueléticos, Análise de Métricas Anatómicas, Aprendizagem Automática, Visão Computacional

Abstract

Musculoskeletal problems and disorders such as low back pain, neck pain and scoliosis are a frequent factor that disturbs quality of life and work activity with high direct and indirect costs. In 2019, 37.3% of the Portuguese population reported suffering from low back pain or other chronic problems associated with back pain, and 27.1% reported having cervical pain or other chronic problems associated with neck pain. However, with early detection during childhood and adolescence, it might be possible to reduce its incidence and impact.

From an early age, children and adolescents are exposed to risk factors that can lead to the development of such musculoskeletal disorders. Some of those risk factors include being female, increasing age, physical inactivity or intensive physical exercise, the weight and type of backpack used at school, and postural behaviors, such as the posture adopted while sitting in the classroom or watching TV or playing video games, or the cervical flexion posture when using the cell phone.

To minimize the impact of those factors, it is important to raise the population's awareness of musculoskeletal problems, for example through campaigns that promote healthy behaviors, or implementing routine medical evaluations in schools for an early detection of musculoskeletal disorders.

Nowadays, there are several methods of postural assessment and multidisciplinary pain treatment programs, however most are accessible only to health professionals in high cost clinical environment. Recently, mobile applications started to appear on the market, which include multidisciplinary pain treatment and analyzes a person's posture.

The aim of this dissertation is to improve a software developed by students of the Integrated Master in Biomedical and Biophysics at FCUL, and later to test its performance and validate the results obtained using a non-pathological population. In a last step, we will analyze the performance of classifiers with the acquired data. This software processes images from 4 different views of the patient (anterior, posterior, left and right lateral views), while the patient uses different markers at specific anatomical locations to obtain a risk indicator value for possible musculoskeletal problems and disorders.

The methodology of this project was divided into four phases. First, with physiotherapy specialists and based on scientific literature, we select the markers and body metrics to be used in the postural assessment, resulting in 29 markers and 38 parameters. Then, we modified the software to include the calculation of these parameters. We also used a rectangle of known length to obtain the scale of the image and a string with a weight at the end to get the vertical line to the floor. Furthermore, we modified the way for detecting the markers by the software, making possible to detect them in images from different types of cameras, using the blob detection method.

In the second phase, we evaluated the software’s performance and run calibration tests. We analyzed the use of markers of different colors and sizes to understand which ones were the most suitable. The different markers were placed on 4 subjects of different heights and volumes while wearing clothes with different colors. We choose the green markers with a diameter of 20 mm, and we defined blue for the rectangle to obtain the image scale, and red for the string to set the vertical line to the ground.

Afterwards, we assessed the software's ability to detect the green markers, the blue rectangle and the red string using different background colors and with two different camera types. Of the backgrounds used, it was possible to detect the markers and the rectangle in all of them, except that, the red string was not detected in the red background.

In the calibration tests, we verified the influence of the use or not of flash, the variation of the distance from the camera to the markers, the variation of the camera height and of the type of camera

used. We concluded that the use of flash had the least influence in the results and in general there was a greater variation in the values obtained when using a photographic camera compared to a cell phone. Also, the anatomical parameters calculated using the red string had a greater variation than when using the blue rectangle. The red string was placed on a different plane from the markers leading to more pronounced parallax errors. Even placing the rectangle next to the markers, the parallax errors were not fully corrected, because the dress form has volume, so the markers are not all at the same plane.

In a third phase, we acquired data from a study population composed of students from the Escola Superior de Saúde da Cruz Vermelha Portuguesa - Lisboa. To validate the results obtained, the images captured from each participant were analyzed with our software and with a commercial clinical software for postural analysis, the software Templo® in the system *Contemplas*, GMHB. The data from both software were compared using hypothesis tests of comparison of means and equivalence tests of means. For most anatomical parameters, we observed with the statistical test of comparison of means that the differences between means were not significant, and in the equivalence tests we obtained a value of the differences in the means between -0.05 and 0.05, both for a confidence level of 95%.

Based on the postural assessment performed by clinical software and by physiotherapy specialists, we considered a population composed of 16 subjects with no evidence of postural problems, 25 subjects with mild evidence and 16 subjects with moderate to severe evidence. With the data collected, we evaluated the performance of classifiers by implementing the Linear Discriminant Analysis and k-Nearest Neighbors models, each with the train/test split validation method and the 10-fold cross-validation method.

Initially, we considered a multiclass classification, and later a two-level binary classification. The first level classifies the subjects between with and without evidence of postural problems, the second level divides the subjects with evidence into with mild evidence and with moderate to strong evidence. In general, the binary classification presented better values than the multiclass one. At both levels of binary classification, the f1-score values were higher for the classes with a greater number of subjects, probably due to more data to train the classifier. Additionally, we obtained higher values using the train/test split validation method than using the 10-fold cross-validation method. This can be explained by the fact that each fold has a maximum of 6 participants, leading to an inefficient distribution of data.

In future projects, it is necessary to improve the method to correct the parallax errors inherent to the images capture. For example, by identifying the vanishing point of the image, and, knowing the field of view or the size of the sensor, reduce the influence of using different types of cameras. Also, it can be tested using a simplified setup, by joining the function of the blue rectangle and the red thread in a single object. Another step is to turn this software, into a mobile application that only stores and processes the results of the patients on a secured server, and automatically deletes the images.

In this dissertation, the study population were over 18 years old, however, it is necessary to test the software in a population composed of children and adolescents to start creating a database with reference values adapted to those ages, which are the target audience of the application.

In the future, we intend, with this software, to develop an assessment kit that will be a backpack containing inside an adjustable full-body harness with the markers sewn on, the object necessary to obtain the scale from the image and the vertical line to the floor, and marks to be placed on the floor to define the position where the subject and the camera should be. The goal is to implement the use of this kit in schools and sports associations, encouraging a screening from an early age, allowing a decrease in the probability of the musculoskeletal disorders or their progression to a chronic state in adult life.

Keywords: Postural Assessment, Musculoskeletal Disorders, Anatomical Metric Analysis, Machine Learning, Computer Vision

Table of Contents

List of Figures

List of Tables

Acronyms

AAD_L	Left Acromion - ASIS Distance
AAD_R	Right Acromion - ASIS Distance
AHA_A	Acromions Horizontal Alignment in Anterior View
AHA_P	Acromions Horizontal Alignment in Posterior View
AIS	Adolescent idiopathic scoliosis
$AJND_L$	Left Acromion - Jugular Notch Distance
$AJND_R$	Right Acromion - Jugular Notch Distance
$AKLA_L$	Left Knee Lateral Angle in Anterior View
$AKLA_R$	Right Knee Lateral Angle in Anterior View
ALA_L	Left Ankle Lateral Angle
ALA_R	Right Ankle Lateral Angle
$ALLL_L$	Left Lower Limb Length in Anterior View
$ALLL_R$	Right Lower Limb Length in Anterior View
APD_L	Left Acromion - PSIS Distance
APD_R	Right Acromion - PSIS Distance
ASA	Acromions - Sternum Angle
ASIS	Anterior superior iliac spine
ASISHA	ASIS's Horizontal Alignment
$ASISLA_L$	ASIS's – Left Leg Angle
$ASISLA_R$	ASIS's – Right Leg Angle
AVA	Acromions - Vertebral Column Angle
C7	7th cervical vertebrae
Contemplas	Templo® from the *Contemplas* system, GMHB
eHealth	Electronic health
ESSCVP-Lisboa	Escola Superior de Saúde da Cruz Vermelha Portuguesa - Lisboa
EU	European Union
FCUL	Faculty of Sciences of the University of Lisbon
GJH	Generalized joint hypermobility
KA_L	Knee Angle in Left Lateral View
KA_R	Knee Angle in Right Lateral View
kNN	k-Nearest Neighbors
LBP	Low Back Pain

LDA	Linear discriminant analysis
LFA_L	Leg – Foot Angle in Left Lateral View
LFA_R	Leg – Foot Angle in Right Lateral View
LLA	Lumbar Lordosis Lateral Angle
LLC_L	Lumbar Lordosis Curvature in Left Lateral View
LLC_R	Lumbar Lordosis Curvature in Right Lateral View
mHealth	Mobile health
NP	Neck Pain
$PKLA_L$	Left Knee Lateral Angle in Posterior View
$PKLA_R$	Right Knee Lateral Angle in Posterior View
PLA_L	Pelvis – Leg Angle in Left Lateral View
PLA_R	Pelvis – Leg Angle in Right Lateral View
$PLLL_L$	Left Lower Limb Length in Posterior View
$PLLL_R$	Right Lower Limb Length in Posterior View
PSIS	Posterior superior iliac spine
PSISHA	PSIS's Horizontal Alignment
$PSISLA_L$	PSIS's – Left Leg Angle
$PSISLA_R$	PSIS's – Right Leg Angle
TKA	Thoracic Kyphosis Lateral Angle
TKC_L	Thoracic Kyphosis Curvature in Left Lateral View
TKC_R	Thoracic Kyphosis Curvature in Right Lateral View

1. Introduction

1.1. Motivation

The human being is, from an early age, exposed to innumerous amount of stimulus and constrictions given by the environment. Some can lead to musculoskeletal pain and disorders, such as neck and low back pain or scoliosis. Considering low back pain (LBP), one of the most common and expensive health care disorder of the industrialized countries in adulthood [1–4], frequently it has its onset in the adolescence, reaching the adult rates by the age of 22 years old [1,5,6], with evidence of quality of life implications [1,3,7]. LBP prevalence in school-age children is high [8], reaching up to 21% to 42% of adolescents reporting LBP episodes in their life's [9]. Another frequent pathology in this age, which also results in high costs to society, is scoliosis, which the causes can be idiopathic, related to joint hypermobility or due to postural behavior, among others [10–14]. Some of the same reasons are pointed out for LBP [1,5,6,15].

The Portuguese National Health Survey of 2019 [16] assessed the prevalence of chronic diseases in the resident population over 15 years old. This survey reveals that 37.3% of the general population, 43.0% of the women population and 30.9% of the men population reported LBP or other chronic back problems. Also, 27.1% of the general population, 33.8% of the women population and 19.5% of the men population reported NP or other chronic neck problems. Additionally, 12% and 6% of people aged between 15 and 24 years old and 34% and 24% of the population between 25 and 64 years old have problems with low back pain and with cervical pain, respectively [17,18].

Furthermore, in 2014, Portugal ranked 22 in the Lower Back Disorder or Other Chronic Back Defect within the European Union, having 25.8% of people between 15 to 64 years old affected, while the EU had 19.6%. Regarding the Neck Disorder or Other Chronic Neck Defect within the European Union, Portugal was ranked 23 with 17,6% of people between 15 to 64 years old affected, while the EU had 14.6% [19].

In addition to the high prevalence of these conditions, they are also an important issue for national health services, because they are one of the most expensive health care disorders, with a constant cost increasing path, especially LBP [1,14].

To adequately address this problem, early diagnosis is needed. This early detection of risk indicators can allow children and adolescents to access the help they need, an approach that might decrease the probability of adult LBP or scoliosis, or even the probability to evolve to a chronic state of these conditions at an early age [10,11]. These risk indicators serve only to point out children to be conducted to a more specialized analysis done by a health professional. This analysis can confirm or not the special attention needed and promote all the means to give the children proper orientations (educational, rehabilitative, lifestyle, etc.).

With this problem in mind, together with the *Escola Superior de Saúde da Cruz Vermelha Portuguesa* (ESSCVP-Lisboa), a software tool is being developed, to be mobile-phone operable, that computes several angles between body segments, with an easy-to-assemble postural assessment kit. The software will provide the scores of the risk factors for LBP or scoliosis to refer the children to a more specialized analysis by a health professional. This tool is developed in a way that facilitates its implementation in schools and youth academies of sports clubs to serve as a population screening in posture health, as others that are already working as oral health in children.

Regarding the possibility of application in schools, the Portuguese government already have another health program in schools, such as the program "Escolas a Sorrir" [20], which aims at dental

specialists to visit all schools that were enrolled in the program in order to educate good hygiene habits and prevent future serious dental problems. This measure was implemented due to Portugal's low position on the European Union's Oral Child Health list, with 6% of children less than 16 years old lacking dental care in Portugal (the average EU value corresponding to 2.8%). It was realized that the lack of such care until the age of 16, would lead to consequences that could be prevented if started at an earlier stage in a person's life.

1.2. Objectives

This project is a continuation of a work carried out by colleagues of the Master Program in Biomedical Engineering and Biophysics at FCUL. The software is based in image processing and metrics determination, which uses markers at intended anatomical points, and calculates various angles and distances between body segments, using these values to give the scores of the risk indicators for LBP, NP or scoliosis.

The objective of this dissertation is first to make some modifications and improvements in the software, so that the software could detect the markers and the anatomical parameters selected. Also, we need to make this detection possible in images obtained with different cameras and with different scales. Afterwards, we want to test its performance, to then validate the measurements made by the software with real measurements in a non-pathological population.

To achieve these objectives, we first determined the best setup to be used with the software to calculate the measurements needed for the postural analysis. After, we validated the results obtained by the software, using a non-pathological population of young-adults, and later we compared them with the results of the postural assessment to the same population made by a clinical software for posture and gait analysis. The last step of this study is to test the performance of classifiers using the data acquired from this study population.

The overall goal of this project is to develop an assessment kit, easily used by any professional present in schools, sport clubs or other institutions where postural assessment can be of importance. This can be understood as a population screening, which will compare the data of each subject with standard values of normality for each age. These standard values can be gathered from the multiple application of this method in different children, increasing the database of the application and allowing better risk indicators detection. The aim is to get a national profile of the children’s posture that justifies policy making actions for school health, to provide health promotion programs and disease prevention actions. A second aim is to provide posture health monitoring. Each institution (school, club) can contribute to the monitoring of the more critical periods of child development in terms of LBP and scoliosis, and to contribute actively to the early detection of risk indicators, promoting school-age child’s postural health.

1.3. Dissertation Structure

This dissertation is composed of six chapters. Chapter 1 explains the motivation and objectives of this thesis. In chapter 2, we address the background of musculoskeletal disorders, focusing on the Portuguese case. This chapter provides a basic description of musculoskeletal disorders, such as LBP, NP and scoliosis, the risk factors associated with their development and the anatomy of the spine and sternum. Chapter 3 presents the types of software for posture analysis currently on the market: software

for use by clinical professionals and for everyone usage. In chapter 4, we present the methodology of the different phases of this work. We explain the process of improving and testing the performance of our software and validating the data obtained from the study population. In chapter 5, we discussed the results obtained during the various phases of this dissertation. Finally, in chapter 6 the general conclusions of this study are drawn, and we present future perspectives for the development of the software.

2. Background

2.1. Impact of Posture Problems

Usually, the costs of a disease are not only the costs related to its treatment, such as medicines consumed, primary and specialized outpatient care, which are called direct costs, there are also indirect costs and intangible costs associated. These indirect costs correspond, for example, to the loss of people's production and reduction of productivity [7,21,22].

Regarding the musculoskeletal disorders, such as the LBP, NP, and scoliosis, in most cases the indirect costs tend to be higher than the direct costs. In these situations, indirect costs includes loss of ability to function in daily life, in adults also include sick leave, loss of ability to function in daily life and work productivity and early retirement disability pension, and in adolescents implies an increase of school absenteeism and physical inactivity, leading to greater use of electronic devices [3,7,9,22].

The loss of productivity usually begins with presenteeism, followed by absenteeism, and leads to disability pension. Absenteeism is the temporary absence from work due to short-term disabilities caused by health problems, and presenteeism is the reduction of productivity while present in the workplace despite being affected by the disabilities, which in the case of back pain is approximately three times higher than absenteeism [7,21,22].

In 2020, around 123 thousand people in Portugal had their work life affected due to bone, joint or muscle problem which mainly affects the back, with at least 42,8% had been absent from work for at least one day, at least 37,2% had been in presentism and around 10,8% did not expect to return to work. Also, around 94 thousand were bone, joint or muscle problems which mainly affects the neck, with at least 34,2% had been absent from work for at least one day, at least 44,9% had been in presentism and around 8,9% did not expect to return to work. This data was collected during the second term of 2020, which was a period of lock down [23].

M. Gouveia and M. Augusto [21] estimated that in 2010 the indirect costs of chronic pain in the lower back and joints in mainland Portugal were €739,85 million, of which €458,90 million corresponds to the costs of employment reduction due to long term disability and €280,95 million was due to short term disability absenteeism, however it was not possible for the study to make the cost estimate of presentism.

2.2. Anatomy

2.2.1. Vertebral Column

The spine has several functions, namely, the support of the head and trunk, allows their movement and absorbs the tensions produced during these movements, protects the spinal cord and serves as a place of attachment of the limbs, rib cage and postural muscles [24,25].

The spine (Figure 2.1) consists of 33 vertebrae divided into 5 regions: 7 cervical vertebrae (C1-C7), 12 thoracic vertebrae (T1-T12), 5 lumbar vertebrae (L1-L5), 5 sacral vertebrae and 4 coccygeal vertebrae, however, during growth to adulthood, the 5 sacral vertebrae and the 4 coccygeal vertebrae merge to form only one sacral bone, the sacrum and one coccygeal bone, the coccyx, respectively [8,24,25]. Between each vertebra there is an intervertebral disc, composed of a gelatinous pulposus nucleus surrounded by a fibrocartilage ring, which increases the flexibility of the spine, provides

additional support to the body, and absorbs suffered shocks. In addition, the spine contains 4 large curvatures, the thoracic and pelvic kyphosis and the cervical and lumbar lordosis, which allow the human being to stand upright [24,25]

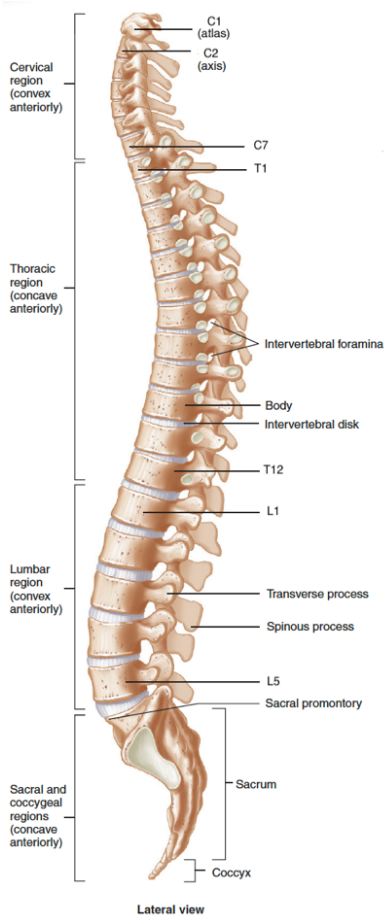


Figure 2.1: Lateral View of the complete vertebral column [25]

2.2.2. Vertebrae

Each vertebra (Figure 2.2) is composed of a main body, a vertebral arch, and several processes. The body has spongy bone and bone marrow inside, which are covered by a layer of compact bone. This has two surfaces, upper and lower, which is where the intervertebral discs are attached [8,24,25].

The vertebral arch protrudes from the back of the body, and it is possible to divide it into two parts, left and right. Each of these contains a pedicle and a lamina, the two blades on each side being joined, from where the spinous process is projected, which corresponds to the protuberances possible to be seen in the midline of the back. In addition, the transverse process is projected on each side of the vertebral arch, at the point of connection between the pedicle and the lamina. Both the transverse and the spinous processes allow the fixation of the spinal muscles and ligaments [24,25].

The vertebrae also contain two upper and two lower processes, which articulate with the lower and upper processes of the adjacent vertebrae, respectively. These are the processes that give the spine greater rigidity, providing greater protection for the spinal cord. The vertebral arch together with the posterior part of the body form the vertebral foramen, and in turn, the various vertebral foramina make up the vertebral canal, where the spinal cord is located and protected [8,24,25].

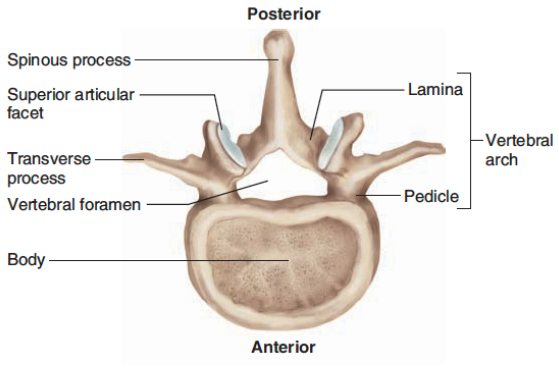


Figure 2.2: General Structure of a Vertebra [24]

2.2.3. Sternum

The sternum is part of the thoracic cage located on the front of the body. This is subdivided into three parts: the manubrium, the body and the xiphoid process, as seen in Figure 2.3 [24–26].

The superior margin of the manubrium has a jugular notch in the midline, which can be easily palpated between the clavicles, with which articulates with. The body is the longest part of the sternum and joins the manubrium at the sternal angle. The manubrium and body have the ribs attached. The xiphoid process provides attachment for some of the abdominal muscles [24,25].

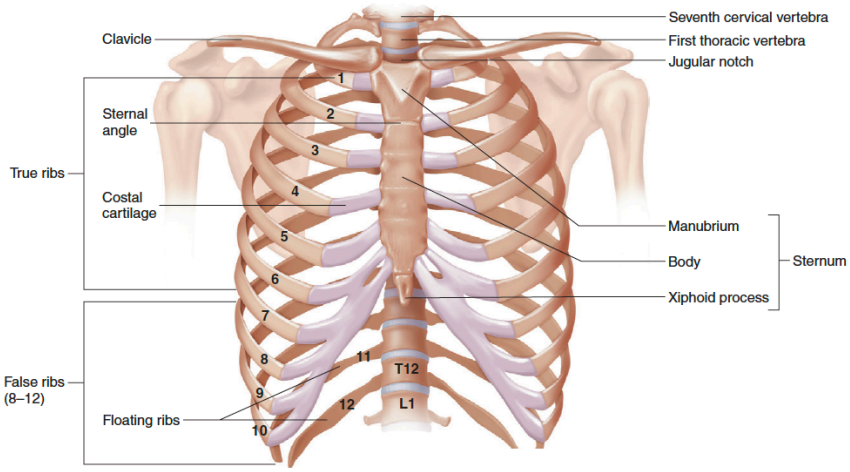


Figure 2.3: Anterior View of the Thoracic Cage, with the sternum in the middle [25]

2.3. Musculoskeletal Disorders

Postural defects are a growing problem that increasingly affects children and adolescents at school age. A person's posture changes throughout life, with the greatest changes occurring during the period of growth. It is in this period, that the person's bone structure develops, and sometimes the growth of the muscles and tendons does not keep up with the bone growth rate leading to a greater rigidity in the joints, which can lead to disorders and to postural defects and cause orthopedic and rheumatologic diseases. Additionally, in the course of time, postural defects can lead to pain complaints, which frequently limit daily physical activity [7,12,15,27].

In children and adolescents, the environment, particularly the environment in schools, can contribute to the development of musculoskeletal disorders. The cause of musculoskeletal disorders in adolescents is multidimensional [28], involving the participation in sports or exercise, long periods of inactivity, poor posture while sitting, and wearing backpacks that are overweight. In the classrooms, the students often adopt incorrect postures, especially during reading and writing [8,12,27,29].

2.3.1. Low back and Neck pain

Low back pain (LBP) and neck pain (NP) is defined as pain and discomfort localized in the lumbosacral region and cervical region of the back, respectively. They are most often of nonspecific origin, musculoskeletal, self-limiting and most causes are benign in their clinical course [1,6,30,31].

LBP and NP are symptoms frequently present in people of all ages, being increasingly common in school-aged children. Their prevalence starts low in children, but strongly increases during adolescence [1,6,9,32,33]. In most cases these symptoms are due to non-specific causes, but there may be some risk factors that influence their onset, however this issue is still controversial [33].

In a European survey [34], it was assessed the prevalence of chronic pain in different body sites, and the results showed that the back in general had the highest prevalence, having 24 % respond the back without specifying, 18 % the lower back, 8 % the neck, and 5 % the upper back. Additionally, in the I. Dianat *et al.* [6], 59.6% of the population in the study (schoolchildren aged 12-14 years) reported neck, shoulder or low back pains during the preceding month of the study. Of this 59.6%, were reported low back, neck, shoulder pain by 33%, 35.3% and 26.1% of the students, respectively.

Two studies of B. Minghelli *et al.* [9,33], with data collected with a time difference of 7 years, performed with a population of schoolchildren living in the south of Portugal, demonstrated a high prevalence of LBP. Both studies showed about 47% of the students reported LBP during the 12-month period before the study. These studies also show there was a slight increase in the values regarding a lifetime prevalence of LBP from 62.1% in the older study to 65.1% in the more recent one [9,27].

Furthermore, studies show that musculoskeletal pain, specifically NP and LBP, in childhood and adolescence is a significant risk factor for experiencing such symptoms in adulthood, possibly becoming chronic in adulthood, with odds ratios up to four. Also, this risk is higher the more days the child/adolescent suffers from LBP [1,9,29,30,32,35,36].

Pain is not the only consequence of LBP and NP, they also have associated: biological factors, which include morphological abnormalities and functional disorders of the spine; psychological factors, since common mental disorders such as anxiety and depression are common in LBP and NP patients; and social factors, which involve sociodemographic (high age and female gender) and socioeconomic variables (less education, lower income and less qualified profession) [3,4,7,28,30,31]. Several studies show that these disorders have an impact on people's quality of life, for example, they are strongly associated with pressure and social suffering in the workplace [1,3,7,12].

2.3.2. Scoliosis

Scoliosis is the one of the most common spinal disorder in children and adolescents, and refers to a structural three-dimensional deformation of the spine that involves a curvature in which a more than 10° lateral flexion (known as Cobbs angle, figure 2.4) is formed in the spinal rotation in the frontal plane [10,11,13,26,37–40].

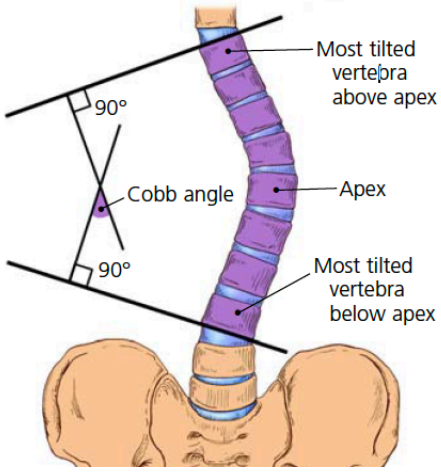


Figure 2.4: Cobb's Angle [39]

Most often, this condition has no known cause and is referred to as "idiopathic", which means that the causes have not been identified, and can be classified in 4 subgroups: infantile idiopathic scoliosis, juvenile idiopathic scoliosis, adolescent idiopathic scoliosis and adult idiopathic scoliosis, which are defined by the age of onset [37,38,40]. The most common types of idiopathic scoliosis are revealed during childhood and adolescence, being the most common the adolescent idiopathic scoliosis (AIS) [10,11,13,26,37]. According to S. Bozkurt *et al.* [11], the prevalence of adolescent idiopathic scoliosis ranges from 1 to 3% among children and adolescents aged 10 to 16 years. Despite being idiopathic, there is evidence of a possible contribution of a wide variety of environmental factors, such as poor posture and asymmetrical distribution of loads, in the development of scoliosis or its association with asymmetries in the length of the skeleton [27].

The scoliotic deformations can be related to generalized joint hypermobility or due to postural behavior, among others. Generalized joint hypermobility (GJH) describes an increased joint range of motion with simultaneous absence of any other systemic disease, involves proprioception impairment and might cause, in the long-term, postural alterations, pain and joint instability [10–13]. Joint hypermobility has been commonly observed in normal school children. S. Bozkurt *et al.* [11] demonstrated a prevalence of 18.4% in a population aged between 10 to 15 years. The prevalence of GJH varies according to age, race and sex, with females having in general a higher prevalence as well as a higher degree of general joint laxity [11].

This disease is accompanied by several symptoms, such as back pain, deformation of the thoracic cage, weakening of the respiratory muscles, and limited range of motion, and can also cause reduced control of balance and gait asymmetry, which therefore leads to subsequently results in a reduced quality of life [10,38,41]. P. Mahaudens *et al.* [42] revealed that AIS patients need to do 30 % more physical effort than normal to ensure habitual locomotion, which requires an increase in oxygen consumption.

The most common diagnosis test performed during physical examination for scoliosis is the Adam's test. This consists in the patient bending forward and the examiner evaluating the symmetry of the back of the patient. A scoliometer is usually used to quantify the curve observed, from which the angle of trunk rotation can be determined [38–40,43–45]. After this examination, a radiography is required to have a more accurate measurement of the Cobb angle, since the initial value of this parameter is one of the most important to predict the progression of the disease [38].

The trunk rotation is one of the main aspects associated with scoliosis and can start at the age of one year, as a possible consequence of the maintenance of primitive reflexes, it is also related with the critical periods of rapid growth [15,16]. According to E. Z. Gieysztor *et al.* [14] most of the trunk rotation

is toward the right, there is a higher frequency of asymmetry among the girls than boys and an early detection of trunk rotation prevents the progression of scoliosis. Besides, the most common curve types of scoliosis are the thoracic curve, the thoracolumbar curve and the lumbar curve [11,37].

Furthermore, E. Kenanidis *et al.* [26] reported in their study, made with a population of scoliotic patients between 14 and 18 years old, that the manubrium of the sternum was more centrally positioned in the axial plane compared to the body of the sternum in both groups. However, the scoliotic group in the study had a more inclined position of the body and xiphoid process than the control group. Also, according to F. Zhu *et al.* [46], the rib lengths on the concave side were significantly longer than those on the convex side in the apical region in AIS patients, contrary to non-scoliotic patients which usually have rib length symmetry.

2.4. Risk Factors of Musculoskeletal Disorders

Considering the risk factors pointed out for the development of these pathologies, some are related with the postural behavior [14,26], but these do not seem to be the only indicators, also the experience of pain, being female, increased age, involvement in sports and stress response strategies seem to be good predictors, since the musculoskeletal disorders previously referred in early ages are multifactorial [4,8,12,27,29]. The identification of these risk factors can facilitate the application of proper prophylactic actions to provide an adequate quality of life without pain, and limitation of physical and motor capabilities [12,35].

2.4.1. Age and Gender

Postural defects are rarely seen before school age, and usually have a higher prevalence in adolescence.

The most common type of scoliosis is the AIS, that corresponds approximately to 90% of the cases of idiopathic scoliosis in minors, and frequently after the start of puberty [13,26,37]. According to E. Kenanidis *et al.* [26], one possible explanation to this higher prevalence, are the variations that occur during the development of the body of the sternum, which during childhood is divided into four segments, but around the age of 15, it merges completely to form one piece, since has connected most of the ribs and vertebrae involved in AIS.

LBP, as well, is a common problem in school age children, with its rates rise until the age of 20-22 years, at which the prevalence of LBP is similar to the one in adults, with an estimated yearly prevalence of 20% and a lifetime prevalence of 70-80% [5,8,30,31,35,36]. According to B. Minghelli *et al.* [33] reveals that older students (between 13-16 years old) have 2.49 times greater probability to have LBP than the students with 10-12 years old, which the study justifies as being possibly caused by pubertal growth and an increased stress on the spine due to longer duration of sitting class.

Additionally, postural defects are more frequent among girls than boys, which could be related to gender differences in development [26,30,31,35,47]. Although male and female infants have an equal risk of scoliosis, with increasing age girls become more at risk compared with their male counterparts [12]. The female to male ratio ranges from 1.5:1 to 3:1 and increases with increasing age. Also, the severity of the curves (higher Cobb's angle) is higher in girls than in boys [37,40].

Furthermore, P. Galozzi *et al.* [1] and I. Dianat *et al.* [6], show an association between gender and from NP, LBP and shoulder pain, with girls more likely to suffer than boys, and the B. Minghelli *et al.* [33] report that girls have 2.5 times more probability to develop LBP. This may be possible due to the differences in growth, due to the greater flexibility of girls' spine compared to boys and to the differences in the perception of pain by the genders [33,48]. Although, it is still unclear if the differences in prevalence between genders are only related to genetic predisposition or whether it is also related to the different behavioral habits shaped by society at school age [12].

2.4.2. Schoolbag weight

Another concern frequently associated with these conditions is the use of the backpack, considering the weight carried, the time carrying it or the bilateral versus asymmetrical use of the backpack itself [6,31,49,50], with some controversial results [27].

There is evidence that overweight schoolbags may be a factor contributing to musculoskeletal complaints in schoolchildren [6,27]. Some studies indicate that carrying overweight schoolbags can be a contributing factor to the development and aggravation of musculoskeletal pain in schoolchildren [13,27,51]. The increased weight on the back results in excessive forward trunk lean, which can lead to increased lordosis and an elevation and retro positioning of the shoulders, causing the spine to be altered from its neutral position [41].

B. Minghelli, *et al.* [33] in a study about a population involving students from Portugal, found that a high percentage of these students (41.1%) carried an excessively heavy backpack. Also, that 49% of students with LBP carried their backpack incorrectly and that these students had 1.83 times greater probability to have LBP. However, according to I. Dianat *et al.* [6], the type of schoolbag was the only significant variable they found associated with LBP, and the schoolbag weight variable was only significantly associated with the occurrence of NP and shoulder complaints. Also, there was significant association found between the time spent carrying a schoolbag and NP, in which longer periods of time contributed to a greater number of reports of NP. Nevertheless, S. Negrini and R. Carabalona [52] reported an association between the time spent carrying a schoolbag and LBP, but did not find a direct association between the load of the schoolbag and LBP.

In respect to scoliosis, K. Watanabe *et al.* [53] and B. Minghelli *et al.* [27] studies found no correlation between the presence of scoliosis and the type of schoolbag and the schoolbag weight.

Regarding the transport method of the school backpack, the use of only one shoulder can lead to postural changes due to the asymmetric load, which leads to the inclination of the spine, elevation and retropositioning of the shoulder on the side of the load and accentuated cervical lordosis, probably contributing to the increase in pressure on the loading side of the intervertebral discs, which leads to increased back pain, so the recommended method is to use both shoulders, although it can cause forward leaning, which leads to an anterior flexion of the trunk [6,13,54,55]. P. Korovessis *et al.* [54] conclude that the students who carried backpacks asymmetrically are more than 4 times likely to suffer from high intensity pain compared with students who carry backpacks symmetrically, specifically they are 2.9 times more likely to suffer from dorsal pain and 5 times more likely to suffer from LBP. According to B. Minghelli *et al.* [27] and J. T. F. Lopes [56], both studies performed in Portugal, showed that most of the students use the backpack with the two straps on the shoulders.

However, while there is still controversy as to whether a heavy backpack has a direct influence on LBP or scoliosis development, the effect of backpack weight on posture in children and adolescents should be still carefully evaluated and the policies of reducing the maximum load recommended for a

weight limit of 10% to 15% of the body weight that children need to carry to school seems a good practice [13,30,41]. One example is the Portuguese National Educational Health Program [57], that develop strategies like “Mochila Leve” with some indication to the parents and educators regarding the control of the load carried out by the children [58], and that puts into action the guidelines proposed by the National Health Plan for School Health.

2.4.3. Exercise and Sedentary Activities

There is a relationship between the physical activity of a child and occurrence of postural defect [12]. The participation in sports and the intensity of the training is shown to be associated with the probability of developing LBP. The intensity of physical activity is associated with the incidence of LBP in school-age children as a U-shaped function, in which inactivity, low and high levels of physical activity are associated with a higher risk of reporting LBP [4,7,8,30,36,51]. Currently, there has been an increase in sedentary activities due to the use of electronic devices. This lack of physical activity implies less flexibility and less muscle strength, as long periods of static posture cause higher pressure on the intervertebral discs and ligaments and do not allow sufficient recovery from local muscle fatigue, which leads to the development of LBP and NP [9,27,31]. On the other hand, high intensity exercise with repeated lifting of heavy loads, insufficient recovery between activities, as well as inadequate positions that cause the rotation and flexion of the spine are risk factors for developing back pain [1,7,15,31].

However, regular physical activity of moderate intensity seems to be associated with a lower risk for developing LBP and have a protective function against LBP [1,2,31], since having a good level of strength in the abdominal musculature and spine extensors and having flexibility in the posterior thigh muscles helps to decrease the risk of developing LBP [27]. The European and the OMS guidelines for prevention of musculoskeletal disorders considers that moderate physical activity with adequate rest periods, is one of the most effective tools for prevention [7,15]. Also, regarding scoliosis, exercise can help to improve neuromotor control, the stability of physical balance, and respiratory functions [10].

2.4.4. Posture

Incorrect postures are other possible factors that contribute to the development of musculoskeletal disorders, but each student adopts incorrect postures at different times of the day, for example, when standing, carrying the backpack, sitting, the latter being one of the situations in which more students complain of pain [9,27]. In addition, K. Grimmer and M. Williams [48] reported that children who sit for long periods of time have a higher risk of low back pain.

B. Minghelli *et al.* [33] revealed that students who sit with the spine position incorrectly have 2.49 times greater probability of having LBP, as well as students who adopt wrongly positions for watching TV or playing video games have 2.01 times greater chances, and students who have an incorrect posture while standing have a 3.39 times higher probability. Additionally, another study from B. Minghelli *et al.* [9], showed a significant association with the time of use of the mobile phone with a wrong posture of cervical flexion and NP, revealing also that adolescents who use the mobile phone, watch television and played video games more than 10 h per week have an increased risk of developing both NP and LBP. Also, A. G. Silva *et al.* [59] showed that the use of the mobile phone for more than 5h per day was significantly associated with increased odds of pain in the last 7-days for the back, the wrists, the hips and the knees. R. Gunzburg *et al.* [44] also reported that children who played videogames for more than 2 hours per day have significantly more risk to develop LBP.

2.4.5. Prevention and Medical Care

Despite the incidence of this pathology, consultations for LBP are not very common, mainly concerning patients with more intense or longer lasting pain [1,31], according to R. Gunzburg *et al.* [44] and K. D. Watson *et al.* [35], approximately only one quarter of the school age children who reported pain sought medical help. Also another study, B. Minghelli [33], made with school age children in the south of Portugal, reported that during the previous year of the study 47.2% of the population sample had experienced LBP, but only 18.2% of them seek medical help.

However, it is important to prevent recurrent episodes of pain, chronicity and functional decline in individuals who suffer from musculoskeletal disorders, both for improving their quality of life, and for society, as it prevents loss of productivity, absences due to sickness and disability pensions and the excessive use of healthcare [7]. This prevention should start in the early stages of a person's life during the school life phase.

Adolescence is the time of shaping health behaviors, and many of the risk factors previously referred to can be modifiable through an intervention, such as implementing educational sessions in schools for parents and children with information about correct health behaviors and the consequences of not having them [9,12]. The content of those sessions should include information, advice from experts, incentives to do exercise, not to be inactive for long periods and to reduce the use of electronic devices [7]. Additionally, children should receive practical sessions to learn the best posture to be adopted while sitting, sleeping, watching television, and playing video games, as well as physical training and stretching exercises to be performed after long periods in a static position. Teach the best way to transport the backpack and what is the acceptable weight, as prevention for musculoskeletal disorders, should also be considered [2,9,27].

These campaigns have the aim to promote health behaviors and increase health literacy of the population towards musculoskeletal pain, especially back pain and scoliosis, in order to decrease the intensity of pain and impairment of workability due to pain [7].

In addition, it is also important to conduct a routine medical evaluation of children at school to get an early detection of musculoskeletal disorders, because the earlier it is treated, the better the outcome [44].

3. Literature Review

3.1. Postural Assessment Methods

The diagnosis of musculoskeletal disorders is based on the use of medical imaging methods, such as X-Ray and MRI, or postural assessment methods [40,45,47]. However, given that most cases don't have an identified organic cause, the use of postural assessment methods has been increasingly recurrent, which can be used to quantify spinal curvature deformities, to monitor changes over time and to evaluate the effectiveness of interventions performed on the patient [8,30,31,60]. Still, these are used mainly in the context of physiotherapy research, being usually only accessible to health professionals. Additionally, multidisciplinary pain treatment programs that combine physical, psychological and educational exercises have been proven to be effective, but these are expensive and usually limited to specialized medical centers [28,61]. Consequently, mobile health apps (mHealth) are starting to appear for anyone to analyze their posture, normally for just a small fee.

3.1.1. Methods for Health Professional

These body assessment methods are usually used by physiotherapy professionals, being used for diagnosis and for research purposes. Due to the variety that exists, there is still no consensus on which is the most suitable, being some more used than others, such as the Postural Assessment Software (PAS), PostureScreen Mobile, Photography and Kendall *et al.* method, Method of Kendall *et al.*, Photogrammetry and Photography [62]. Furthermore, there is no consensus also in the way to present the assessment results, some of the studies using these software present quantitative results (which the results are values of angles relative to posture alignment) and others present qualitative results (which describes the changes observed), however the quantitative analysis is the one most used [62].

3.1.1.1. Postural Assessment Software (PAS)

The PAS is the most used software of body postural analysis and produces quantitative results. This is a computer program that analyzes digitalized pictures acquired from the subjects, in which they have to be close to a plumb line for vertical calibration, wear appropriate clothing and have marked the intended anatomical points [62–64]. The anatomical points used can vary from study to study, for example, two different studies [63,64] used this software, but one of them marked 50 anatomical landmarks and the other marked 32, having some in common such as acromion, anterior superior iliac spine (ASIS), greater trochanter of femur, knee, articular line, patella, tibial tuberosity, lateral and medial malleoli, medium point between second and third metatarsal, the spinal process of C7 and T3, scapula, inferior angle, posterior superior iliac spine (PSIS), medial line between the calcaneus and the popliteal fossa, midpoint between the two malleoli and calcaneus. Both studies used polystyrene balls as markers.

The software measures distances and angles of the body segments in the frontal, posterior and lateral view, and allows archiving results and comparing them to observe the evolution of the patient. Positive aspects of the PAS software protocol have been its practicality and the ability to standardize measures, however it is believed that this method should not replace clinical judgment but complement it [62–64].

The PAS are usually computer software, which make them not movable. They are also normally for the use of health professionals, having high costs.

Contemplas

The *Contemplas*, an example of a PAS, is a computer software with the purpose to be used by professionals in a clinical environment. This software analyzes posture, movement or gait through images or video. Regarding the postural analysis, the software allows a 2D or 3D assessment of the static posture, in which it measures the body axes, angles and measurements. In 2D analysis, the software analyzes images from the posterior and side views. In the 3D analysis, the evaluation of an image of the anterior view is added and requires a minimum of 17 markers. Though, it is possible to use more markers to obtain additional metrics [65].

3.1.1.2. PostureScreen Mobile

A software app (tablet or phone app) for evaluation of posture and movement specifically and only designed for health professionals. This software calculates posture variables, such as angulation and displacement, using the anatomical landmarks selected from 2 to 4 photographs (from different views) of the subject taken with the camera on the mobile or tablet. The output file of the analyses are the values of the posture variables and the photographs with the locations of the anatomical landmarks relative to a neutral position. Besides being only for the use of health professionals, it also has high costs around 50 US dollars [66,67].

3.1.1.3. Method of Kendall *et al.*

In this method the posture of the patient is observed naturally standing in 4 different views, the anterior, the posterior and both profile views. In the first one, it is observed the position of the head, the differences between the height of the two shoulders and between the iliac crests and if the knees present valgum (causes inward curve of the legs and the knee touch each other) or varum (causes outward curve of the legs and the knee are furthest away than normal) deformities. In the posterior view assess if the calcaneus is vertical align with the Achilles tendon, if the column is vertical align without presenting any deviation, and if both medial malleoli are with the same height, as well as, both inferior angle of the scapula, the anterior posterior iliac spines and both popliteal fossae. In the posterior views it is analyzed the convexity of the dorsal region and the concavity of the lumbar region, if the shoulders present protrusion and if the knees are too bent back [62,68].

According to Kendal *et al.* [68], there are 4 possible classifications of the posture alignment: ideal alignment, flat back (corresponds to a posture in which the lumbar spine and lower part of the thoracic spine are straight and the upper part of the thoracic spine has an increased flexion), kyphosis/lordosis (which means that the dorsal kyphosis or the lumbar lordosis is accentuated, respectively) and sway back (occurs when there is a flattening of the lumbar spine and an accentuation of the thoracic kyphosis).

One evolution of this method is the Photography and Kendall *et al.* method [68] which innovates to the extent that the patient's posture is analyzed through photographs of the patients in the 4 views, being subsequently compared with the images of the control postures presented in the figure 3.1, defined in the postural classification by Kendall *et al.* [62,68].

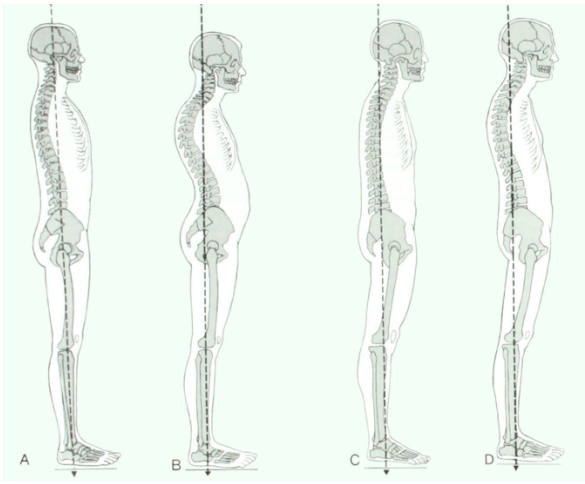


Figure 3.1: Kendall's Posture Classifications: (A) Ideal alignment; (B) Kyphotic-lordosis posture; (C) Flat-back posture; (D) Sway-back posture [68]

3.1.1.4. Photogrammetry

The photogrammetry method also consists in taking pictures of the posture of the subjects, that afterwards are analyzed with a graphical software, for example the Posture Pro software, that scans the image to identify anatomical landmarks and quantify linear and angular measures generated by the intersection of the anatomical landmarks [60,62]. This method compares in the frontal and posterior plane the symmetry between the left and right side and the straightness of the spine is analyzed, and in the sagittal plane assess the magnitude of the physiological curvature characteristics of the spine [69].

3.1.1.5. Photography Evolution

In this method photographs of patients are taken, also each in the 4 views, and later shown to qualified physiotherapists who perform a descriptive analysis of the patient's posture in the images, namely assessing changes in posture of the shoulders, spine in the thoracic and lumbar area, the knees and the scapular and pelvic girdles [62,70].

3.1.2. Online Apps

In recent years, the use of electronic health (eHealth) and mobile health (mHealth) apps for everybody's usage, has become increasingly common. The appearance of these allowed patients to be active participants in maintaining or improving their health and becoming more aware of their behavioral patterns, as they can normally be used by anyone. The eHealth and mHealth can be diagnostic or treatment or both [71].

3.1.2.1. Kaia App

The Kaia app (for iOS, Android, and native Web solutions) is a mHealth app with an evidence-based multidisciplinary pain treatment, classified as a medical product of class I, for people with non-specific LBP, which includes patient education, video-guided physical therapy exercises and mindfulness guiding, all authored by board-certified by physicians and psychologist (Figure 3.2). This app also uses biomarkers during the exercises to analyze and quantify the state of health of the user [28,61,72].

This app is free to download and allows a free trial of 7 days, however, to continue to use after this period and to unlock all the functionalities it is needed to purchase the pro-version with a cost of 9.99€/month [61]. Until now, the app is only available in English or in German [72].

To start use the app, it is necessary to register, confirm that you have already visited a physician who did not give a contraindication for physiotherapy and perform a mandatory self-test on the health situation. The self-test evaluates the distribution, duration, and intensity of pain, as well as general fitness, and depending on the results, the algorithm customizes the regime and the content of the exercises for the user from a set of 120 exercises. Daily pain and sleep levels should be recorded in the diary, using numerical rating scales from 0 (worst) to 10 (best), in order to update progress and content day by day. The data of the user is stored in the company servers. Additionally, there is a chat function that connects users to a physical therapist for motivational and exercise-related issues [61].

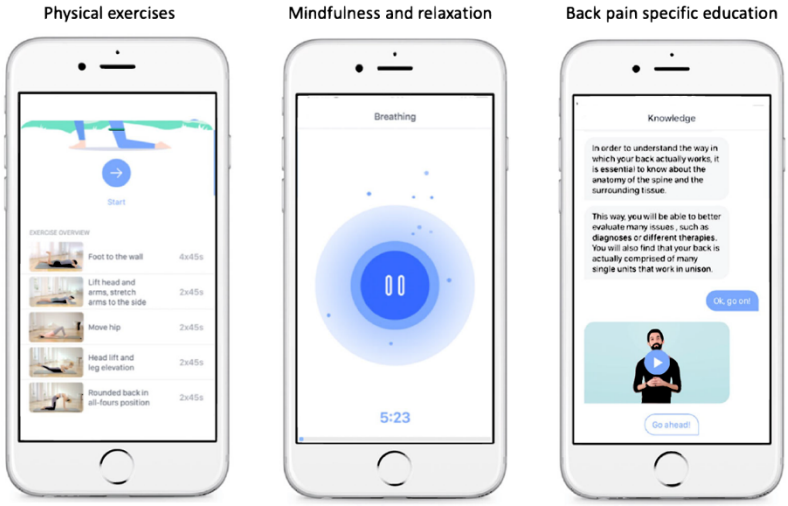


Figure 3.2: Examples of the main modules of the Kaia app, the physical exercises page on the left, the mindfulness page in the center and the educational page on the right [28]

3.1.2.2. Kiio

The Kiio is a digital MSK solution with an evidence-based therapy app for people with pain in the low back, knee, hip, and neck. The app allows the users to create a profile to give personalized therapy exercises based on the user symptoms, educational articles, behavioral modification tools to ease pain and strengthen muscles, track their progress and access to the Kiio Care Team for questions and support [73]. Some of these functionalities are present in the figure below, figure 3.3.

Figure 3.3: Examples of different contents of the Kiio app. **On the left:** Example of the therapy exercises page. In the center: Example of the educational articles page. **On the right:** Example of a user progress page and chat page with Kiio Care [73].

Nowadays there are several methods of postural analysis, however most are expensive and usually used in a clinical environment by professionals. Currently, there are also some mobile applications for the patient use. Furthermore, postural assessment performed by these various methods is usually performed after the onset of musculoskeletal pain in patients, as a form of diagnosis or treatment.

However, early detection of musculoskeletal disorders and reduction of associated risk factors during childhood and adolescence decreases the likelihood of these disorders in adulthood. The reduction of these disturbances makes it possible to reduce the impacts on people's quality of life and reduce the existing costs associated with their incidence. As well, implementing prevention screenings from an early age makes it possible to identify these problems and obtain medical help for children and adolescents at an early stage. As well as increasing the population's attention to musculoskeletal problems through awareness campaigns.

4. Methodology

The methodology used in the project is divided into 4 major phases: improvement of the software already created; assessing the software's performance; acquisition of data from an experimental group; and implementation of classifiers. Each of these phases were also divided into several progressive steps.

During the development of this project, we submitted an application to the Born from Knowledge program of the National Innovation Agency, in which we participated as one of the teams in the first phase. As well, this project was approved by the ethics committee of ESSCVP – Lisboa.

4.1. Software Improvement

The software created uses specific anatomical sites that need to be identified by markers, to measure various anatomical parameters, with the purpose to evaluate the subject's posture.

In a first phase, the anatomical sites where the markers are placed were selected for all the different views (anterior, posterior, left and right lateral view), namely, the jugular notch, xiphoid appendix, spinous processes of C7, most prominent point of thoracic kyphosis, deepest point of lumbar lordosis, acromion, ASIS, PSIS, greater trochanter of femur, lateral condyle of the femur, popliteal fossa, patella, tibial tuberosity, lateral malleolus, posterior midpoint between the lateral and medial malleolus, calcaneus and the fifth metatarsal. The last 12 anatomical sites mentioned are bilateral. These anatomical sites are presented in figure 4.1 and were chosen in conjunction with physiotherapy specialists and also based on scientific articles about other postural analysis software [60,63,64].

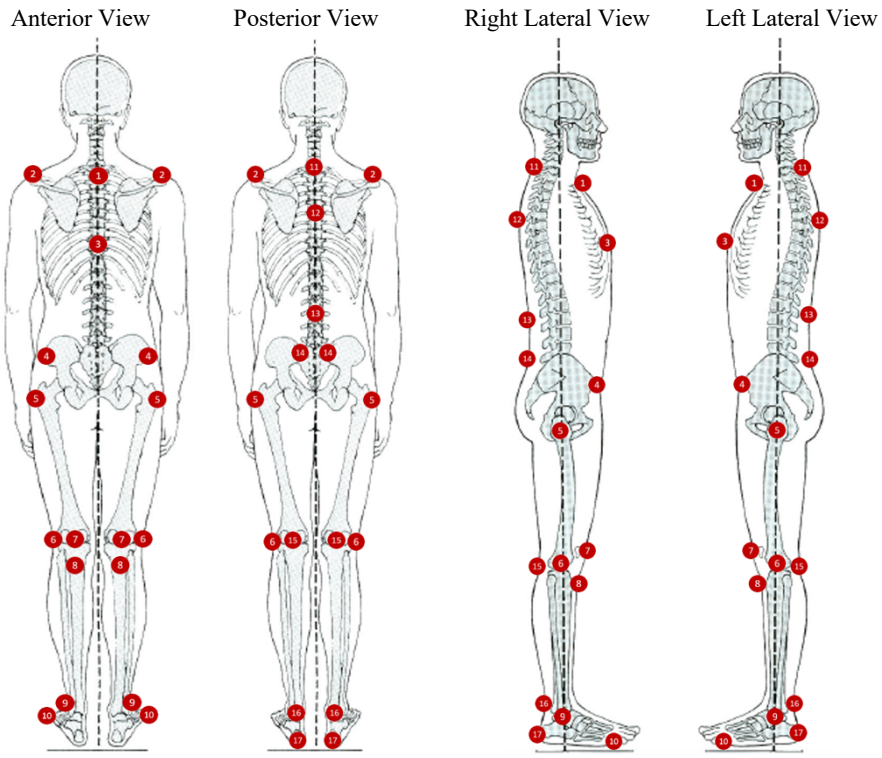


Figure 4.1: Anatomical sites selected to place the markers: (1) jugular notch, (2) acromion, (3) xiphoid appendix, (4) ASIS, (5) greater trochanter of femur, (6) lateral condyle of the femur, (7) patella, (8) tibial tuberosity, (9) lateral malleolus, (10) fifth metatarsal, (11) spinous processes of C7, (12) most prominent point of thoracic kyphosis, (13) deepest point of lumbar lordosis, (14) PSIS, (15) popliteal fossa, (16) posterior midpoint between the lateral and medial malleolus and (17) calcaneus.

In a second phase, we needed to improve the code of the program previously started by some colleagues in a course of the Integrated Master in Biomedical Engineering and Biophysics at FCUL. We included new marker positions selected and the anatomical parameters we wanted the software to measure, such as the angles and the length of body segments. Based on those values, the software gives the scores of the risk indicators as a result. The software determines a total of 38 features, of which 15 are bilateral. These parameters are described in table 4.1, table 4.2 and table 4.3 for the posterior view, anterior view, and the lateral view, respectively.

Anterior View

Table 4.1: Anterior View – Anatomical parameters calculated by the software and descriptions of the anatomical points used to determine each parameter

Parameter	Description of Anatomical Landmarks Used
1. Acromions Horizontal Alignment (AHA_A)	Angle between the line that passes through the 2 acromions and the horizontal line
2. Acromions - Sternum Angle (ASA)	Angle between the line of the 2 acromions and the line that passes jugular notch and xiphoid appendix
3. ASIS's Horizontal Alignment (ASISHA)	Angle between the line that passes through the 2 ASIS's and the horizontal line
4. ASIS's - Leg Angle (Left and Right Leg – $ASISLA_L$ and $ASISLA_R$)	Angle between the line of the 2 ASIS's and the line that passes through trochanter and patella
5. Knee Lateral Angle (Left and Right Knee – $AKLA_L$ and $AKLA_R$)	Angle between the line that passes the trochanter and patella and the line that passes tibial tuberosity and lateral malleolus
6. Acromion - Jugular Notch Distance (Left and Right Acromion – $AJND_L$ and $AJND_R$)	Distance between jugular notch and each acromion
7. Acromion - ASIS Distance (Left and Right Side – AAD_L and AAD_R)	Distance between acromion and ASIS of the same side
8. Lower Limb Length (Left and Right Side – $ALLL_L$ and $ALLL_R$)	Distance between trochanter and lateral malleolus of the same side

Posterior View

Table 4.2: Posterior View – Anatomical parameters calculated by the software and descriptions of the anatomical points used to determine each parameter

Parameter	Description of Anatomical Landmarks Used
9. Acromions Horizontal Alignment (AHA_P)	Angle between the line that passes through the 2 acromions and the horizontal line
10. Acromions - Vertebral Column Angle (AVA)	Angle between the line of the 2 acromions and the line that passes C7 and most prominent point of thoracic kyphosis
11. Thoracic Kyphosis Lateral Angle (TKA)	Angle formed by C7, most prominent point of thoracic kyphosis and deepest point of lumbar lordosis
12. Lumbar Lordosis Lateral Angle (LLA)	Angle formed by most prominent point of thoracic kyphosis, deepest point of lumbar lordosis and the midpoint between the PSIS's
13. PSIS's Horizontal Alignment (PSISHA)	Angle between the line that passes through the 2 PSIS's and the horizontal line
14. PSIS's - Leg Angle (Left and Right Leg – $PSISLA_L$ and $PSISLA_R$)	Angle between the line of the 2 PSIS's and the line that passes through trochanter and popliteal fossa
15. Knee Lateral Angle (Left and Right Knee – $PKLA_L$ and $PKLA_R$)	Angle formed by trochanter, popliteal fossa, and posterior midpoint between the lateral and medial malleolus
16. Ankle Lateral Angle (Left and Right Ankle – ALA_L and ALA_R)	Angle formed by popliteal fossa, posterior midpoint between the lateral and medial malleolus, and calcaneus
17. Acromion - PSIS Distance (Left and Right – APD_L and APD_R)	Distance between the acromion and the PSIS of the same side
18. Lower Limb Length (Left and Right – $PLLL_L$ and $PLLL_R$)	Distance between trochanter and calcaneus of the same side

Lateral View

Table 4.3: Lateral View – Anatomical parameters calculated by the software and descriptions of the anatomical points used to determine each parameter. These parameters are measured in both left and right lateral view

Parameter	Description of Anatomical Landmarks Used
19. Thoracic Kyphosis Curvature (TKC)	Angle formed by C7, most prominent point of thoracic kyphosis and deepest point of lumbar lordosis
20. Lumbar Lordosis Curvature (LLC)	Angle formed by most prominent point of thoracic kyphosis, deepest point of lumbar lordosis and the PSIS on the side of the respective lateral view.
21. Pelvis – Leg Angle (PLA)	Angle between the line that passes through the ASIS and PSIS and line that passes trochanter and lateral condyle of the femur
22. Knee Angle (KA)	Angle formed by trochanter, lateral condyle of the femur, and lateral malleolus
23. Leg – Foot Angle (LFA)	Angle between the line that passes trochanter and lateral condyle of the femur and the line that passes calcaneus and fifth metatarsal

The angles are calculated using the dot product between imaginary lines that connect the markers. The lengths are calculated using an object with known dimensions, to obtain the scale of the image, the number of pixels in the image equivalent to 1 cm in real life.

Some of the angles and distances that will be calculated are represented in purple and green lines in figure 4.2, respectively.

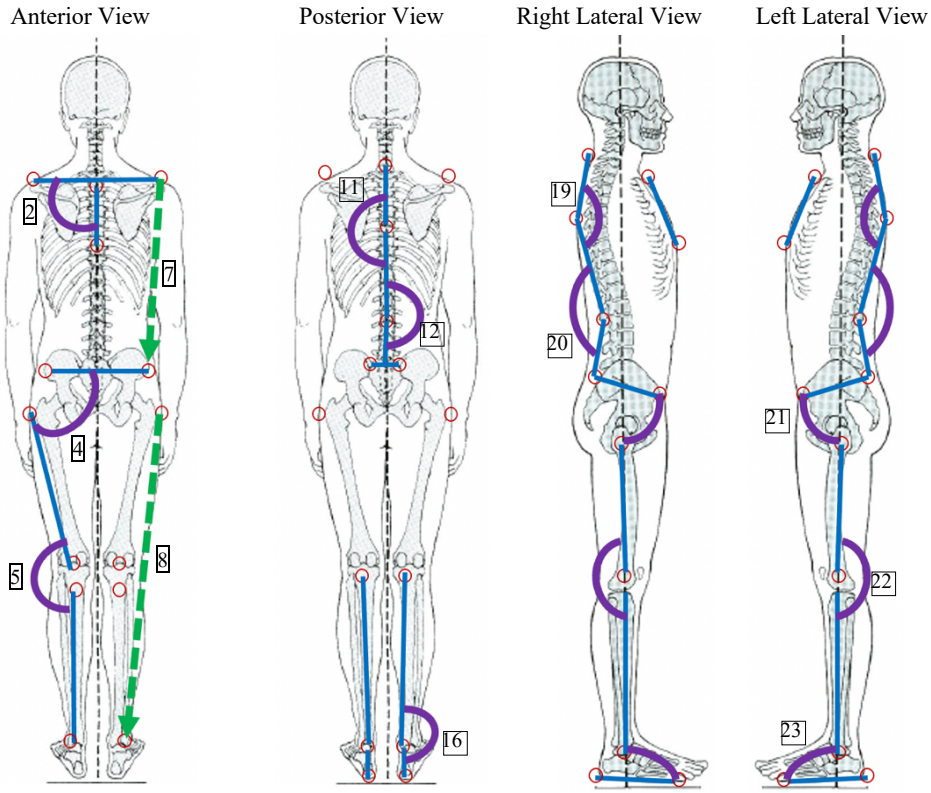


Figure 4.2: Examples of the angles (in purple) and distances (in green) between body segments that the software will calculate, in the posterior, anterior and the right and left lateral view.

Moreover, it was necessary to modify the code with the purpose to allow the software to detect the markers in images taken from different types of cameras, using the blob detection method. One type was a camera of a mobile phone, an *iPhone 8*, and the second was a photographic camera, a *Canon EOS 40D*. These camera models were the same throughout all phases of this project.

4.1.1. Image Processing and Metrics Determination

The software processes the images of the patients, first detecting the markers in the anatomical landmarks previously defined and the respective locations in the image. Afterwards it uses these locations to calculate the anatomical parameters, namely the angles and distances between body segments, giving these values and the scores of the risk indicators as a result.

In the image processing, the software initially uses a thresholding to isolate only the color regions defined for the markers, as exemplified in figure 4.3. Later it applies the blob detection method to detect and identify only the regions corresponding to the markers, drawing a circle on the approximate contours of the marker in the original image (Figure 4.3). After detecting the contours, the code determines the center coordinates of each marker. As well, the code will also have as output the number of markers detected.

The blob detection method consists of detecting regions in digital images that contain connected pixels that share the same light properties and that are brighter or darker than the surrounding region; these regions are called blobs [74,75]. In this project, the color property of the pixels was also considered, and before applying this method, a threshold was applied to isolate in the image only the regions with the colors of interest. This method, in addition to using the properties of light for the detection of blobs, can also filter the identification of these regions through additional parameters, such as area, circularity, convexity and the ratio of inertia. For each one of these is possible to define a maximum and minimum value different from those defined by default in order to reduce the possible regions to be identified by the method [76].

Figure 4.3: **On the left:** Example of image taken to the patient with the markers previously placed. In the center: Image obtained after applying the threshold to the original image to isolate the color regions of interest. **On the right:** Output image with the anatomical markers identified with green circles after using the blob detection method.

The area parameter allows to select the minimum and maximum area value that the detected region may have, being possible to ignore regions that, despite having the desired color property, do not have

an area within the desired range. Regarding the circularity parameter, it varies between 0 and 1 by default. The more this parameter has a value close to 1, the closer is the shape of the region of a circle. Convexity is defined by the ratio between the area and the area of the convex hull of the blob, which by default also varies between 0 and 1. The convex hull of a shape is the minimum convex set that contains all the pointes of that shape, and in turn, the convex set is a region where all points on a line segment lie entirely within that region [77,78]. The more the convexity parameter has a value close to 1, the more the area of the convex hull is closer to the total area of the blob. Finally, the ratio of inertia, which also varies between 0 and 1, defines how elongated the blob is. When this value is 1 it means that the region is a circle, however, as it approaches to 0, the region is shaped like an ellipse until it reaches the shape of a line that corresponds to the value 0 [76].

In addition to detecting the various markers, this method also identifies their respective positions and the approximate center of the markers. Subsequently, with the coordinates of the centers of the markers obtained, the software will calculate different angles and distances between them. Each angle is calculated considering two vectors formed by the markers or between one vector formed by the markers and the vertical line perpendicular to the floor. The calculation of the internal product is applied to these segments to obtain the amplitude of the angle. Distances are calculated using two markers and considering an imaginary line segment between them, obtaining their length.

To calculate the distances, it is necessary to know the scale of the image. To obtain the scale we decided to use a rectangle with a known length, put in the left side of the image during image acquisition. As well, to identify the vertical line to the floor, we add the use of a string attached to the wall, with a weight at the bottom end, as it will be seen in the results.

4.2. Software Performance Assessment

To understand the best solution for placing the markers we used ping pong balls and styrofoam balls which were glued with adhesive tape to the subject's clothes.

In this phase we performed calibration tests on the markers. For this we took images with a mobile phone camera, from 4 subjects with different heights and volumes using different combinations of clothes, while using markers with different colors and sizes. In these tests we used markers with the colors blue, green, orange, red, yellow, and white, which we put in clothes with the colors white, brown, red, blue, green, yellow, and gray. Two types of white markers were tested, ping pong balls with a diameter of 40 mm and styrofoam balls with a diameter of 20 mm. All the other colored markers were styrofoam spheres with a diameter of 20 mm. In this step, the background was white.

Afterwards, we also assessed which are the best conditions for image acquisition, such as using flash or not, and using different backgrounds colors to be able to define which are the best controlled background and light. The colors of background that we tried were white, brown, green, red, black, gray, and blue. In this phase, we used a dress form and put the markers in the anatomical landmarks equivalent to a human, namely, both acromions, the jugular notch, the xiphoid appendix, both ASIS and both trochanters. In this step, in addition to taking the images with a mobile phone camera, we also used a photographic camera, with the purpose of understanding the influence of the type of camera used on the software's perception of the image's colors.

The next step included calibration tests and for that we also used a dress form with the markers in the same positions. Then we took different pictures with the camera at different distances from the dress form and at different heights, to understand if the values obtained of the angles and distances between

the markers were kept constant. For this step we also used the previous phone camera and photographic camera, and we analyzed again the use of flash or not. The distances from the wall and the camera height were measured considering the center of the camera lens as the reference point. The dress form was used to guarantee that the markers are always in the same position in all the photos.

4.3. Data Acquisition

After completing the software performance testing phase, we started collecting data from a population composed of the students from the ESSCVP-Lisboa. Although the final objective of the work is to apply this software in the diagnosis of children and adolescents, due to the need of getting consent of each parent, we applied as first test population the adult students of ESSCVP-Lisboa. We chose these students because they were all older than 18 years old, thus avoiding the operationally heavy process of collecting the consent of children's parents. Also, due to Covid-19 and confinement measures during the development of this dissertation, the process of collecting images and data from the participants took longer than expected. This acquisition had to be carried out in person in a closed room with at least 2 people inside, maintaining the safety standards related to the pandemic situation. This also did not allow us to consider the use of an experimental population composed of more participants. This allowed the use of an initial database for the testing and improvement of the software.

With this population, we were also able to test the results obtained with the software on a larger scale of variability of different skin colors, heights, volumes, and clothes. The images of the participants were taken in a laboratory in the facilities of the ESSCVP-Lisboa, to assess their posture. In that laboratory, we could also use the *Contemplas* software (which is a clinical software for posture, motion, and gait analysis), and compare the results obtained between the *Contemplas* software and our software, so we could validate the data calculated with our software.

4.3.1. Description of Participants

The experimental group included a total of 60 participants, however we only selected 57 due to errors (it was missing the red string in some of the images, making impossible to obtain some angle measures) in the acquisition of images from 3 of the participants. These subjects had ages between 18 and 38 years old, heights between 153 and 198.5 cm, weights between 48.3 and 117 kg and 39 people were females and 18 were males. For this experience, the participants had to fulfill a questionnaire that is present in Appendix A.

This questionnaire allowed us to obtain the demographic characteristics of the study population, such as the average and standard deviation of the age, height, and weight, as well as the number of participants, the feminine and masculine gender, are shown in Table 4.4.

Table 4.4: Demographic characteristics of the experimental population

Age	Gender (F/M)	Height (cm)	Weight (kg)
20.9 ± 4.5	39 / 18	166.4 ± 9.3	69.4 ± 16.3

The values related to age allowed us to assess that the average age of the study population was close to that expected from a population composed of young university students. With the other 3 demographic parameters we could verify the variability of participants, volume and heights.

Of the 57 subjects, 30 claimed to have pain in at least one of the cervical, dorsal, and lumbar regions. Of these 30 people, only 8 went to a health professional, from which 4 were diagnosed with a pathology prior to the study. These data from the questionnaire were initially collected in order to have an idea of the existence of evidence or the diagnosis of postural problems in the study population. We would use this last information to classify the subjects' posture. However, few answered that they had gone to a health professional because of these disorders, not having enough data to be used for this purpose. On the other hand, these values also show that less than 30% of the study population who suffered from back pain seek medical help.

4.3.2. Materials and Experimental Setup

The acquisition of the images involved the use of a camera placed on a tripod, facing the subjects. The participants were positioned in front of a wall with green styrofoam markers placed with double tap glue on the anatomical sites previously referred to. The markers were set in the anatomical sites with the help of a physiotherapy specialist.

On the left side of the image next to the subject, a blue rectangle with a length of 20 cm was placed to be used for the scale of the image. And on the right side of the image, we placed on the wall a red string with a weight at the end, to know the perpendicular line to the floor.

For each participant, an anterior view, a posterior view, a left lateral view, and a right lateral view were taken.

4.3.3. Technical Validation

After the phase of image acquisition, these were analyzed with our software and with the *Contemplas* software. The *Contemplas* software was used to calculate the same parameters that our software calculates for posterior comparison of the values obtained. The *Contemplas* software also gave an evaluation of the posture.

During this process, we noticed that some of the markers used were not painted correctly and did not appear as fully colored green circles. Therefore, before processing the images with both software, we had to manually draw green circles with the same dimension as the markers present in the images.

In addition, we also had to manually draw green circles on the markers that only appeared half a circle or less because they were covered by body parts. This situation occurred especially with the markers on the back in the profile images. These circles were drawn in the approximate locations where the actual markers would be.

Furthermore, for image and data protection reasons, we had to cover the faces of all participants in all images with black rectangles.

To compare the values obtained with the *Contemplas* software and our software, we tested the significance difference between the means for each parameter calculated with both software.

Firstly, we analyzed separately the normality of the data of each feature of each software to understand if we should use a parametric or non-parametric statistical test to compare the means. To evaluate the normality of the data we used the Shapiro-Wilk test [79].

Thereafter, we proceed to the analysis of the means for each pair of features. For the comparison of means, we applied the paired t-test if the data of the variable from both software's were parametric. When at least one of the datasets in the pair of features was non-parametric, we applied the Wilcoxon

signed-rank test. All the statistical tests previously mentioned were carried out with a significance level threshold of 0.05 and 0.1 [80].

After a comparison of the parameters, we performed this analysis again for the variables that result as having a statistically significant difference at a significance level of 0.05, this time removing in each pair of datasets the outlier differences between the values.

The paired t-test and the Wilcoxon signed-rank test allow us to know whether we can reject the null hypothesis or not, however not rejecting the null hypothesis is not the same as accepting it. To this extent, we also applied an equivalence test for each pair of data, using as lower and upper boundary values: -0.01 and 0.01; -0.05 and 0.05.

The equivalence statistical test considers as an alternative hypothesis that the difference between the means is inside of a selected interval, being statistically smaller than the boundary values chosen. This statistical test allows us to consider that the two data samples are similar to each other, when accepting the alternative hypothesis and rejecting the null hypothesis [81,82].

In the equivalence test, we also considered paired data and we took into account the results of the Shapiro-Wilk test regarding the normality of the data. After this step, we repeated these analyzes removing the outliers from the parameters that we could not reject the null hypothesis at a significance level threshold of 0.05, with the lower and upper boundary values of -0.05 and 0.05.

The images were also presented to physiotherapy teachers at the ESSCVP-Lisboa to evaluate the posture of the participants. Considering the variables that our software analyzed, the participants' postures were label as: without evidence of postural problems; mild evidence of postural problems; and with moderate to severe evidence.

With the results from the *Contemplas* software and the physiotherapy teachers' assessment we could validate the results obtained with our software.

4.4. Classification

After the assessments done by the physiotherapy specialists and the *Contemplas* software results, we proceed to apply a classifier to the data, which uses the previously defined anatomical parameters calculated by our software, to show a result that indicates the existence or absence of evidence of posture problems.

First, we performed a feature selection to reduce and select the best features to use in the classifiers. For that, we ranked the 33 features using the Orange toolbox. To select the best set of features we ranked them using the Information Gain method, the Chi-squared method, and the Relief method.

Second, we implemented two different widely used algorithms to train the classifier: Linear Discriminant Analysis (LDA) and k-Nearest Neighbors (kNN) [83].

The LDA is a supervised technique used for dimensionality reduction and classification that enables instances of different classes to be far away while samples within the same group as close as possible [84–86]. For this, the algorithm first calculates the distance between the means of different classes and then calculates the distance between the mean and the samples of each class. Subsequently, the LDA method constructs a lower dimensional space that maximizes the ratio of the between-class distance to the within-class variance, achieving maximum discrimination [87,88]. LDA assumes that all classes are linearly separable and have the same covariance matrix and considers that the data of all

classes are described by Gaussian distribution [89]. The performance of this method may decrease due to its sensitivity to peripheral samples, which can increase the number of misclassified samples [85]. This method is common in binary classification but is also appropriate for multiclass classification [86].

The kNN is a non-parametric classification method that considers that objects similar in the input space are also similar in the output space and uses the whole dataset as the model. For the data to be classified, the algorithm finds its k nearest neighbors and returns the category that has the largest category probability [84,90–92]. The algorithm uses a training data set to calculate the distance of the unclassified instance from each one of the training instances, and then ranks them from the nearest to the farthest [83,93]. The proximity between the points is calculated by a distance function, such as Euclidean distance [90,92]. The k value is chosen by the user and will affect the success of classification. With a very small k value, the results will be more affected by the data sparseness and the noisy, ambiguous, or mislabeled points. A larger value of k can help in reducing the change of errors, but at the same time increases the complexity. However, if the k value is too large can decrease the classification performance with the introduction of the outliers from other classes. So, it is necessary to test different k values and choose the one with the best performance [90,94,95].

We started by considering a multiclass classification, labeling the 57 participants into the three groups for the selection of features: no evidence of postural problems, mild evidence, and moderate to severe evidence. After this selection, to obtain the best performance when we applied the kNN algorithm, we obtained the values of error rate using values of k between 2 and 39. We chose the five values of k with which we got the lowest error rate. We then tested the kNN model for each of the five chosen values of k and compared the results obtained with this model and with the LDA model.

Following, we repeated the same procedures of feature selection and subsequent training of the LDA and kNN algorithms, however this time we considered two steps of binary classification. A first level, in which we carried out this process with the 57 participants labeled as without evidence of postural problems or with evidence of postural problems. In the second level, we selected only the 41 participants who were previously labeled as having evidence of postural problems and we divided them into subjects with mild evidence of postural problems and with moderate to severe evidence of postural problems. This process is exemplified in the diagram in figure 4.4. Regarding the implementation of the kNN model, for the first level, we obtained the values of error rate using k values between 2 and 39. In the second level, in which we considered only 41 participants, the k values used were between 2 and 28.

Both LDA and kNN were always trained first with a randomly split of 70% of the data used for training and 30% for testing, and secondly it was applied with a 10-fold cross-validation for later comparison. In the 10-fold cross-validation the data set is divided into 10 subsets of approximately equal size. Then, iteratively, for each fold, one of the subsets is used to assess the performance of the classifier model, and the other nine serve as a training set. The accuracy obtained with this validation method is the average of all ten accuracies of the testing sets [96,97].

Figure 4.4: Diagram of the binary classification of the participants, divided in two levels. In the first level, the 57 participants were classified as belonging to class “without evidence” of postural problems or “with evidence” of postural problems. In the second level, the participants who were labeled as having postural problems were further split into class “with mild evidence” or into class “with moderate to severe evidence”.

5. Results and Discussion

The results were divided in three steps: the results of the testing of the software performance; the comparison of the results obtained with the *Contemplas* software and with our software in the analysis of the data acquired from the study population compose by the students of the ESSCVP-Lisboa; and the training of classifiers with the data calculated with our software from the study population.

5.1. Testing the Software Performance

5.1.1. Selection of the Marker Color

To study the performance of the software, we analyzed what would be the best marker option for the software to detect, in terms of color and size. To this end, we used as markers white ping pong balls with 40 mm of diameter, and styrofoam balls with the colors white, yellow, orange, red, blue, and green with 20 mm of diameter. The markers were placed on four different subjects with different shapes, sizes and wearing clothes of different colors. One participant was wearing a brown shirt and blue pants, another was wearing a red shirt and green pants, another subject was wearing a yellow shirt and grey pants and the last subject was wearing a black shirt and jeans. Each participant used all the different markers.

For each acquisition we used the 29 markers previously chosen during the phase of creation of the software and we took images of the anterior, posterior, left and right lateral views of the subject. All images of each subject with the different colored markers, identified by green circles by the software, are presented in Appendix B.

For each color we tried to reduce the threshold range so the software would only detect the color tones of the markers, differentiating from the background and clothes. Nevertheless, the threshold range had to allow the software to identify the greatest number of markers and the greatest amount of area of each marker. Some of the markers were not detected because less than half a circle was shown in the images, and the software only detects figures from half to full circles.

We first tried using white ping pong ball markers, which have a 40 mm diameter. These were the markers that had been used previously in the work of my classmates in a course of the Integrated Master in Biomedical Engineering and Biophysics at FCUL. Later, we tested using white markers with 20 mm diameter, to understand if it was possible to reduce the size of the marker. By reducing the size, we were able to attach the marker to the clothes more easily without falling because of weight. We were also able to place the center of the marker more accurately at the intended anatomical site.

In the pictures of the figure 5.1, we can observe that when using the ping pong balls some of the markers were not detected in their total area. Some markers had areas with shadows caused by other markers or body parts, which leads to these regions of the marker being considered as having the color gray by the software. It was not possible to increase the threshold to detect these regions without detecting the tone of the gray pants worn by one of the subjects. If we increased the range of the neutral tones detected, the markers would not be detected at all, as they would be confused with gray pants or with the shadows on the wall. This problem was also verified with the white styrofoam markers, as we can observe on the right image of the figure 5.2, where the marker in the ASIS is not detected due to the arm's shadow.

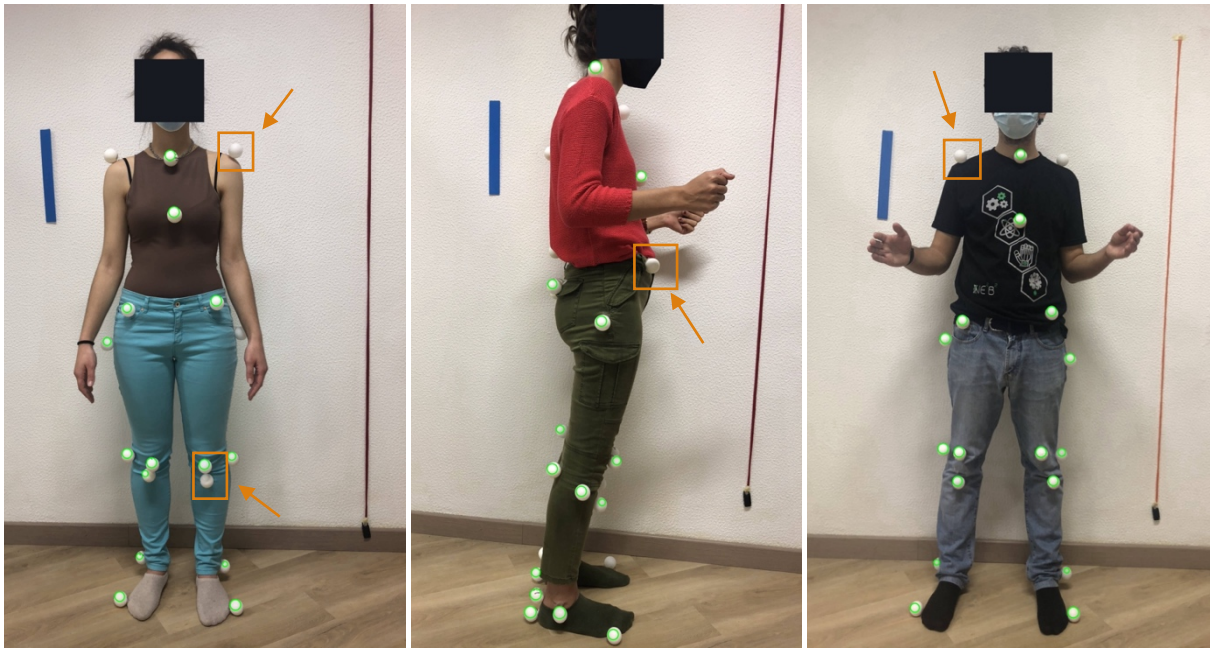


Figure 5.1: Images of the participants using white ping pong balls as markers, taken with a mobile phone camera. The green circles spot the detected area of the markers that were identified by the software. It demonstrates that when these markers have a white wall as a background or when they have shadows caused by other markers or body parts, they are not fully distinguished by the software. These situations are identified with an orange box and arrow. **On the left:** Anterior view of the participant wearing a brown shirt and blue pants. **In the middle:** Right lateral view of the participant wearing a red shirt and green pants. **On the right:** Anterior view of the participant wearing a black shirt and jeans.

Furthermore, as shown by the images in the figure 5.1 and in the figure 5.2, some of the white styrofoam markers and some of the white ping pong ball markers were also not detected, due to the white wall. As the wall was white, a lot of the markers that were in front of the wall were not differentiated from the background by the software because they had the same color. We can observe this situation on the left picture of figure 5.1, where the markers in the acromions were not detected because their color was similar to the color of the background, making it difficult for the software to distinguish. Also, due to the color searched by the software being close to the background color, we verified that in the middle picture of the figure 5.2 the software considered a false marker forming by the hand of the subject.

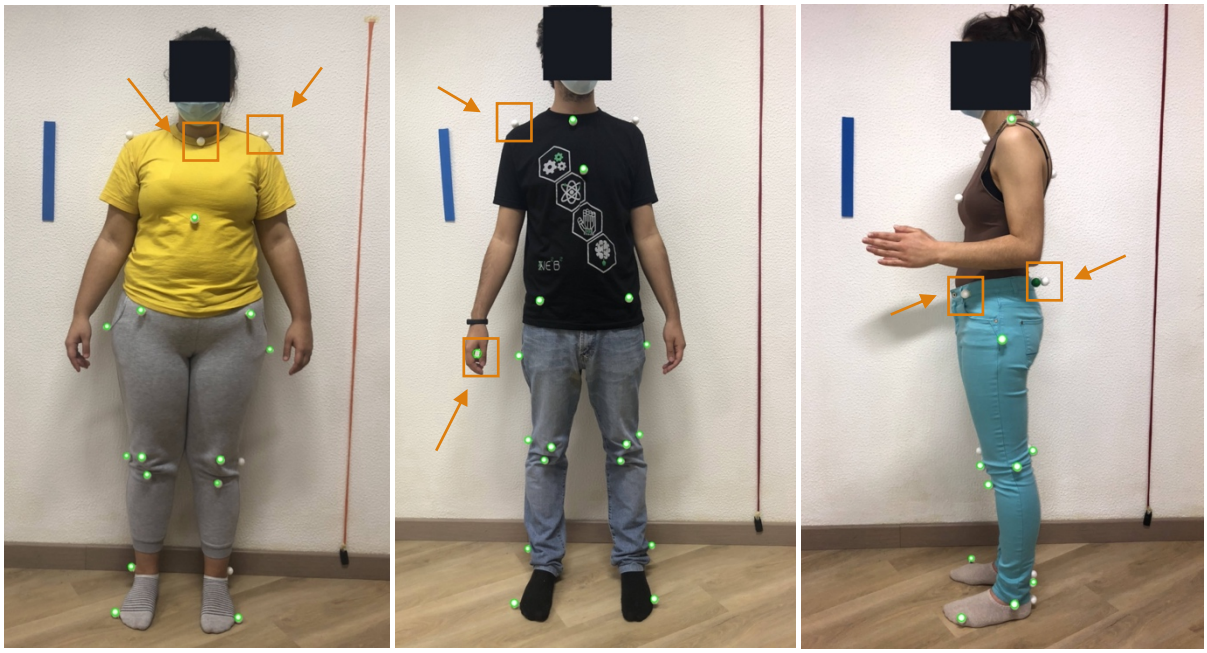


Figure 5.2: Images of the participants using white styrofoam balls as markers, taken with a mobile phone camera. The green circles spot the detected area of the markers that were identified by the software. It demonstrates that when these markers have a white wall as a background or when they have shadows caused by body parts, they are not fully distinguished by the software. Also, it shows the detection of false markers due to similar color between the background and the markers. These situations are identified with an orange box and arrow. **On the left:** Anterior view of the participant wearing a yellow shirt and gray pants. **In the middle:** Anterior view of the participant wearing a black shirt and jeans. **On the right:** Left lateral view of the participant wearing a brown shirt and blue pants.

When using the ping pong balls, there was a greater number of markers detected than when using the 20 mm styrofoam markers. Nonetheless, when observing only the number of markers with their detected area close to the total area, there was a greater amount when using white markers with 20 mm. Therefore, we selected the 20 mm diameter as the markers' size. When testing the other color markers, we decided to use only markers with 20 mm.

The use of the yellow color implied that some of the markers were not correctly detected. From the figure 5.3, we can see that the markers located on the left shoulder are not detected because it is confused with the color of the wall. The confusion with the background color also happens with the yellow markers, because the color of the wall is not pure white and may appear to be yellowish in the images due to the light tone in the room. Also, in this figure 5.3, the marker situated in the xiphoid appendix is not detected because its color is very similar to the color of the t-shirt.

The identification of yellow markers was also affected the use or not of flash, which leads the software to consider different sizes of the marker radius, as we can see by the markers identified with orange rectangles in the figure 5.3. This difference could mean that the calculated center of the marker is not equal.

Figure 5.3: Images of the participants using yellow styrofoam balls as markers, taken with a mobile phone camera. The green circles spot the detected area of the markers that were identified by the software. It demonstrates that the software does not fully distinguish these markers from the background and from the yellow clothes. It also shows that some marker sizes detected by the software are not the same when using flash or not, leading to a different value of radius and center of the marker. These situations are identified with an orange box and arrow. **On the left:** Anterior view of the participant wearing a yellow shirt and gray pants, using flash. **In the middle:** Right lateral view of the participant wearing a red shirt and green pants, using flash. **On the right:** Right lateral view of the participant wearing a red shirt and green pants without using flash.

The threshold values for the different colors were defined so that the largest amount of marker area could be detected. These values were set to detect either regions of the markers that are affected by direct light, becoming brighter, or affected by existing shadows making them darker.

However, in the case of orange and red markers, these threshold values lead to the detection of false markers on the hands and arms of subjects, as can be seen in the figure 5.4 and figure 5.5. This situation occurs especially when flash is not used for the orange markers, and when the flash is used for the red markers, since there are skin tones within the color threshold range selected.

Additionally, the red markers are not correctly detected when the subject is wearing red colored clothes. From the figure 5.5, we can verify that some markers were not detected because they were in contact with the red clothes, which makes the software not consider them to have an approximate shape of a circle. We can observe this situation in the markers located on the xiphoid process and on the left shoulder.

Figure 5.4: Images taken with a mobile phone camera without using flash of the participant wearing a black shirt and jeans, while using orange styrofoam balls as markers. The green circles spot the detected area of the markers that were identified by the software. It demonstrates the detection of false markers on the hands and arms of subjects due to the skin tone being inside the color threshold range, which are identified with an orange box and arrow. **On the left:** Anterior view. **On the right:** Left lateral view.

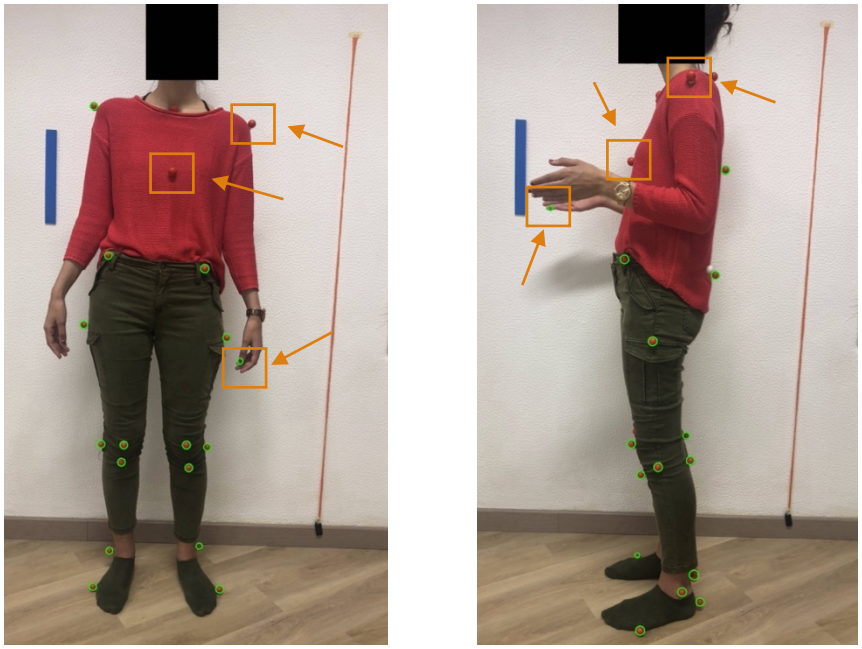


Figure 5.5: Images taken with a mobile phone camera using flash of the participant wearing a red shirt and green pants, while using red styrofoam balls as markers. The green circles spot the detected area of the markers that were identified by the software. It demonstrates the detection of false markers on the hands and arms of subjects due to the skin tone being inside the color threshold range. Also, it shows that the software does not fully distinguish these markers from the red clothes. These situations are identified with an orange box and arrow. **On the left:** Anterior view. **On the right:** Left lateral view.

Regarding the use of blue and green color for the markers, both did not present problems in their detection due to the selected color threshold range. In both cases, there was no confusion between the color of the markers and the color of the background, or the clothes used. However, some of the blue and green markers were not detected because they were partially or completely covered by body parts or folds of clothing. This situation occurred especially in the profile views. We can observe on the left picture in the figure 5.6, that the marker located on the right malleolus is only visible a small part, because it is partially covered by the fold of jeans. We can see in the figure 5.7, that the marker of the xiphoid appendix, on the right picture, and the kyphosis curve marker of both pictures, are also not detected because only a small part of them was visible. This situation also happened with the marker located in the xiphoid appendix on the right picture of the figure 5.6, because it is covered by the arm.

Figure 5.6: Images of the participants using blue styrofoam balls as markers, taken with a mobile phone camera. The green circles spot the detected area of the markers that were identified by the software. It demonstrates that when these markers are covered by body parts or by folds of the clothing, they are not detected by the software, which is identified with an orange box and arrow. **On the left:** Anterior view of the participant wearing a black shirt and jeans. **On the right:** Right lateral view of the participant wearing a yellow shirt and gray pants.

Figure 5.7: Images of the participants using green styrofoam balls as markers, taken with a mobile phone camera. The green circles spot the detected area of the markers that were identified by the software. It demonstrates that when these markers are covered by body parts they are not detected by the software, which is identified with an orange box and arrow. **On the left:** Left lateral view of the participant wearing a red shirt and green pants. **On the right:** Left lateral view of the participant wearing a yellow shirt and gray pants.

Overall, the success rate in detecting the markers for each color was: 75.2% when using white ping pong balls, however approximately 25% of the detected markers had only part of their area detected, causing the software to consider the center of the marker different from the actual marker center. When using the styrofoam balls, the success rate was 64.8% when the color was white; 89.9% when the color was yellow; 95.4% when using the orange color; 93.9% when we used red; 94.2% when the color used was blue; and 96.7% when the markers were green. We obtained a high success rate with the colors orange and red, however when these colors were used the software considered false markers in the skin of some subjects.

After analyzing the different colors to use as markers, we decided to use the green color in the continuation of this work. Since we needed to use a rectangle of known length to obtain the scale of the images, and a string with a weight at the end to know which is the vertical to the floor, we chose the color blue for the rectangle and the color red for the string. We decided to use different colors from the markers and each other, so they would not be confused by the software.

5.1.2. Background Testing

After we selected the color green for the markers, we analyzed the use of these markers in a dress form while varying the color of the background, namely white, brown, blue, red, black, gray, and green.

For each different background, we took one image of the anterior view with flash and another without flash, so we could analyze the influence of luminosity in the detection of the markers. We also took these pairs of photographs with a phone camera and another pair with a photographic camera.

We concluded that the markers could be detected for all the tested background colors, even if we used or not the flash or when we used the two different types of cameras, as can be observed in the figure 5.9, figure 5.10, figure 5.11 and figure 5.12. When we used the green color as background (Figures

5.9f, figure 5.10f, figure 5.11f and figure 5.12f), despite having the same color as the markers, the shades of green were not the same, so it was possible to reduce the threshold to detect only the green of the markers, without confusing it with the background.

As well, the blue rectangle used for the image scale was also detected in all backgrounds, in the images with and without the use of flash captured with both cameras. It is possible that the blue rectangle has been detected in the figure 5.9e, figure 5.10e, figure 5.11e and figure 5.12e, as it is placed on a surface other than the blue background. The blue rectangle was not placed on the wall, because to reduce parallax errors we placed it in approximately the same plane as the markers.

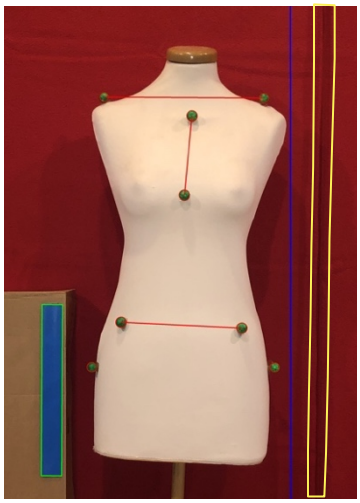
Regarding the red rope used as reference to have the vertical line to the ground, it was detected for most of the backgrounds while using flash and without flash. When using the red background, the software could not identify it (Figures 5.8g, figures 5.9g, figure 5.10g, and figure 5.11g), because the color of the rope and the background were too similar and could not be differentiated. The figure 5.8, figure 5.9, figure 5.10 and figure 5.11 mark with a blue line what the software considered the vertical reference line. The figure 5.8g, figures 5.9g, figure 5.10g, and figure 5.11g have also identified the actual location of the red string with a yellow rectangle. In the pictures with the red background, the blue line should have been inside the yellow rectangle if the software had correctly detected the rope.

The software's ability to detect the green markers, the blue rectangle and the red rope was not significantly influenced by the variation in whether to use flash or not, since for the same background the outcome was similar when analyzing an image with flash and an image without flash.

In addition, the use of the two types of cameras also did not influence the results obtained. With both cameras, the software detected the total area of the green markers and the blue rectangle with all tested backgrounds. As well, the software could not detect the red rope when using the red background in the images of both cameras, succeeding with all the other backgrounds.

(a) (b) (c)

(d) (e) (f)

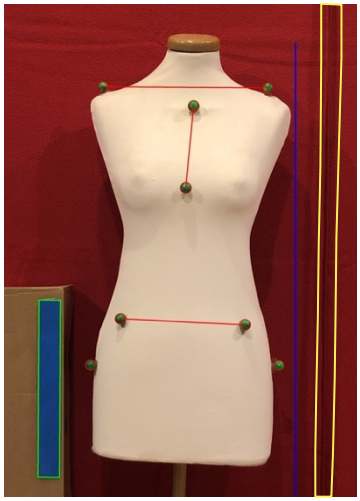


(g)

Figure 5.8: Images captured with a mobile phone camera using flash. We varied the backgrounds of the images to understand the influence of their color on the detection of the markers, the blue rectangle, and the red string by our software. The detected markers are shown circled in red and are connected by red lines that represent the vectors used to calculate the angles. The blue rectangle is displayed with a green outline and the where the software considered the red string to be is represented by a dark blue line. In image (g) the true position of the red string is outlined in yellow. We tested as background the colors (a) white, (b) brown, (c) gray, (d) black, (e) blue, (f) green and (g) red.

(a) (b) (c)

(d) (e) (f)



(g)

Figure 5.9: Images captured with a mobile phone camera without using flash. We varied the backgrounds of the images to understand the influence of their color on the detection of the markers, the blue rectangle, and the red string by our software. The detected markers are shown circled in red and are connected by red lines that represent the vectors used to calculate the angles. The blue rectangle is displayed with a green outline and the where the software considered the red string to be is represented by a dark blue line. In image (g) the true position of the red string is outlined in yellow. We tested as background the colors (a) white, (b) brown, (c) gray, (d) black, (e) blue, (f) green and (g) red.

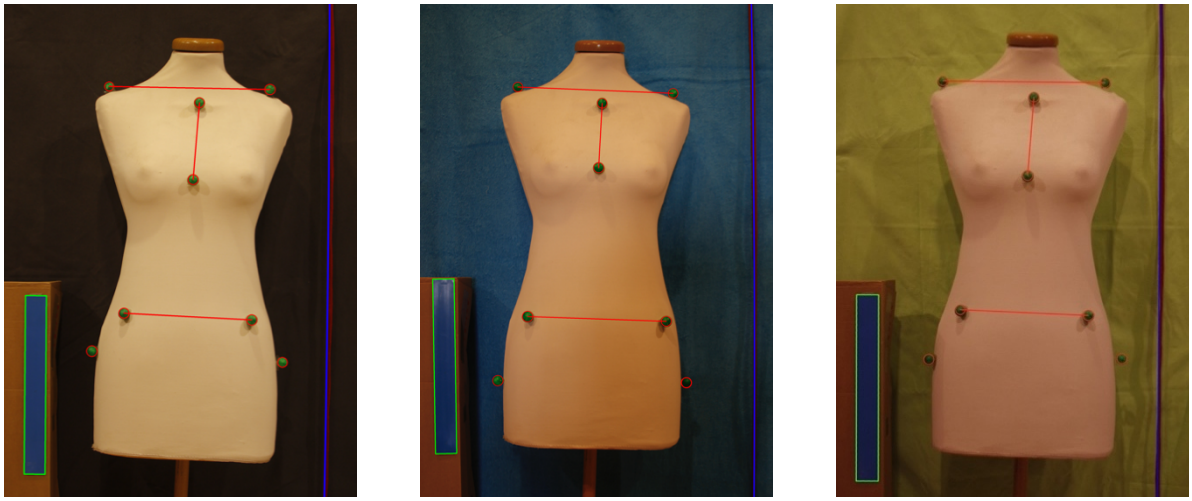
(a) (b) (c)

(d) (e) (f)

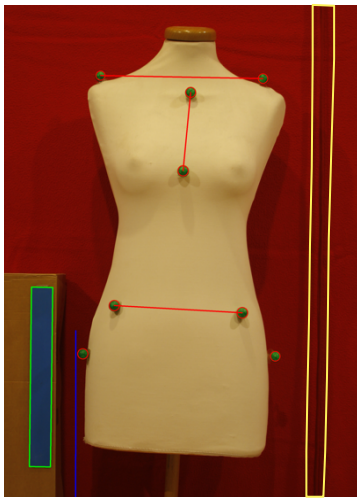
(g)

Figure 5.10: Images captured with a photographic camera using flash. We varied the backgrounds of the images to understand the influence of their color on the detection of the markers, the blue rectangle, and the red string by our software. The detected markers are shown circled in red and are connected by red lines that represent the vectors used to calculate the angles. The blue rectangle is displayed with a green outline and the where the software considered the red string to be is represented by a dark blue line. In image (g) the true position of the red string is outlined in yellow. We tested as background the colors (a) white, (b) brown, (c) gray, (d) black, (e) blue, (f) green and (g) red.

(a) (b) (c)



(d) (e) (f)



(g)

Figure 5.11: Images captured with a photographic camera without using flash. We varied the backgrounds of the images to understand the influence of their color on the detection of the markers, the blue rectangle, and the red string by our software. The detected markers are shown circled in red and are connected by red lines that represent the vectors used to calculate the angles. The blue rectangle is displayed with a green outline and the where the software considered the red string to be is represented by a dark blue line. In image (g) the true position of the red string is outlined in yellow. We tested as background the colors (a) white, (b) brown, (c) gray, (d) black, (e) blue, (f) green and (g) red.

5.1.3. Software Calibration Tests

We have also performed calibrations tests of the distances and angles measured by the software. We took images of the anterior view of the dress form with the camera at different distances from the wall (125 cm, 180 cm, and 250 cm). For each one of these distances, we also took the pictures with the camera at three different heights (at 100 cm, 110 cm, and 120 cm of the floor), and we also captured a version of each image using flash and another without using the flash. This set of images was first obtained with a mobile phone camera and then with a photographic camera. The dress form stayed always in the same position.

Afterwards, for each image presented in the Appendix C, we calculated the following features: AHA_A, ASA, $AJND_L$, $AJND_R$, ADD_L, ADD_R and ASISHA, for comparison of the values. The values obtained for each image are present in Appendix D.

We first analyzed whether or not the use of flash influenced in the parameter's calculation. For this, the data were divided into two sets, with flash and without flash, and for each set the mean, standard deviation and 95% confidence interval were calculated for comparison of the results. Then we did the same procedure to evaluate the influence of the variation of the camera distance from the dress form. We divided the data into three groups, one group for each distance (125 cm, 180 cm, and 250 cm). Finally, we examined the influence of the camera height variation, also dividing the data into three groups, one group for each height (100 cm, 110 cm, and 120 cm).

The 95% confidence intervals for the mean for each variable analyzed are presented for each feature in the plots in the figure 5.12, figure 5.13, figure 5.14, figure 5.15, figure 5.16, figure 5.17 and figure 5.18. In each graph, the influence of luminosity is shown in orange, the influence of distance is in blue, and the influence of height is in green.

We can see from these graphs that for both cameras used, most of the intervals related to the use of flash or not were the most similar among the three analyzed metrics. This is, they presented in general the smallest differences between the means, and they had the most similar ranges of the intervals. This aspect was most noteworthy in the ASA plots (Figure 5.13) and in the $AJND_R$ plot (Figure 5.15). Of the analyzed features, the ones that showed the greatest variations in values, whether using flash or not, for both cameras, were the AHA_A (Figure 5.12) and the ASISHA (Figure 5.18). The parameter that had the smallest amplitude, for the mobile phone camera, was the $AJND_R$ (Figure 5.15a) and for the photographic camera was the AAD_R (Figure 5.17b)

Of the three variables analyzed, the variation of the distance to wall resulted in the smallest amplitudes, this is, for the same distance the values calculated were similar. These intervals were also the most separated from each other, i.e., the mean value obtained at different distances presented the biggest differences among themselves. This can be seen in the ASISHA with the photographic camera plot (Figure 5.18b) and in the AHA_A with the mobile phone plot (Figure 5.12a). The intervals with the biggest amplitudes occurred in the AHA_A (figure 5.12) for both cameras, and with the mobile phone is also observed in the ASISHA plot (Figure 5.18a). For the mobile phone camera, the smallest amplitude was in the calculation of the ASA (Figure 5.13a), and for the photographic camera the lowest amplitudes were in the AAD_R plot (Figure 5.17b) and in the ASA plot (Figure 5.13b). For all the features, the amplitudes obtained were lower with the mobile phone camera than with the photographic camera.

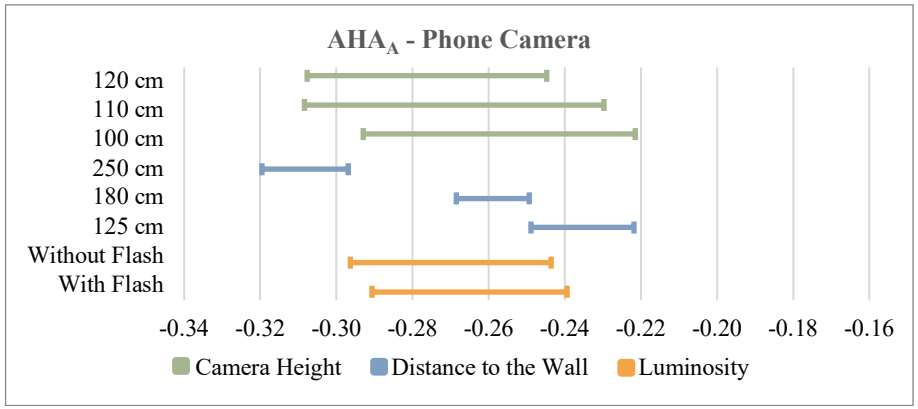
Furthermore, in the analysis of the variation of the distance to the wall, when comparing the three distances with each other, the amplitudes of the intervals were on average quite similar between the three. However, we verified that for the calculation of the angles the values of the measurements obtained were smaller for a distance of 250 cm and larger for a distance of 125 cm. On the contrary, when calculating the distances between markers, the mean of the measurements obtained was smaller

when the camera was 125 cm from the wall, and increased with the increasing of the distance, having been greater for the distance of 250 cm. In the features that are distances of body segments, we also found that the amplitudes of the intervals were smaller when the camera was at 250 cm, and that they increased as the camera approached the wall.

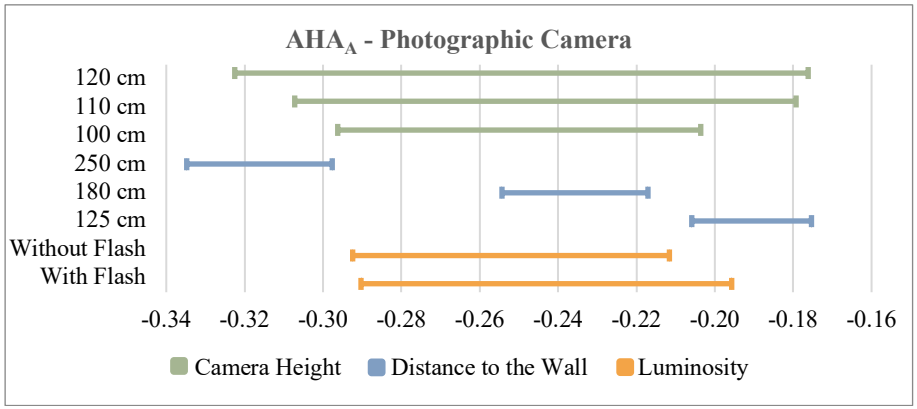
In opposition, the intervals related to height variability have the largest amplitude of the three variables under analysis, which is noticeable, for example, in the AAD_L with the mobile phone camera plot (Figure 5.16a) and ASISHA with the photographic camera plot (Figure 5.18b). In addition, when we varied the height of the camera, the features that are calculations of distances of body segments had the smallest amplitude of the intervals, while the calculation of the two horizontal alignments had the greatest ranges. Regarding the comparison of the values obtained between the three heights, there was a slightly smaller variation of the values when the camera was placed at a height of 120 cm with the photographic camera, and a slightly smaller variation at a height of 110 cm with the mobile phone camera.

These greater variations in the results that occurred when we varied the camera height and the distance to the wall, can be explained due to the fact that the rectangle and the string were not exactly in the same plane as the markers, which can lead to parallax errors. Especially, there are greater differences in the values obtained from the AHA_A and ASISHA, as they are calculated using the string that was placed on the wall plane. The features which are distances of body segments presented in general, smaller differences between their values, as they only use the blue rectangle for their calculation. This blue rectangle was already placed next to the dress form in order to have fewer errors in the calculations made by the software. However, as the dress form has volume, not all the markers were exactly aligned with the rectangle, which led to variations in their values when we changed the position of the camera. Another aspect that may have influenced the results is the fact that the cameras may not have been positioned completely parallel to the dress form, having some lateral rotation of the camera in relation to the wall. One possible way to address this problem is to find the vanishing point of each image to be able to have the blue rectangle and the red sting on different planes of the subject. As well, another possibility is through the images specifications to know the size of the camera sensor, in order to be able to use different cameras and obtain more exact and similar results.

Comparing the results obtained with photographic camera and mobile phone camera, we can observe that in general the former had greater amplitudes in their values, however both presented similar influences on the results due to the variation of each one of the variables analyzed: distance to the wall, camera height and the use or not of the flash.

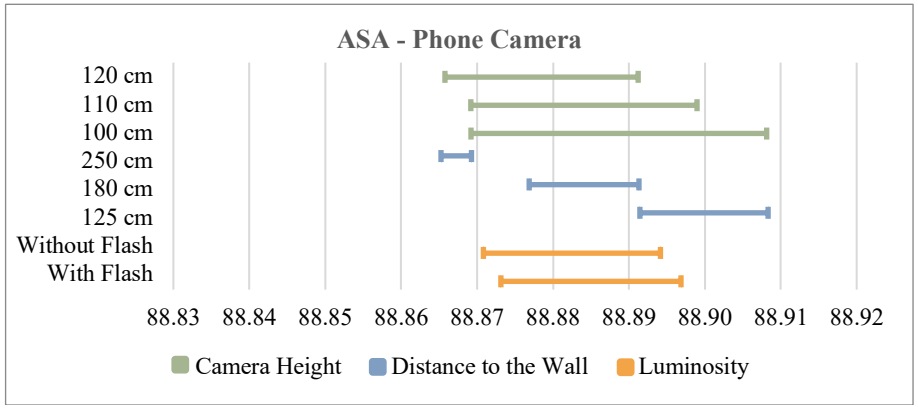


(a)

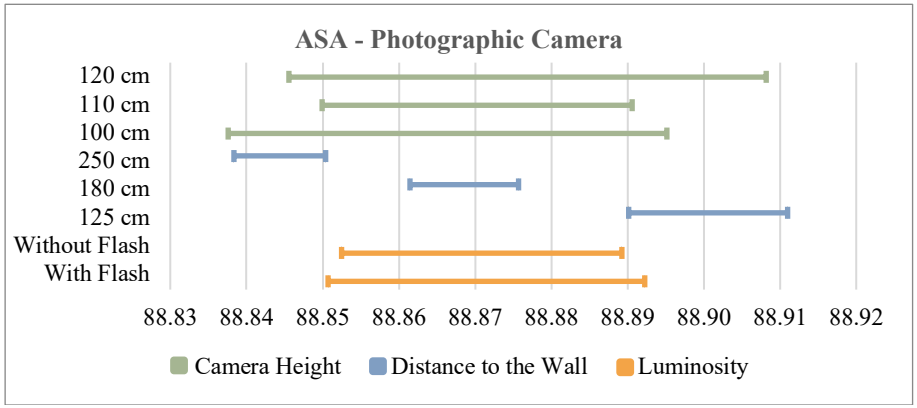


(b)

Figure 5.12: Results of the 95% confidence interval for the mean obtained in the AHA$_A$ calculation. When using the mobile phone camera (a) the values were between: -0.308º and -0.245º for the height "120 cm"; -0.308º and -0.230º for the height "110 cm"; -0.293º and -0.221º for the height "100 cm"; -0.320º and -0.297º for the distance "250 cm"; -0.269º and -0.249º for the distance "180 cm"; -0.249º and -0.222º for the distance "125 cm"; -0.296º and -0.244º "without flash"; -0.291º and -0.239º "with flash". When using the photographic camera (b) the values were between: -0.323º and -0.176º for the height "120 cm"; -0.307º and -0.179º for the height "110 cm"; -0.296º and -0.204º for the height "100 cm"; -0.335º and -0.298º for the distance "250 cm"; -0.254º and -0.217º for the distance "180 cm"; -0.206º and -0.175º for the distance "125 cm"; -0.292º and -0.212º "without flash"; -0.290º and -0.196º "with flash".

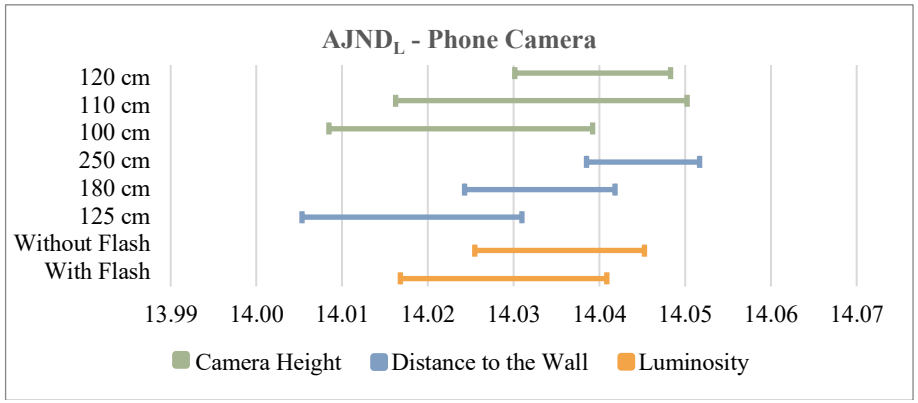


(a)

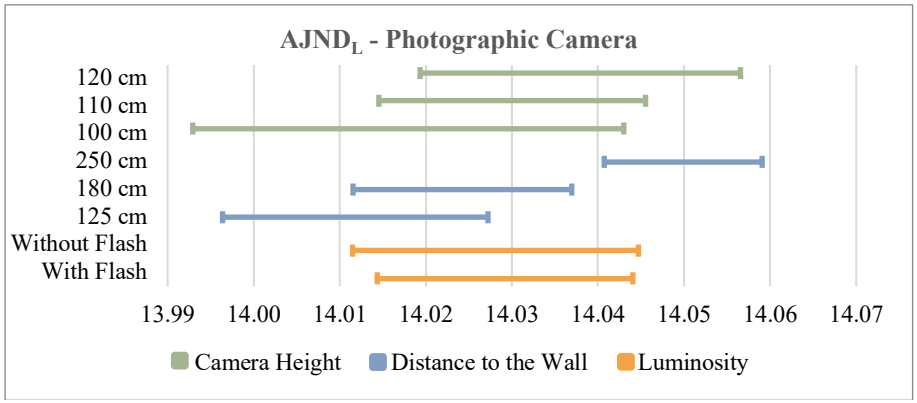


(b)

Figure 5.13: Results of the 95% confidence interval for the mean obtained in the ASA calculation. When using the mobile phone camera (a) the values were between: 88.866 cm and 88.891 cm at the height "120 cm"; 88.869 cm and 88.899 cm for the height "110 cm"; 88.869 cm and 88.908 cm for the height "100 cm"; 88.865 cm and 88.869 cm for the distance "250 cm"; 88.877 cm and 88.891 cm for the distance "180 cm"; 88.891 cm and 88.908 cm for the distance "125 cm"; 88.871 cm and 88.894 cm "without flash"; 88.873 cm and 88.897 cm "with flash". When using the photographic camera (b) the values were between: 88.846 cm and 88.908 cm at the height "120 cm"; 88.85 cm and 88.891 cm for the height "110 cm"; 88.838 cm and 88.895 cm for the height "100 cm"; 88.838 cm and 88.85 cm for the distance "250 cm"; 88.861 cm and 88.876 cm for the distance "180 cm"; 88.89 cm and 88.911 cm for the distance "125 cm"; 88.852 cm and 88.889 cm "without flash"; 88.851 cm and 88.892 cm "with flash".

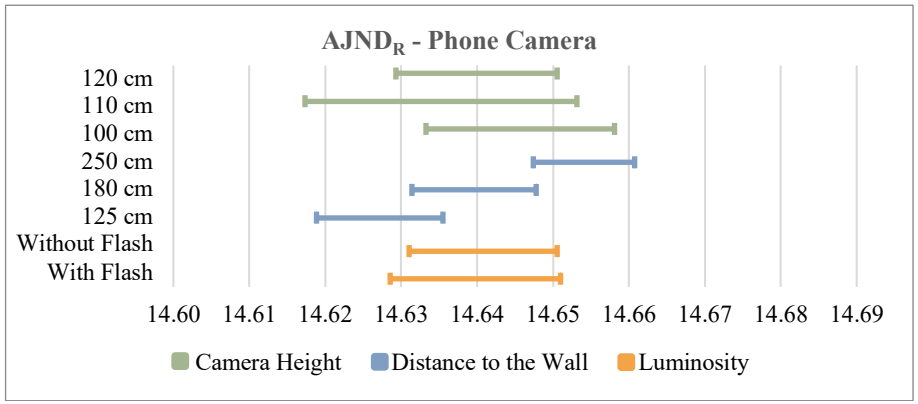


(a)

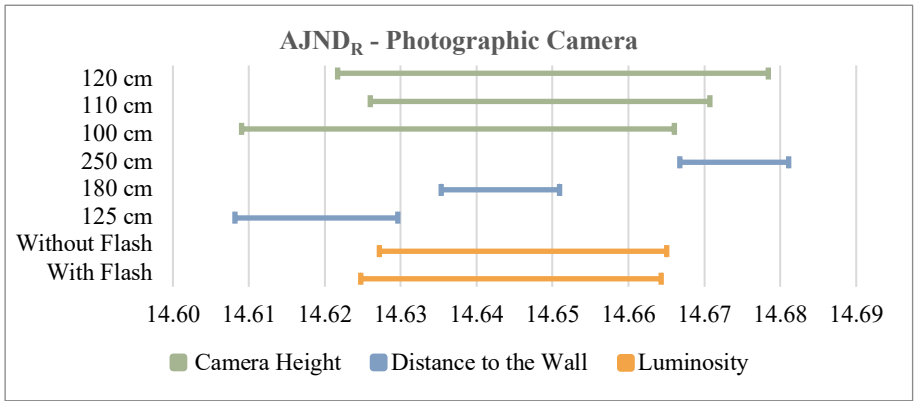


(b)

Figure 5.14: Results of the 95% confidence interval for the mean obtained in the AJND$_L$ calculation. When using the mobile phone camera (a) the values were between: 14.03 cm and 14.048 cm for the height "120 cm"; 14.016 cm and 14.05 cm for the height "110 cm"; 14.008 cm and 14.039 cm for the height "100 cm"; 14.038 cm and 14.052 cm for the distance "250 cm"; 14.024 cm and 14.042 cm for the distance "180 cm"; 14.005 cm and 14.031 cm for the distance "125 cm"; 14.025 cm and 14.045 cm "without flash"; 14.017 cm and 14.041 cm "with flash". When using the photographic camera (b) the values were between: 14.019 cm and 14.057 cm for the height "120 cm"; 14.015 cm and 14.046 cm for the height "110 cm"; 13.993 cm and 14.043 cm for the height "100 cm"; 14.041 cm and 14.059 cm for the distance "250 cm"; 14.012 cm and 14.037 cm for the distance "180 cm"; 13.996 cm and 14.027 cm for the distance "125 cm"; 14.011 cm and 14.045 cm "without flash"; 14.014 cm and 14.044 cm "with flash".

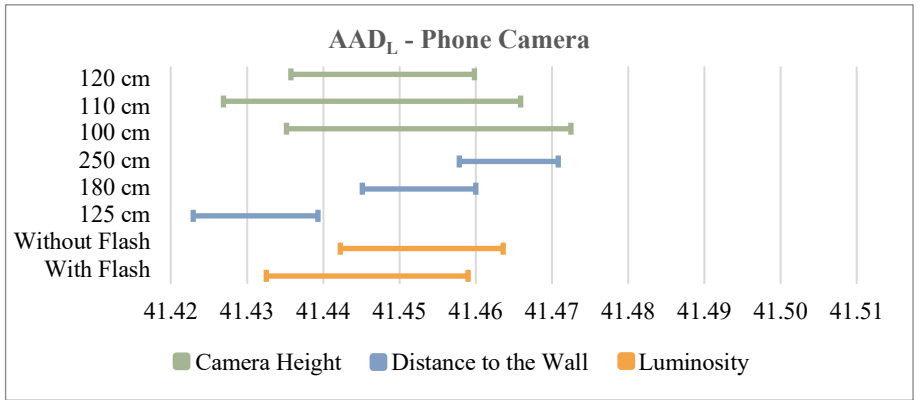


(a)

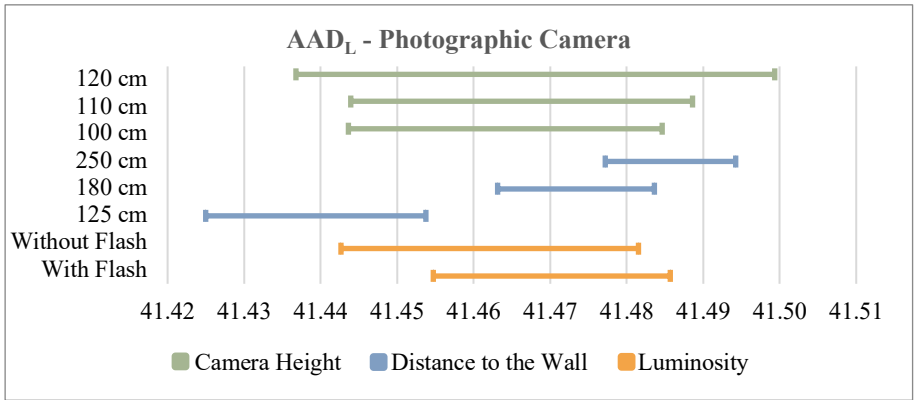


(b)

Figure 5.15: Results of the 95% confidence interval for the mean obtained in the AJND$_R$ calculation. When using the mobile phone camera (a) the values were between: 14.629 cm and 14.651 cm for the height "120 cm"; 14.617 cm and 14.653 cm for the height "110 cm"; 14.633 cm and 14.658 cm for the height "100 cm"; 14.647 cm and 14.661 cm for the distance "250 cm"; 14.631 cm and 14.648 cm for the distance "180 cm"; 14.619 cm and 14.636 cm for the distance "125 cm"; 14.631 cm and 14.651 cm "without flash"; 14.629 cm and 14.651 cm "with flash". When using the photographic camera (b) the values were between: 14.622 cm and 14.678 cm for the height "120 cm"; 14.626 cm and 14.671 cm for the height "110 cm"; 14.609 cm and 14.666 cm for the height "100 cm"; 14.667 cm and 14.681 cm for the distance "250 cm"; 14.635 cm and 14.651 cm for the distance "180 cm"; 14.608 cm and 14.63 cm for the distance "125 cm"; 14.627 cm and 14.665 cm "without flash"; 14.625 cm and 14.664 cm "with flash".

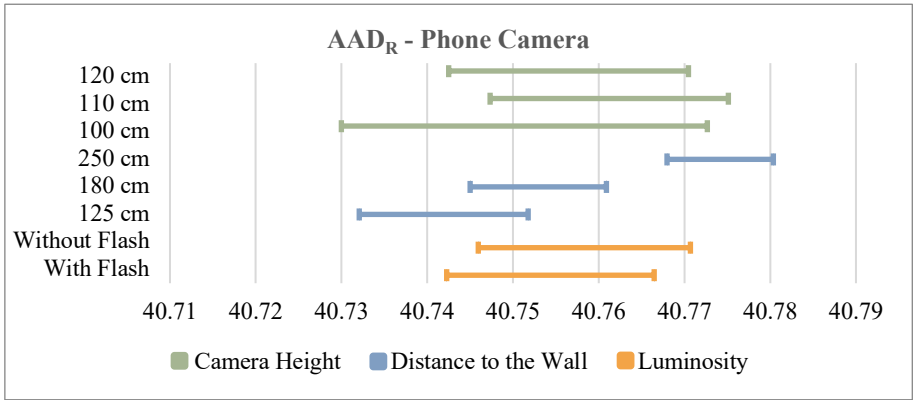


(a)

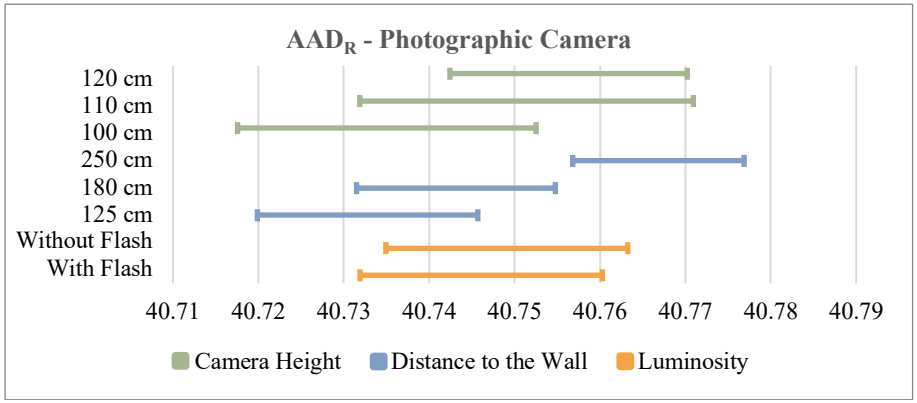


(b)

Figure 5.16: Results of the 95% confidence interval for the mean obtained in the AAD_L calculation. When using the mobile phone camera (a) the values were between: 41.436 cm and 41.46 cm for the height "120 cm"; 41.427 cm and 41.466 cm for the height "110 cm"; 41.435 cm and 41.473 cm for the height "100 cm"; 41.458 cm and 41.471 cm for the distance "250 cm"; 41.445 cm and 41.46 cm for the distance "180 cm"; 41.423 cm and 41.439 cm for the distance "125 cm"; 41.442 cm and 41.464 cm "without flash"; 41.433 cm and 41.459 cm "with flash". When using the photographic camera (b) the values were between: 41.437 cm and 41.499 cm for the height "120 cm"; 41.444 cm and 41.489 cm for the height "110 cm"; 41.444 cm and 41.485 cm for the height "100 cm"; 41.477 cm and 41.494 cm for the distance "250 cm"; 41.463 cm and 41.484 cm for the distance "180 cm"; 41.425 cm and 41.454 cm for the distance "125 cm"; 41.443 cm and 41.482 cm "without flash"; 41.455 cm and 41.486 cm "with flash".

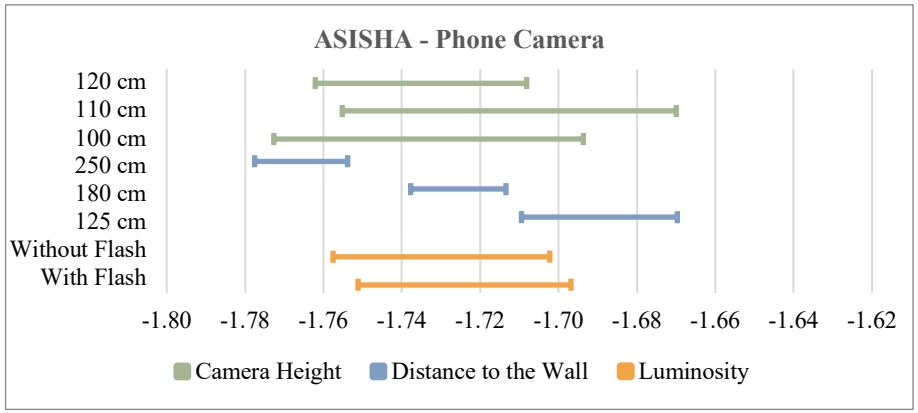


(a)

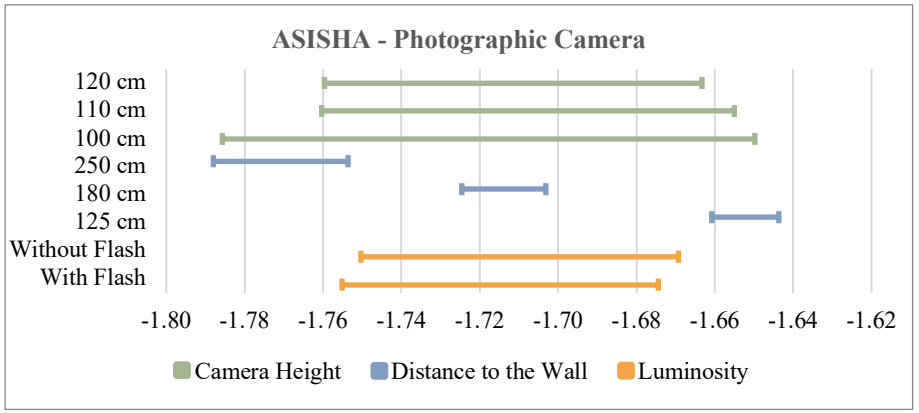


(b)

Figure 5.17: Results of the 95% confidence interval for the mean obtained in the AAD_R calculation. When using the mobile phone camera (a) the values were between40.743 cm and 40.77 cm for the height "120 cm"; 40.747 cm and 40.775 cm for the height "110 cm"; 40.73 cm and 40.773 cm for the height "100 cm"; 40.768 cm and 40.78 cm for the distance "250 cm"; 40.745 cm and 40.761 cm for the distance "180 cm"; 40.732 cm and 40.752 cm for the distance "125 cm"; 40.746 cm and 40.771 cm "without flash"; 40.742 cm and 40.766 cm "with flash". When using the photographic camera (b) the values were between: 40.742 cm and 40.77 cm for the height "120 cm"; 40.732 cm and 40.771 cm for the height "110 cm"; 40.718 cm and 40.753 cm for the height "100 cm"; 40.757 cm and 40.777 cm for the distance "250 cm"; 40.732 cm and 40.755 cm for the distance "180 cm"; 40.72 cm and 40.746 cm for the distance "125 cm"; 40.735 cm and 40.763 cm "without flash"; 40.732 cm and 40.76 cm "with flash".



(a)



(b)

Figure 5.18: Results of the 95% confidence interval for the mean obtained in the ASISHA calculation. When using the mobile phone camera (a) the values were between: -1.762º and -1.708º for the height "120 cm"; -1.755º and -1.67º for the height "110 cm"; -1.773º and -1.694º for the height "100 cm"; -1.778º and -1.754º for the distance "250 cm"; -1.738º and -1.713º for the distance "180 cm"; -1.709º and -1.67º for the distance "125 cm"; -1.758º and -1.702º "without flash"; -1.751º and -1.697º "with flash". When using the photographic camera (b) the values were between: -1.76º and -1.663º for the height "120 cm"; -1.76º and -1.655º for the height "110 cm"; -1.786º and -1.65º for the height "100 cm"; -1.788º and -1.754º for the distance "250 cm"; -1.725º and -1.703º for the distance "180 cm"; -1.661º and -1.644º for the distance "125 cm"; -1.75º and -1.669º "without flash"; -1.755º and -1.674º "with flash".

5.2. Comparison with the *Contemplas Software*

Following the testing of the software performance, we proceeded to collect data from a population composed of students from the ESSCVP-Lisboa.

After collecting the images of the study population, we had to manually draw green circles at the location of some markers that would not be correctly detected by the software. These markers would not be detected correctly because they were not fully painted as green or because they were only visible as half circle or less due to being covered by body parts. These circles were drawn with the same dimension and same location as the markers present in the images. Also, we covered all faces in all images with black rectangles due to image and data protection reasons.

Following this step, we ran our software to evaluate the images of the study population. To validate the values calculated, we also used the *Contemplas* software, which is a professional software for posture and movement analysis. After acquiring the results of both software, we needed to verify whether these outcomes were significantly different or not, so we performed statistical hypothesis testing. In this step, we could not compare the features from the left profile since *Contemplas* software does not use this view to evaluate the posture of subjects. We were only able to compare the data from the anterior, posterior, and right lateral view.

For the statistical hypothesis testing, we first apply the Shapiro-Wilk test to verify if the data set of each parameter from each software was normally distributed or not. If both data sets of the same feature were considered normally distributed, a paired sample t-test was then applied to compare the difference in means of the two samples. Otherwise, if at least one of the samples of the data pair was considered not normally distributed, we would use the Wilcoxon signed-rank test. The feature was considered not normal distributed if the p-value obtained was lower than the significance level thresholds α=0.05. The results of the Shapiro-Wilk test and the consequent decision to apply the paired sample t-test or the Wilcoxon signed-rank test for each one of the features are summarized in the table 5.1, table 5.2 and table 5.3, for the anterior view, posterior view, and right lateral view, respectively.

Table 5.1: Anterior View: Results of the p-value obtained in the statistical analysis with the Shapiro-Wilk test to study the normality of the data related to each parameter, acquired with the *Contemplas* software and with our software. Based on the p-value, we selected the statistical test to analyze the difference in means of each pair of data related to the same feature.

	Software	P-value	Normally Distributed	Statistical Test
AHA_A	*Contemplas*	0.473	Yes	Paired Sample T-Test
	Our	0.432	Yes	
ASA	*Contemplas*	0.321	Yes	Paired Sample T-Test
	Our	0.285	Yes	
$AJND_L$	*Contemplas*	0.344	Yes	Paired Sample T-Test
	Our	0.369	Yes	
$AJND_R$	*Contemplas*	0.050	Yes	Wilcoxon Signed-Rank Test
	Our	0.048	No	
AAD_L	*Contemplas*	0.785	Yes	Paired Sample T-Test
	Our	0.735	Yes	
AAD_R	*Contemplas*	0.903	Yes	Paired Sample T-Test
	Our	0.918	Yes	

	Software	P-value	Normally Distributed	Statistical Test
ASISHA	*Contemplas*	0.837	Yes	Paired Sample T-Test
	Our	0.820	Yes	
$ASISLA_L$	*Contemplas*	0.040	No	Wilcoxon Signed-Rank Test
	Our	0.038	No	
$ASISLA_R$	*Contemplas*	0.680	Yes	Paired Sample T-Test
	Our	0.695	Yes	
$AKLA_L$	*Contemplas*	0.137	Yes	Paired Sample T-Test
	Our	0.123	Yes	
$AKLA_R$	*Contemplas*	0.092	Yes	Paired Sample T-Test
	Our	0.110	Yes	
$ALLL_L$	*Contemplas*	0.087	Yes	Paired Sample T-Test
	Our	0.116	Yes	
$ALLL_R$	*Contemplas*	0.601	Yes	Paired Sample T-Test
	Our	0.599	Yes	

Table 5.2: Posterior View: Results of the p-value obtained in the statistical analysis with the Shapiro-Wilk test to study the normality of the data related to each parameter, acquired with the *Contemplas* software and with our software. Based on the p-value, we selected the statistical test to analyze the difference in means of each pair of data related to the same feature.

	Software	P-value	Normally Distributed	Statistical Test
AHA_P	*Contemplas*	0.588	Yes	Paired Sample T-Test
	Ours	0.690	Yes	
AVA	*Contemplas*	0.023	No	Wilcoxon Signed-Rank Test
	Ours	0.024	No	
TKA	*Contemplas*	2.86E-0.4	No	Wilcoxon Signed-Rank Test
	Ours	3.38E-0.4	No	
LLA	*Contemplas*	0.224	Yes	Paired Sample T-Test
	Ours	0.814	Yes	
PSISHA	*Contemplas*	0.004	No	Wilcoxon Signed-Rank Test
	Ours	0.005	No	
APD_L	*Contemplas*	0.831	Yes	Paired Sample T-Test
	Ours	0.814	Yes	
APD_R	*Contemplas*	0.086	Yes	Paired Sample T-Test
	Ours	0.073	Yes	
$PSISLA_L$	*Contemplas*	0.086	Yes	Paired Sample T-Test
	Ours	0.073	Yes	
$PSISLA_R$	*Contemplas*	0.206	Yes	Paired Sample T-Test
	Ours	0.210	Yes	
$PKLA_L$	*Contemplas*	0.090	Yes	Paired Sample T-Test
	Ours	0.092	Yes	
$PKLA_R$	*Contemplas*	0.857	Yes	Paired Sample T-Test
	Ours	0.865	Yes	
ALA_L	*Contemplas*	0.757	Yes	Paired Sample T-Test
	Ours	0.812	Yes	
ALA_R	*Contemplas*	0.818	Yes	Paired Sample T-Test
	Ours	0.812	Yes	
$PLLL_L$	*Contemplas*	0.460	Yes	Paired Sample T-Test
	Ours	0.465	Yes	
$PLLL_R$	*Contemplas*	0.026	No	Wilcoxon Signed-Rank Test
	Ours	0.026	No	

Table 5.3: Right Lateral View: Results of the p-value obtained in the statistical analysis with the Shapiro-Wilk test to study the normality of the data related to each parameter, acquired with the *Contemplas* software and with our software. Based on the p-value, we selected the statistical test to analyze the difference in means of each pair of data related to the same feature.

	Software	P-value	Normally Distributed	Statistical Test
TKC_R	*Contemplas*	0.033	No	Wilcoxon Signed-Rank Test
	Ours	0.037	No	
LLC_R	*Contemplas*	0.927	Yes	Paired Sample T-Test
	Ours	0.938	Yes	
PLA_R	*Contemplas*	0.046	No	Wilcoxon Signed-Rank Test
	Ours	0.045	No	
KA_R	*Contemplas*	0.575	Yes	Paired Sample T-Test
	Ours	0.614	Yes	
LFA_R	*Contemplas*	0.405	Yes	Paired Sample T-Test
	Ours	0.412	Yes	

The table 5.4, table 5.5 and table 5.6 contain the values of mean and correspondent standard deviation, as well as the range between maximum and minimum values of each parameter analyzed with our software and with the *Contemplas* software. These tables also present the result of the p-value given by the statistical hypothesis testing to compare the difference in means of the pair of the same feature, for the significance level thresholds α=0.05 and α=0.1. This analysis assesses whether the values acquired with the *Contemplas* software and with our software for each feature are statistically significant or not. If the pair of parameters is statistically significant, i.e., if the p-value is less than the significance level threshold, it means that we reject the null hypothesis that the difference in means is equal to zero.

Table 5.4: Anterior View: Results of the mean, standard deviation and extremes of each parameter acquired with *Contemplas* software and with our software, and result of the p-value obtained in the statistical analysis to evaluate the difference in means of the pair composed of the same parameter acquired by the two software. Shows the conclusion based on the p-value, whether the feature pair is significantly different or not when α=0.05 and α=0.1. The $ALLL_L$ was statistically significant when α=0.1, as is identified by the red shading.

		AHA_A	**ASA**	**$AJND_L$**	**$AJND_R$**	**AAD_L**	**AAD_R**	**ASISHA**	**$ASISLA_L$**	**$ASISLA_R$**	**$AKLA_L$**	**$AKLA_R$**	**$ALLL_L$**	**$ALLL_R$**
Mean ± Standard Deviation	Our Software	0.44 ±2.05	89.68 ±2.12	15.68 ±1.92	15.45 ±2.00	41.16 ±3.60	40.64 ±3.41	-0.17 ±2.11	73.25 ±3.37	73.00 ±3.38	159.1 ±4.19	159.11 ±3.9	68.65 ±5.25	68.87 ±4.97
	Contemplas	0.44 ±2.05	89.67 ±2.12	15.67 ±1.93	15.45 ±2.01	41.15 ±3.61	40.64 ±3.43	-0.18 ±2.09	73.25 ±3.39	72.99 ±3.37	159.09 ±4.16	159.11 ±3.88	68.62 ±5.25	68.86 ±4.98
Values Range	Our Software	[-4.33; 4.74]	[85.5; 94.93]	[12.03; 20.30]	[12.12; 20.46]	[32.62; 52.05]	[32.78; 48.11]	[-4.56; 5.03]	[64.72; 79.24]	[65.55; 79.60]	[148.76; 167.06]	[148.66; 165.66]	[58.35; 80.05]	[57.76; 80.82]
	Contemplas	[-4.4; 4.8]	[85.5; 95.0]	[12.0; 20.2]	[12.1; 20.5]	[32.5; 51.9]	[32.7; 48]	[-4.7; 4.9]	[64.6; 79.3]	[65.5; 79.5]	[148.7; 166.9]	[148.6; 165.5]	[58.5; 80.0]	[57.9; 80.9]
P-value		0.677	0.529	0.554	0.438	0.496	0.594	0.355	0.971	0.374	0.393	0.801	0.058	0.298
Statistically Significant α=0.05 / α=0.1		No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No / Yes	No/No

Table 5.5: Posterior View: Results of the mean, standard deviation and extremes of each parameter acquired with *Contemplas* software and with our software, and result of the p-value obtained in the statistical analysis to evaluate the difference in means of the pair composed of the same parameter acquired by the two software. Shows the conclusion based on the p-value, whether the feature pair is significantly different or not when α=0.05 and α=0.1. The PSISHA was statistically significant for both α values, as is identified by the red shading.

		AHA_P	**AVA**	**TKA**	**LLA**	**PSISHA**	**APD_L**	**APD_R**	**$PSISLA_L$**	**$PSISLA_R$**	**$PKLA_L$**	**$PKLA_R$**	**ALA_L**	**ALA_R**	**$PLLL_L$**	**$PLLL_R$**
Mean ± Standard Deviation	Our Software	-0.68 ±1.95	90.03 ±3.93	179.24 ±4.65	178.88 ±5.67	-0.58 ±2.72	40.28 ±3.47	40.11 ±3.3	77.73 ±4.08	78.22 ±3.52	173.51 ±4.53	172.47 ±4.14	173.15 ±5.86	173.59 ±6.44	84.99 ±6.05	84.58 ±5.66
	Contemplas	-0.68 ±1.93	90.04 ±3.96	179.23 ±4.64	178.89 ±5.68	-0.55 ±2.70	40.28 ±3.47	40.11 ±3.29	77.71 ±4.06	78.21 ±3.5	173.5 ±4.54	172.46 ±4.15	173.14 ±5.84	173.59 ±6.44	84.99 ±6.04	84.61 ±5.66
Values Range	Our Software	[-5.66; 3.25]	[77.88; 102.5]	[160.7; 192.66]	[163.23; 190.81]	[-9.90; 5.04]	[33.02; 48.46]	[33.39; 47.73]	[68.87; 84.85]	[70.35; 87.97]	[162.23; 183.23]	[163.43; 182.02]	[155.67; 185.95]	[158.01; 187.95]	[76.23; 99.91]	[72.56; 98.12]
	Contemplas	[-5.5; 3.2]	[77.8; 102.6]	[160.6; 192.6]	[163.2; 190.8]	[-9.8; 5.1]	[33.1; 48.6]	[33.3; 47.8]	[68.9; 84.8]	[70.5; 87.9]	[162.1; 183.3]	[163.5; 182]	[155.7; 185.8]	[157.9; 187.8]	[76.2; 99.9]	[72.7; 98.2]
P-value		0.724	0.700	0.354	0.628	0.047	0.905	0.758	0.141	0.582	0.557	0.407	0.915	0.807	0.95	0.17
Statistically Significant α=0.05 / α=0.1		No/No	No/No	No/No	No/No	Yes/Yes	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No

Table 5.6: Right Lateral View: Results of the mean, standard deviation and extremes of each parameter acquired with *Contemplas* software and with our software, and result of the p-value obtained in the statistical analysis to evaluate the difference in means of the pair composed of the same parameter acquired by the two software. Shows the conclusion based on the p-value, whether the feature pair is significantly different or not when α=0.05 and α=0.1.

		TKC_R	**LLC_R**	**PLA_R**	**KA_R**	**LFA_R**
Mean ± Standard Deviation	Our Software	146.8 ±7.24	151.58 ±12.87	79.57 ±8.12	175.83 ±6.18	95.9 ±5.07
	Contemplas	146.79 ±7.26	151.58 ±12.88	79.59 ±8.13	175.84 ±6.18	95.9 ±5.07
Values Range	Our Software	[131.89; 164.36]	[116.38; 179.81]	[60.92; 98.5]	[161.11; 192.89]	[79.19; 108.68]
	Contemplas	[132; 164.2]	[116.3; 179.9]	[61; 98.7]	[161; 193]	[79.3; 108.6]
P-Value		0.649	0.492	0.278	0.396	0.491
Statistically Significant α=0.05 / α=0.1		No/No	No/No	No/No	No/No	No/No

Since our objective is to verify that the data of both software are approximately equal, we do not want a big type II error. This error occurs when we do not reject the null hypothesis when we should. One way to reduce this type of error is using a greater significance level threshold, so we decided to use the significance level thresholds α=0.05 and α=0.1. The probability of committing a type II error with the size of our study population would still be around 70%.

The statistical hypothesis test for most features allowed us to conclude that they were not significantly different for both significance level thresholds, that is, we could not reject the null hypothesis. For the parameter $ALLL_L$, when the significance level threshold used was α=0.1, the null hypothesis was rejected, meaning that the two populations were different with 90% confidence. As well, for the parameter PSISHA, the null hypothesis was rejected with both alpha values.

Following, we removed the outlier differences from these two features, and we re-analyzed them with the previous statistical hypothesis testing. With the outliers removed, we could conclude that we could not reject the null hypothesis with both significance level thresholds for both parameters. The table 5.7 shows the new results of the features $ALLL_L$ and PSISHA.

Table 5.7: Results of the mean, standard deviation and extremes after removing the outliers from the parameters $ALLL_L$ and PSISHA acquired with *Contemplas* software and with our software, and result of the p-value obtained in the statistical analysis to evaluate the difference in means of the pair composed of the same parameter acquired by the two software. Shows the conclusion based on the p-value, whether the feature pair is significantly different or not when α=0.05 and α=0.1.

		$ALLL_L$	**PSISHA**
Mean ± Standard Deviation	Our Software	68.64 ±5.3	-0.58 ±2.76
	Contemplas	68.62 ±5.3	-0.56 ±2.74
Values Range	Our Software	[58.35; 80.05]	[-9.9; 5.04]
	Contemplas	[58.5; 80.0]	[-9.8; 5.1]
P-Value		0.108	0.106
Statistically Significant α=0.05 / α=0.1		No/No	No/No

Table 5.8: Anterior View: Result of the p-value obtained when applying the equivalence test to evaluate the equivalence between the means of the pair composed of the same parameter acquired by the two software. Shows the conclusion based on the p-value, whether the feature pair is significantly equivalent or not for the boundaries between ΔL=-0.01 and ΔU = 0.01 and between ΔL=-0.05 and ΔU = 0.05, when α=0.05 and α=0.1. The $ALLL_L$ was not statistically significant between ΔL=-0.01 and ΔU = 0.01 when α=0.1, as is identified by the red shading.

		AHA_A	**ASA**	**$AJND_L$**	**$AJND_R$**	**AAD_L**	**AAD_R**	**ASISHA**	**$ASISLA_L$**	**$ASISLA_R$**	**$AKLA_L$**	**$AKLA_R$**	**$ALLL_L$**	**$ALLL_R$**
ΔL = -0.01 ΔU = 0.01	P-Value	0.37	0.42	0.42	0.36	0.49	0.43	0.57	0.21	0.55	0.57	0.33	0.89	0.59
	Statistically Significant α=0.05/α=0.1	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No
ΔL = -0.05 ΔU = 0.05	P-Value	8.29E-0.4	3.75E-0.4	7.23E-0.4	1.05E-0.3	2.93E-0.3	1.91E-0.3	3.40E-0.3	1.27E-0.3	2.56E-0.3	6.01E-0.3	1.21E-0.3	0.05	1.73E-0.3
	Statistically Significant α=0.05/α=0.1	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes / No	Yes/Yes

Table 5.9: Posterior View: Result of the p-value obtained when applying the equivalence test to evaluate the equivalence between the means of the pair composed of the same parameter acquired by the two software. Shows the conclusion based on the p-value, whether the feature pair is significantly equivalent or not for the boundaries between ΔL=-0.01 and ΔU = 0.01 and between ΔL=-0.05 and ΔU = 0.05. The PSISHA was not statistically significant for both boundaries, as is identified by the red shading.

		AHA_P	**AVA**	**TKA**	**LLA**	**PSISHA**	**APD_L**	**APD_R**	**$PSISLA_L$**	**$PSISLA_R$**	**$PKLA_L$**	**$PKLA_R$**	**ALA_L**	**ALA_R**	**$PLLL_L$**	**$PLLL_R$**
ΔL = -0.01 ΔU = 0.01	P-Value	0.30	0.33	0.50	0.39	0.91	0.23	0.30	0.77	0.45	0.38	0.57	0.26	0.28	0.24	0.74
	Statistically Significant α=0.05/α=0.1	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No	No/No
ΔL = -0.05 ΔU = 0.05	P-Value	8.20E-0.5	3.70E-0.3	1.08E-0.3	6.87E-0.4	0.07	5.99E-0.5	1.36E-0.4	1.13E-0.2	3.47E-0.3	1.58E-0.4	7.30E-0.3	2.22E-0.4	1.44E-0.4	2.11E-0.3	3.32E-0.2
	Statistically Significant α=0.05/α=0.1	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/No	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes

Table 5.10: Right Lateral View: Result of the p-value obtained when applying the equivalence test to evaluate the equivalence between the means of the pair composed of the same parameter acquired by the two software. Shows the conclusion based on the p-value, whether the feature pair is significantly equivalent or not for the boundaries between ΔL=-0.01 and ΔU = 0.01 and between ΔL=-0.05 and ΔU = 0.05, when α=0.05 and α=0.1.

		TKC_R	**LLC_R**	**PLA_R**	**KA_R**	**LFA_R**
ΔL = -0.01 ΔU = 0.01	P-Value	0.417	0.49	0.634	0.567	0.449
	Statistically Significant α=0.05 /α=0.1	No/No	No/No	No/No	No/No	No/No
ΔL = -0.05 ΔU = 0.05	P-Value	2.74E-0.3	1.86E-0.3	4.99E-0.3	6.33E-0.3	5.82E-0.4
	Statistically Significant α=0.05 /α=0.1	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes

To better prove the similarity between the values calculated with both software, we applied an equivalence test to evaluate if the difference in means is between two pre-selected boundaries. The alternative hypothesis that this statistical test analyzes is H_1: $\Delta L < \mu_2 - \mu_1 < \Delta U$, with ΔL being the lower boundary and ΔU the upper boundary, and the null hypothesis which we want to reject is H_0: $\mu 2 - \mu 1 \leq \Delta L$ or $\mu 2 - \mu 1 \geq \Delta U$. For the equivalence test, we considered the results of the analysis of the normal distribution of the data with the Shapiro-Wilk, performed previously.

We first select as margins values ΔL=-0.01 and ΔU = 0.01, however as can be seen from the table 5.8, table 5.9 and table 5.10, we could not reject the null hypothesis when the significance level threshold was α=0.05 or when was α=0.1, for any of the parameters. Subsequently, we increase the boundary interval to ΔL=-0.05 and ΔU = 0.05, which allowed us to reject the null hypothesis when α=0.05 for most of the features, except for two of them. For the parameters $ALLL_L$ and the PSISHA, as it is indicated in red shading in table 5.9 and table 5.10, we could only reject the null hypothesis when the significance level threshold was α=0.1.

Then, we repeated the equivalence test for these two parameters, increasing the range to ΔL=-0.06 and ΔU=0.06, and as shown in table 5.11, we were able to reject the null hypothesis with a significance level threshold of 0.05. Therefore, we can say that the differences in the means are within these intervals, and we can consider that the features of each pair are equivalent to each other.

Table 5.11: Result of the p-value obtained when applying the equivalence test to evaluate the equivalence between the means of the features acquired by the two software, for the pair $ALLL_L$ and the pair PSISHA. Shows the conclusion based on the p-value, whether the feature pair is significantly equivalent or not for the boundaries between ΔL=-0.06 and ΔU = 0.06, when α=0.05 and α=0.1

		$ALLL_L$	**PSISHA**
ΔL = -0.06 ΔU = 0.06	P-Value	0.01	0.02
	Statistically Significant α=0.05 / α=0.1	Yes / Yes	Yes / Yes

The parameters $ALLL_L$ and the PSISHA were only not significantly different after we remove the outliers. Outliers have a big impact on the final result, although they have a low frequency, being possible to question whether or not they are significant. Thus, it was decided to remove them, and in the future, it might be worth considering an analysis with a larger sample to assess whether their relative frequency increases and whether or not they are statistically significant. In this results' validation, the outliers demonstrated specific cases in which the absolute difference between the anatomical metric measured by our software and by *Contemplas* software was greater than the limits using the interquartile

range method. In the future, a possible step is to verify whether or not these outliers are a significant representation of a problem in calculations made by our software in specific situations in the image analysis.

Another possible limitation concerns the use of the *Contemplas* software. We used the *Contemplas* as the clinical software for postural analysis because there was already prior knowledge of how it operates, being one of the software used in the laboratories of ESSCVP-Lisboa. Nevertheless, there are few scientific literature that refers the use of this software, or that presents information about its reliability in postural assessment. Most of the information accessible is present on the website of the company that owns this software. Although recurrently used by physiotherapy specialists, this lack of scientific studies can be considered a limitation.

5.3. Classification

Following the comparison of the data acquired between our software and the *Contemplas* software, the images of the participants were presented to two specialists in physiotherapy for them to do the posture assessment of each subject. Afterwards, we proceed to use the study population data to evaluate the performance of the classifiers. First, we considered a multiclass classification, and then a binary classification with two levels.

5.3.1. Postural Assessment by Physiotherapy Specialists and *Contemplas*

The specialist assessment and the *Contemplas* software assessment allowed us to consider that from a population of 57 participants, 16 did not present any evidence of posture problems and 41 present evidence of posture problems. From these 41 subjects, 25 presented mild evidence of posture problems and 16 presented moderate to severe evidence.

In most evaluations of the participants' postures made by the physiotherapy specialists, there was agreement in the assessment made. However, there were cases in which there was a slight discrepancy in the intensity of evidence of postural problems presented. In these situations, the assessment most similar to the assessment made by the *Contemplas* was the one taken into account. Our software is able to carry out an analysis with greater certainty in the measurement of angles and lengths of body segments than the analysis carried out directly by a health professional only using the photographs of the subjects. Thus, our software allows identifying evidence of postural problems when these are still very slight, enabling detection at an earlier stage.

5.3.2. Feature Selection

In order to analyze the performance of the classifiers, we previously needed to perform a feature selection. To optimize the classification process, we needed to obtain the best set of the parameters, as some of the parameters calculated by the software may become redundant for the classification of the subject. To perform the selection of the features to apply in the classifier, we worked with the Orange toolbox in Python.

We started to consider a classifier of multiclass composed of the 57 subjects labeled in 3 classes: without evidence, with mild evidence and with moderate to severe evidence of postural problems. We ranked the features using the following feature selection methods: the information gain method, the chi-

squared method and the ReliefF method. After we tested different sets of features for each method, we selected the set that presented the best confusion matrix and the best value of accuracy (61.40%) and kappa coefficient (0.40), which using the ReliefF method were: ASA, $AJND_L$, AAD_L, AAD_R, PSISHA, $PSISLA_{R,}$ APD_L, APD_R and LLC_L.

Afterwards, we studied a binary classification, for which we divided the feature selection in two levels: first we selected the features with the 57 participants labels as without evidence of postural problems and with evidence of postural problems; secondly, we took only the 41 participants that were previously labels as with evidence of postural problems, and we divided in subjects with mild evidence of postural problems and with moderate to severe evidence of posture problems.

In each one of these two cases, we repeated the procedure that we performed for the multiclass situation to select the best set of parameters. For the first case, we selected the following parameters with the information gain method: AHA_A, ASA, AAD_L, TKA, LLA, $PKLA_L$, $PKLA_R$, APD_L, APD_R, $PLLL_R$, TKC_L and LLC_L, which gave an accuracy of 78.95% and a kappa coefficient of 0.48. For the second case the best accuracy of 75.61% was obtained using the chi-squared method when the selected parameters were: $ASISLA_R$, $AKLA_R$, AAD_L, AAD_R, APD_L, APD_R and LLC_L, which had a kappa coefficient of 0.44. These values of the kappa coefficient present moderate strength of agreement.

As well, the feature selection allowed us to know which parameters were the least important to be obtained, and in this way, we could understand which markers were not that crucial. Therefore, we could reduce the number of markers used, making the setup process to obtain the images simpler. Observing the three sets of features that were selected, we can verify that none contained parameters that use the markers of the following anatomical locations: right patella, right tibial tuberosity, left calcaneus, the two lateral condyles of the knee and the two fifth metatarsal. On the other hand, it is also necessary to consider that this feature selection was based on a population with subjects over 18 years old. It is possible that in a population composed of school-age children, these markers may prove to be necessary for their postural classification. In addition, they may also be important for the evaluation of other possible postural problems. Thus, it might be worth considering keeping these anatomical markers when testing the software with a population of school-age children.

Following this step, we implemented algorithms of classification for each one of the 3 previous situations separately. First, we used the LDA model and then the kNN model. For each classification model, we first tested the selected dataset by applying the train/test split method, and after with the 10-fold cross-validation method.

5.3.3. Classification Performance

In the analysis of the classification performance, for each one of the three situations we first applied the train/test split method, with which we randomly generated a training data frame with 70% of data and a test data frame with the remaining 30%. After, we applied the 10-fold cross-validation method. For both validation methods we implemented first the LDA algorithm and then the kNN algorithm. We evaluated the performance of the classifiers using the confusion matrices, and the respective values of accuracy, and f1-scores values of each class.

When we implemented the kNN method, we first plotted the values of error rate in function of the k value, with the k value ranging between 2 and 39 for the multiclass classification and for the first level of the binary classification. For the second level of the binary classification, we varied the value of k between 2 and 28, as the number of subjects in this dataset was smaller. After, we selected the first five

k values that corresponded to the five lowest values of error rate, and we ran the kNN algorithm for these five k values, to later compare the confusion matrices obtained for each one.

This entire process was applied in the case of multiclass classification and in the case of binary classification, in the later it was applied at both levels separately.

5.3.3.1. Multiclass classification

Our first step was to analyze the multiclass classification. We started with the train/test split validation method, and later we used the 10-fold cross-validation method.

Train/Test Split Validation Method

In this part, we evaluated the results obtained when we used the train/test split validation method, in which 70% of the data were for training and 30% for testing. Then, we first implemented the LDA algorithm, with which we get the confusion matrix of figure 5.19.

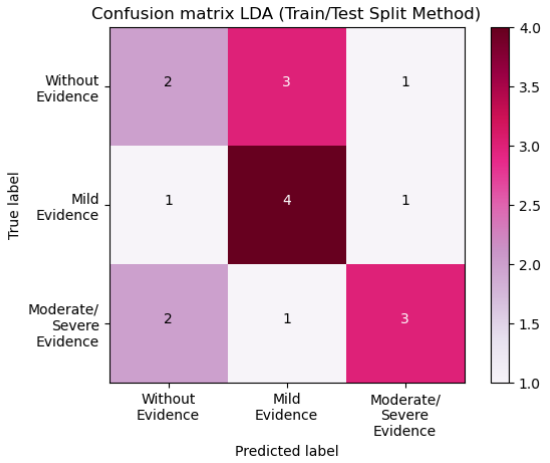


Figure 5.19: Confusion matrix for the three classes ("without evidence", "mild evidence" and "moderate to severe evidence") when the LDA model was applied with the train/test split validation method. It shows the number of correct predictions for each of the three classes (2 for "without evidence", 4 for "mild evidence" and 3 for "moderate to severe evidence") and the number of the wrong predictions (4 for "without evidence", 2 for "mild evidence" and 3 for "moderate to severe evidence"). The kappa coefficient is 0.25.

For the kNN model, we first plotted values of the error rate when we varied the k value from 2 to 39 (Figure 5.20), to assess which k values corresponded to the lowest error rate and select the first five best values. By figure 5.20, we selected the k=3, k=4, k=5, k=6 and k=9.

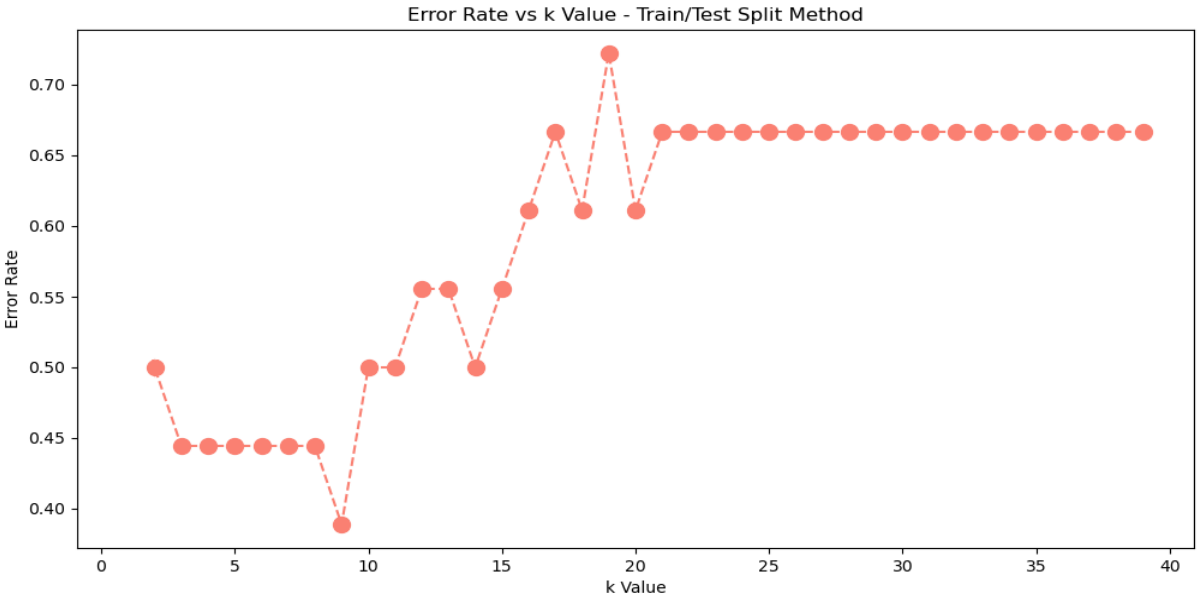


Figure 5.20: Graph of the error rate variation in function of k value when implementing the kNN algorithm with the train/test split validation method to classify the subjects into the three classes: “without evidence”, “with mild evidence” and “with moderate to severe evidence”. The five k values that correspond to the first five lowest error rates are k=3, k=4, k=5, k=6 and k=7.

Afterwards, we analyzed the confusion matrix obtained for each one of the k values selected: k=3 and k=4 (Figure 5.21); k=5 and k=6 (Figure 5.22) and k=9 (Figure 5.23).

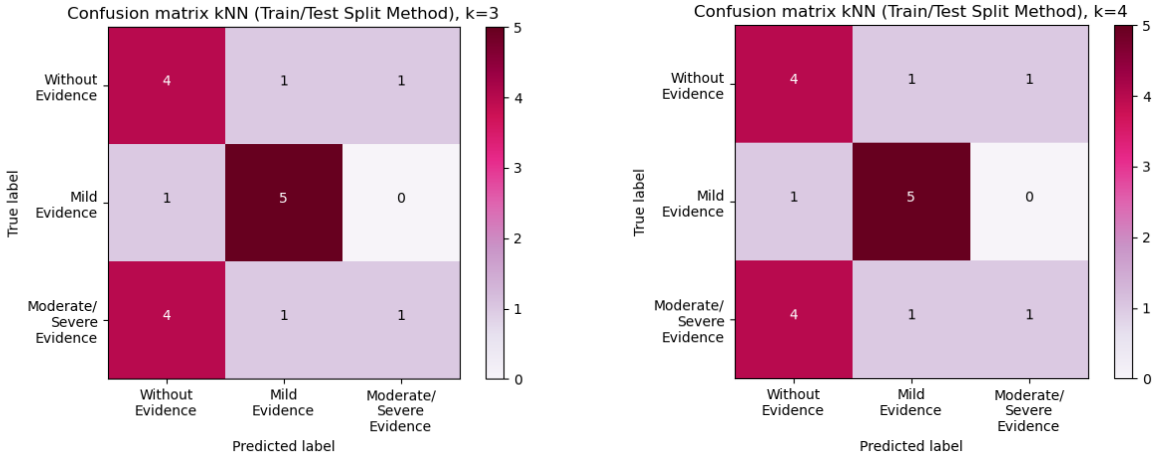


Figure 5.21: Confusion matrix for the three classes (“without evidence”, “mild evidence” and “moderate to severe evidence”) when the kNN model was applied with the train/test split validation method. It shows the number of correct predictions for each of the three classes (4 for “without evidence”, 5 for “mild evidence” and 1 for “moderate to severe evidence”) and the number of the wrong predictions (2 for “without evidence”, 1 for “mild evidence” and 5 for “moderate to severe evidence”). The kappa coefficient is 0.33. **On the left:** we used k=3. **On the right:** we used k=4.

Figure 5.22: Confusion matrix for the three classes ("without evidence", "mild evidence" and "moderate to severe evidence") when the kNN model was applied with the train/test split validation method. It shows the number of correct predictions for each of the three classes (4 for "without evidence", 5 for "mild evidence" and 1 for "moderate to severe evidence") and the number of the wrong predictions (2 for "without evidence", 1 for "mild evidence" and 5 for "moderate to severe evidence"). The kappa coefficient is 0.33. **On the left:** we used k=5. **On the right:** we used k=6.

Figure 5.23: Confusion matrix for the three classes ("without evidence", "mild evidence" and "moderate to severe evidence") when the kNN model with k=9 was applied with the train/test split validation method. It shows the number of correct predictions for each of the three classes (4 for "without evidence", 6 for "mild evidence" and 1 for "moderate to severe evidence") and the number of the wrong predictions (2 for "without evidence", 0 for "mild evidence" and 5 for "moderate to severe evidence"). The kappa coefficient is 0.42.

With these confusion matrices obtained with the kNN model and with the confusion matrix of the LDA model (Figure 5.19), we calculated for each one the accuracy value and the f1-score values of each class. These results are present in the chart of the figure 5.24, the accuracy values are in orange, the f1-score values of the class "without evidence" are in pink, the f1-score values of the class "with mild evidence" are in blue and f1-score values of the class "with moderate to severe evidence" are in red.

We can observe in the figure 5.24, that the highest accuracy value (61.11%) was obtained when we used the kNN algorithm with k=9, and the lowest accuracy value (50.00%) with the LDA model.

When analyzing the f1-scores of the kNN models, we could conclude that the f1-score values of the class "with mild evidence" were the highest among the three classes (70.59% – 76.92%). The f1-score values of the class "with moderate to severe evidence" were considerably lower compared with the other two classes, between 25.00% and 28.57%. The f1-scores values of the class "without evidence" were between 53.33% and 66.67%.

In addition to the kNN model with k=9 having the best the accuracy value, it also reached the highest f1-score value of the class “with moderate to severe evidence” compared to the other four kNN models. As well, it had the highest f1-score value of the class “without evidence” of all the models tried in this step. In relation to the f1-score metric for the class “with mild evidence”, the kNN model with k=9 had the lowest values of all models tested.

The LDA model, despite having the lowest accuracy values, presented the smallest difference between the f1-score values of three classes and the highest value of f1-score of the class “with moderate to severe evidence” of all the models. However, this model also had lowest values of f1-score of the classes “without evidence” and “with mild evidence”.

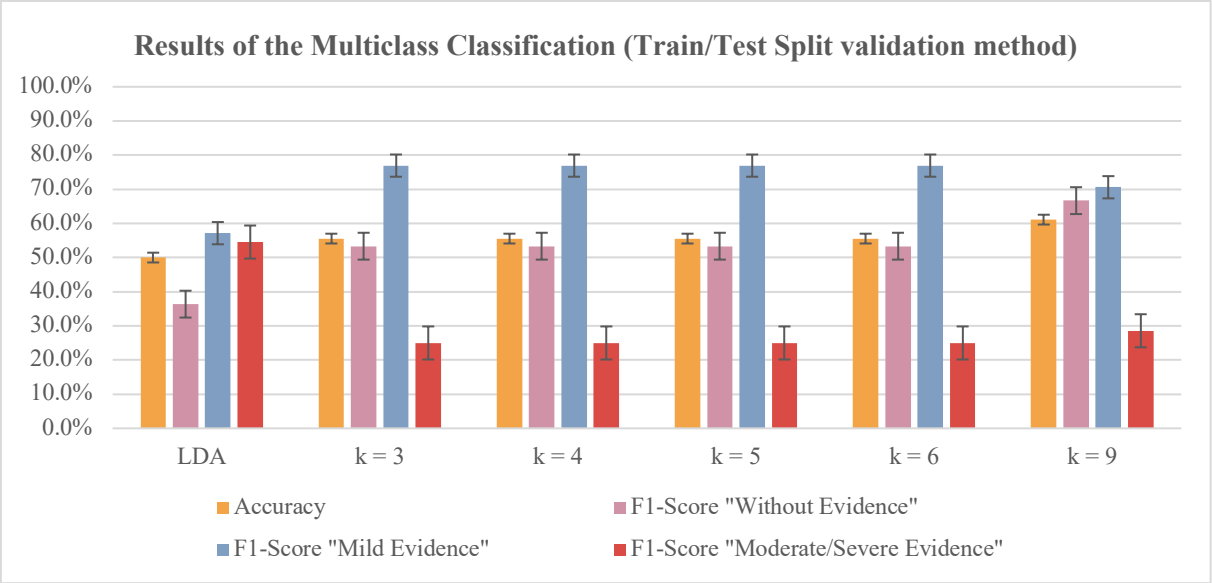


Figure 5.24: Results of the accuracy and f1-scores of each class obtained after the multiclass classification of all subjects with the LDA model and the five selected kNN models, when the train/test split validation method was applied. The maximum accuracy (61.11%) was obtained with the kNN model with k=9, and the minimum accuracy (50.00%) with the LDA model. The highest f1-score for the “without evidence” class (66.67%) was reached with the kNN model with k=9, and the lowest value (53.33%) with the LDA model. The maximum f1-score for the “mild evidence” class (76.92%) was achieved with the kNN models with k=3, k=4, k=5 and k=6, and the minimum value (57.14%) with the LDA model. The highest f1-score for the “moderate to severe evidence” class (54.55%) was reached with the LDA model, and the lowest values (25.00%) with the kNN models with k=3, k=4, k=5 and k=6.

Cross-Validation Method

After testing the LDA model and the kNN model with the train/test split validation method, we assessed the same classifier models this time using 10-fold cross-validation.

We first ran the LDA algorithm, with which we obtained the confusion matrix presented in the figure 5.25.

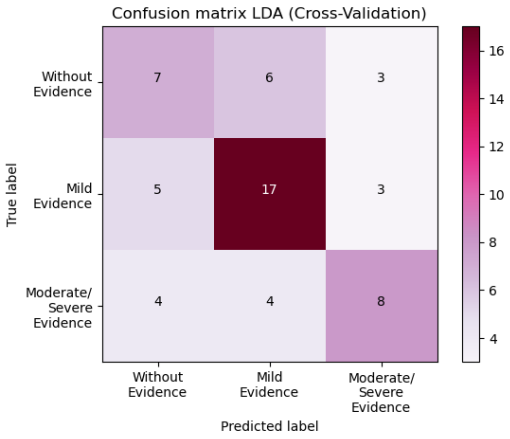


Figure 5.25: Confusion matrix for the three classes ("without evidence", "mild evidence" and "moderate to severe evidence") when the LDA model was applied with the train/test split validation method. It shows the number of correct predictions for each of the three classes (2 for "without evidence", 4 for "mild evidence" and 3 for "moderate to severe evidence") and the number of the wrong predictions (4 for "without evidence", 2 for "mild evidence" and 3 for "moderate to severe evidence"). The kappa coefficient is 0.32.

Regarding the implementation of the kNN model, the plotting of the error rate in function of the k value (Figure 5.26), allowed us to choose as the five values of k to test the following: k = 4, k=5, k=6, k=8 and k=12. The confusion matrices of the kNN algorithm using each one of selected k values are shown in the figures: figure 5.27, figure 5.28 and figure 5.29.

Figure 5.26: Graph of the error rate variation in function of k value when implementing the kNN algorithm with the cross-validation method to classify the subjects into the three classes: "without evidence", "with mild evidence" and "with moderate to severe evidence". The five k values that correspond to the first five lowest error rates are k=4, k=5, k=6, k=8 and k=12.

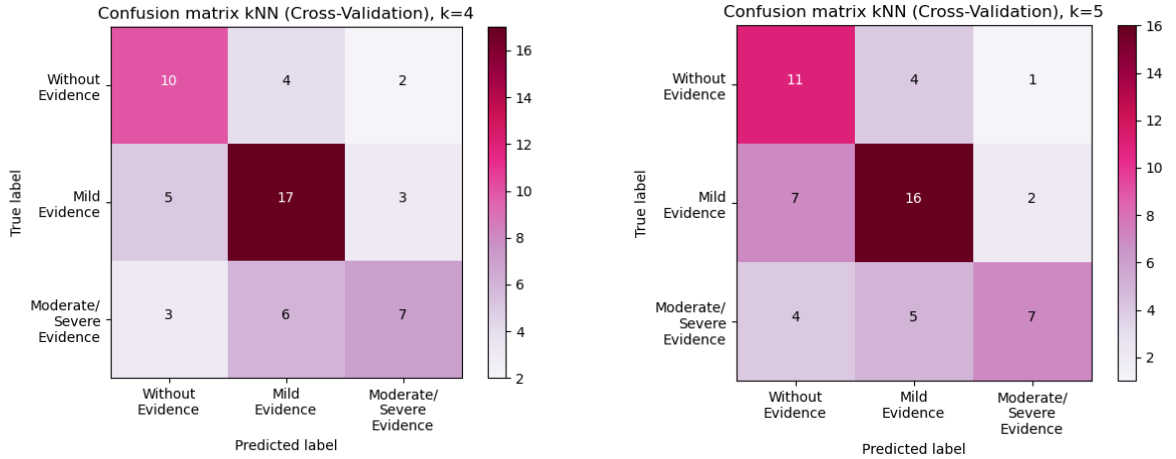


Figure 5.27: Confusion matrix for the three classes ("without evidence", "mild evidence" and "moderate to severe evidence") when the kNN model was applied with the cross-validation method. **On the left:** we used k=4. It shows the number of correct predictions for each of the three classes (10 for "without evidence", 17 for "mild evidence" and 7 for "moderate to severe evidence") and the number of the wrong predictions (6 for "without evidence", 8 for "mild evidence" and 9 for "moderate to severe evidence"). The kappa coefficient is 0.37. **On the right:** we used k=5. It shows the number of correct predictions for each of the three classes (11 for "without evidence", 16 for "mild evidence" and 7 for "moderate to severe evidence") and the number of the wrong predictions (5 for "without evidence", 9 for "mild evidence" and 9 for "moderate to severe evidence"). The kappa coefficient is 0.38.

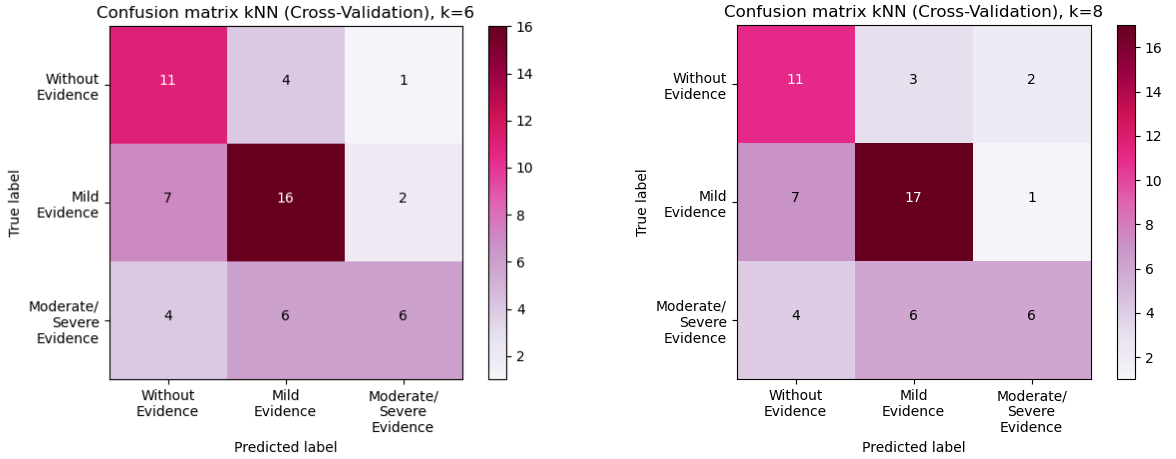


Figure 5.28: Confusion matrix for the three classes ("without evidence", "mild evidence" and "moderate to severe evidence") when the kNN model was applied with the cross-validation method. **On the left:** we used k=6. It shows the number of correct predictions for each of the three classes (11 for "without evidence", 16 for "mild evidence" and 6 for "moderate to severe evidence") and the number of the wrong predictions (5 for "without evidence", 9 for "mild evidence" and 10 for "moderate to severe evidence"). The kappa coefficient is 0.35. **On the right:** we used k=8. It shows the number of correct predictions for each of the three classes (11 for "without evidence", 17 for "mild evidence" and 6 for "moderate to severe evidence") and the number of the wrong predictions (5 for "without evidence", 8 for "mild evidence" and 10 for "moderate to severe evidence"). The kappa coefficient is 0.38.

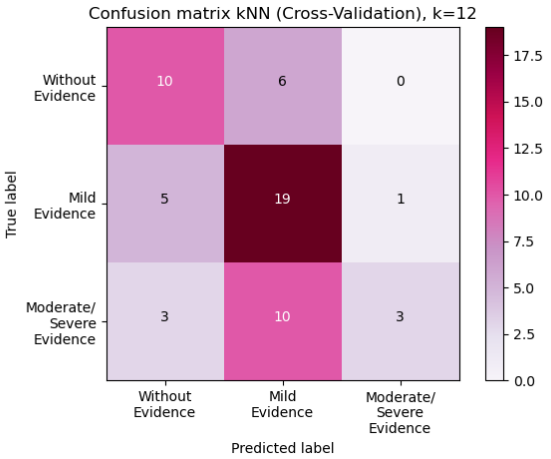


Figure 5.29: Confusion matrix for the three classes ("without evidence", "mild evidence" and "moderate to severe evidence") when the kNN model with k=12 was applied with the cross-validation method. It shows the number of correct predictions for each of the three classes (10 for "without evidence", 19 for "mild evidence" and 3 for "moderate to severe evidence") and the number of the wrong predictions (6 for "without evidence", 6 for "mild evidence" and 13 for "moderate to severe evidence"). The kappa coefficient is 0.30.

Subsequently, we analyzed the f1-score results for each class (pink – class "without evidence", blue – class "with mild evidence" and red – class "with moderate to severe evidence") and accuracy (orange) for each of the tested models, which are presented in the chart in figure 5.30.

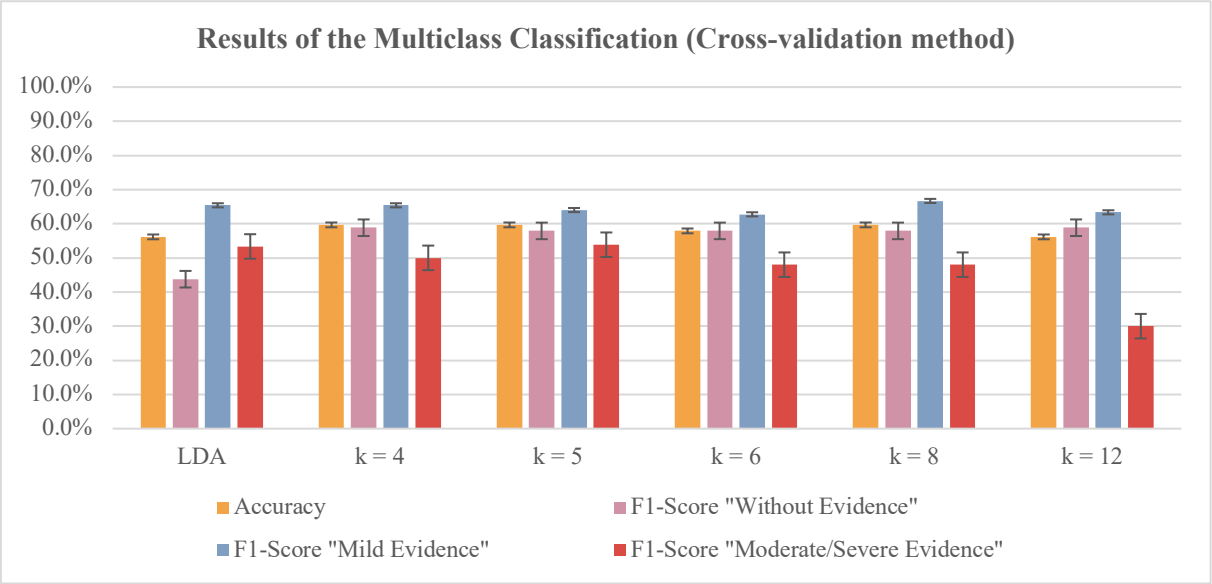


Figure 5.30: Results of the accuracy and f1-scores of each class obtained after the multiclass classification of all subjects with the LDA model and the five selected kNN models, when the cross-validation method was applied. The maximum accuracy (59.65%) was obtained with the kNN models with k=4, k=5 and k=8, and the minimum accuracy (56.14%) with the LDA model and the kNN with k=12. The highest f1-score for the "without evidence" class (58.82%) was reached with the kNN models with k=4 and k=12, and the lowest value (43.75%) with the LDA model. The maximum f1-score for the "mild evidence" class (66.67%) was achieved with the kNN model with k=8, and the minimum value (62.75%) with the kNN model with k=6. The highest f1-score for the "moderate to severe evidence" class (53.85%) was reached with the kNN model with k=5, and the lowest value (30.00%) with the kNN models with k=12.

As can be seen from the figure 5.30, the accuracy values obtained were between 56.14% and 59.65%, with the lowest value being with the LDA model and the kNN model when k=12, and the highest with the kNN model when k=4, k=5 and k=8.

Concerning the f1-score values, the values of the class "without evidence" were between 43.75% and 58.82%, those of the class "with mild evidence" were between 62.75% and 66.67% and the f1-score values of the class "with moderate to severe evidence" were between 30.00% and 53.85%.

The LDA model had the lowest f1-score value of the class "without evidence", and the kNN model when k=4 and k=12 had the highest values of this metric. The highest f1-score value of the class "moderate to severe evidence" was obtained with the kNN model when k=5 and lowest value when k=12. Regarding the f1-score of the class "mild evidence", the highest value was obtained with the kNN model when k=8.

The f1-score values of the kNN models had smaller differences between the classes using the 10-fold cross-validation method than when using the train/test split validation method.

5.3.3.2. Binary Classification: Level 1 - Without or With Evidence of Postural Problems

After testing the multiclass classification, we evaluated the performance of a binary classification with two levels of classification, to understand if the accuracy and f1-score values would be higher. The improvement of these metrics means a lower probability of generating misclassified results. The first level of this binary classification consisted in classifying the 57 participants of the study population into without or with evidence of postural problems.

Train/Test Split Validation Method

In this step, we also started with the application of the train/test split validation method with the data divided in 70% for training and 30% for testing, for later we implement the LDA and kNN models.

For the kNN model, we analyzed the graph in the figure 5.31, which shows the variation of error rate when we vary the value of k. With this graph we have selected the following five values of k, which correspond to the lowest values of error rate: k=9, k=11, k=14, k=15 and k=17. The confusion matrix obtained with each one of these kNN algorithms are displayed in the figures: figure 5.32, figure 5.33, and figure 5.34. The figure 5.34 also shows the confusion matrix for the LDA model.

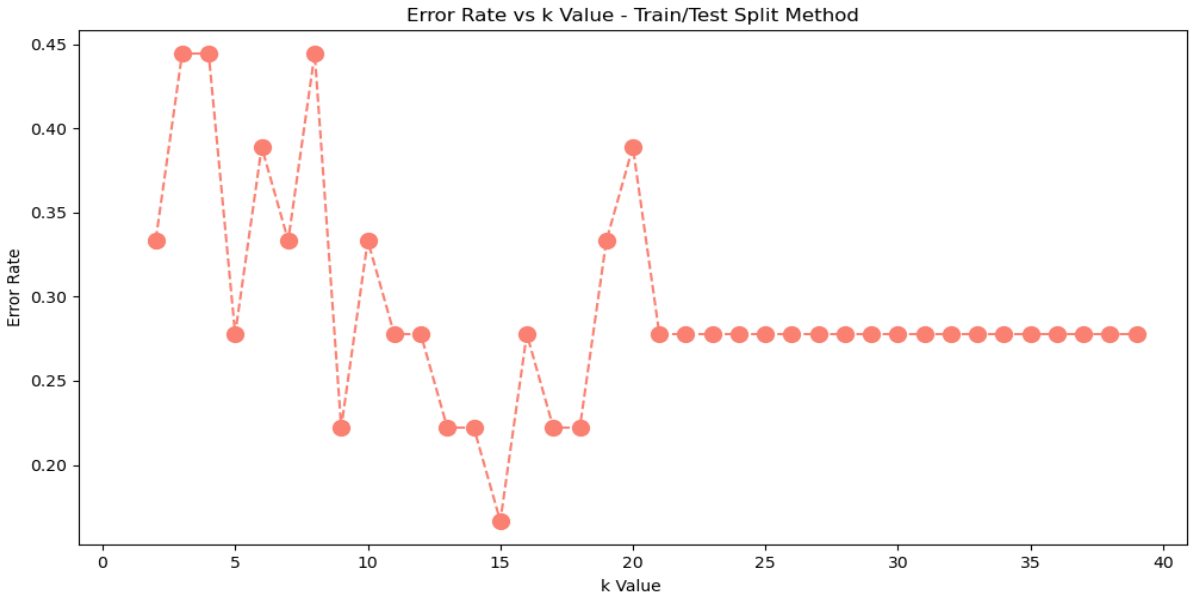


Figure 5.31: Graph of the error rate variation in function of k value when implementing the kNN algorithm with the train/test split validation method to classify the subjects into the two classes of the level 1 of the binary classification: "without evidence" and "with evidence". The five k values that correspond to the first five lowest error rates are k=9, k=13, k=14, k=15 and k=17.

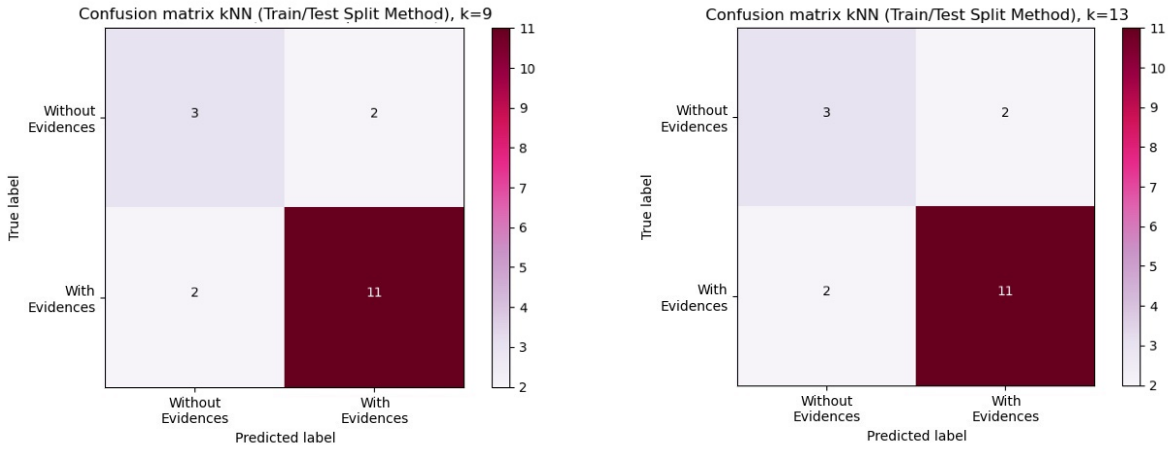


Figure 5.32: Confusion matrix for the two classes of the level 1 of the binary classification ("without evidence", "with evidence") when the kNN model was applied with the train/test split validation method. It shows the number of correct predictions for each of the three classes (3 for "without evidence", 11 for "with evidence") and the number of the wrong predictions (2 for "without evidence", 2 for "with evidence"). The kappa coefficient is 0.45. **On the left:** we used k=9. **On the right:** we used k=13.

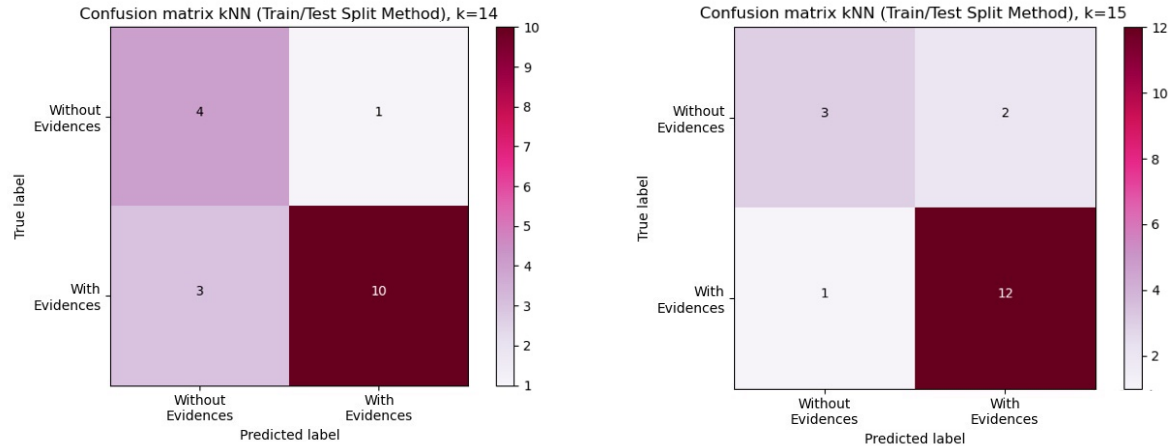


Figure 5.33: Confusion matrix for the two classes of the level 1 of the binary classification ("without evidence", "with evidence") when the kNN model with k=15 was applied with the train/test split validation method. **On the left:** we used k=14. It shows the number of correct predictions for each of the three classes (4 for "without evidence", 10 for "with evidence") and the number of the wrong predictions (1 for "without evidence", 3 for "with evidence"). The kappa coefficient is 0.51. **On the right:** we used k=15. It shows the number of correct predictions for each of the three classes (3 for "without evidence", 12 for "with evidence") and the number of the wrong predictions (2 for "without evidence", 1 for "with evidence"). The kappa coefficient is 0.56.

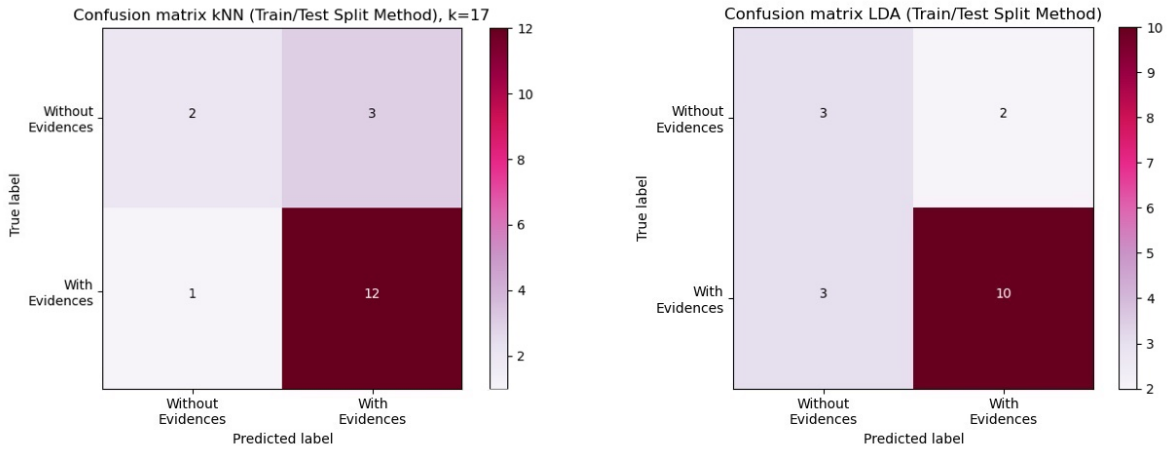


Figure 5.34: Confusion matrix for the two classes of the level 1 of the binary classification ("without evidence", "with evidence") when was applied the train/test split validation method. **On the left:** we used the kNN model with k=17. It shows the number of correct predictions for each of the three classes (2 for "without evidence", 12 for "with evidence") and the number of the wrong predictions (3 for "without evidence", 1 for "with evidence"). The kappa coefficient is 0.37. **On the right:** we used the LDA model. It shows the number of correct predictions for each of the three classes (3 for "without evidence", 10 for "with evidence") and the number of the wrong predictions (2 for "without evidence", 3 for "with evidence"). The kappa coefficient is 0.35

For each one of these 6 confusion matrices, we calculated the values of accuracy and the f1-score values of the two classes since both classes are important in our study. The results of these metrics are displayed in the chart of the figure 5.35, the accuracy in orange, the f1-score of the class "with evidence" in green and the class "without evidence" in blue.

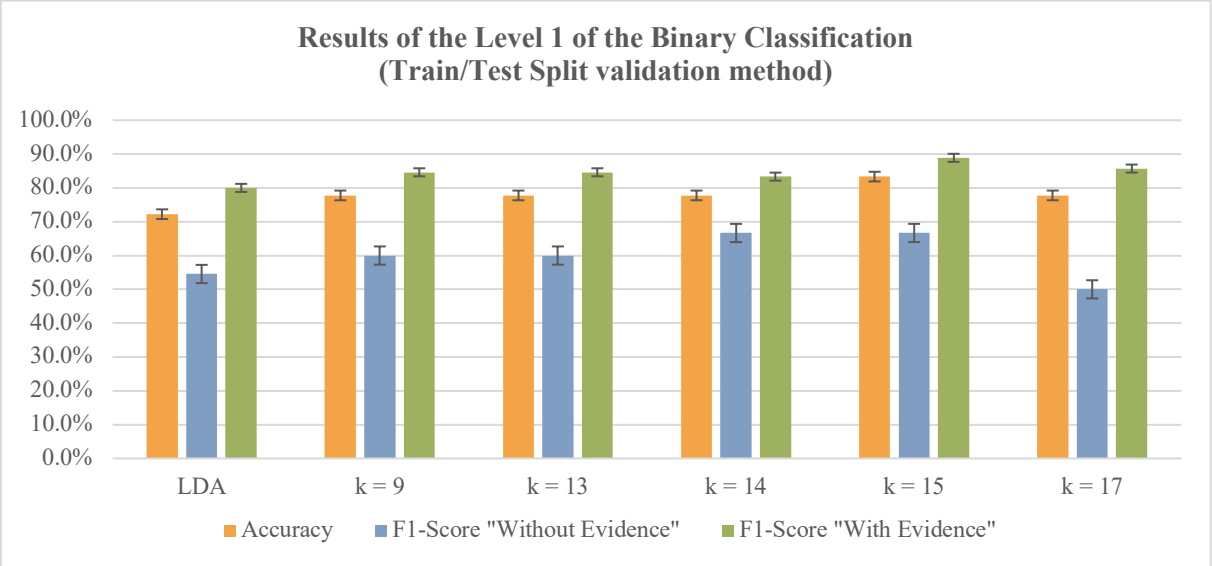


Figure 5.35: Results of the accuracy and f1-scores of each class obtained after the level 1 of the binary classification of all subjects with the LDA model and the five selected kNN models, when the train/test split validation method was applied. The maximum accuracy (83.33%) was obtained with the kNN model with k=15, and the minimum accuracy (72.22%) with the LDA model. The highest f1-score for the "without evidence" class (66.67%) was achieved with the kNN models with k=14 and k=15, and the lowest value (50.00%) with the kNN model when k=17. The maximum f1-score for the "with evidence" class (88.89%) was reached with the kNN model with k=15, and the minimum value (80.00%) with the LDA model.

According to the chart of the figure 5.35, the accuracy values are between 72.22% and 83.33%, and the highest accuracy value was obtained with the kNN model when k=15. The values of f1-score of the class "with evidence" varied between 80.00% and 88.89%, and the values of f1-score of the class

"without evidence" were between 50.00% and 66.67%. The model which had the best values for each metric of f1-score was also the kNN model when k=15. The kNN model when k=14, had as well the highest value of f1-score of the class "without evidence", however, the values of accuracy and f1-score of the class "with evidence" were lower.

The LDA model had lowest values of accuracy and f1-score of the class "with evidence", and second lowest value of f1-score of the class "without evidence".

Cross-Validation Method

Afterwards, we applied the 10-fold cross-validation method and repeated the previous procedures of implementation of the LDA algorithm and the kNN algorithm. With the graph of the figure 5.36, we observed that the five k values which corresponded to the highest accuracy values were k=8, k=9, k=10, k=17 and k=21. The confusion matrix obtained with the LDA algorithm is represented in the figure 5.37, and the confusion matrices resulting from the kNN algorithm for each k value selected are presented in figure 5.37, figure 5.38 and figure 5.39.

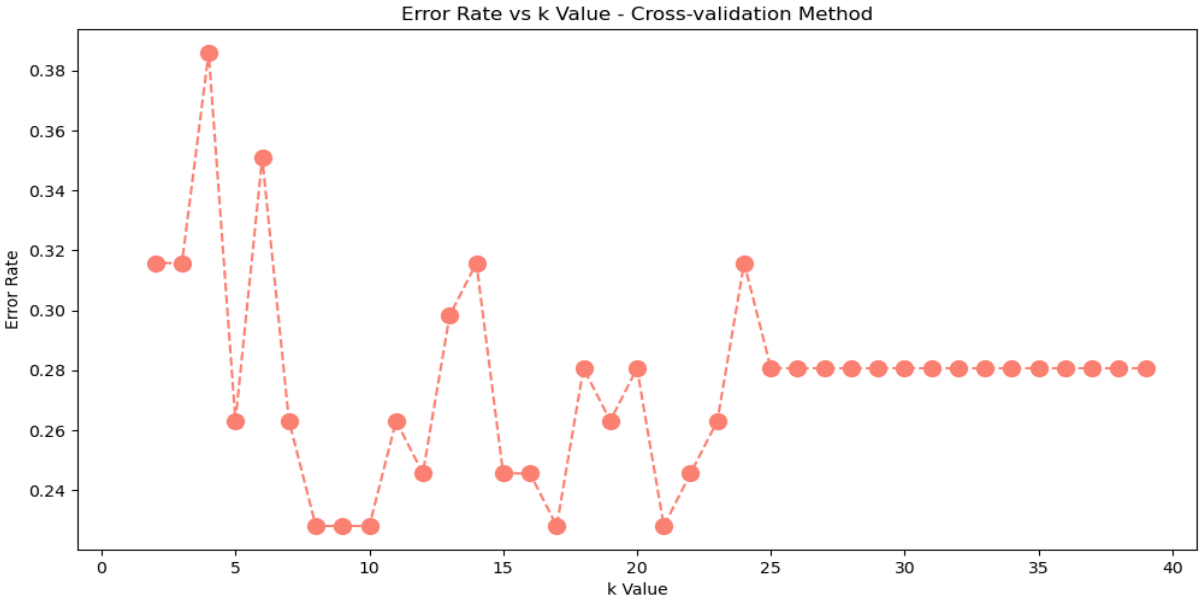


Figure 5.36: Graph of the error rate variation in function of k value when implementing the kNN algorithm with the cross-validation method to classify the subjects into the two classes of the level 1 of the binary classification: "without evidence" and "with evidence". The five k values that correspond to the first five lowest error rates are k=8 k=9, k=10, k=17 and k=21.

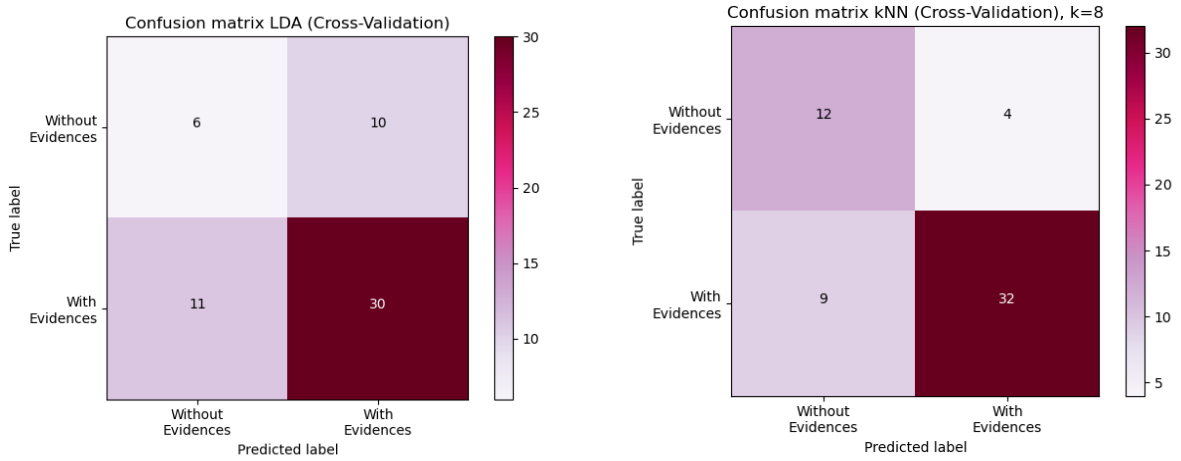


Figure 5.37: Confusion matrix for the two classes of the level 1 of the binary classification ("without evidence", "with evidence") when was applied the cross-validation method. **On the left:** we used the LDA model. It shows the number of correct predictions for each of the three classes (6 for "without evidence", 30 for "with evidence") and the number of the wrong predictions (10 for "without evidence", 11 for "with evidence"). The kappa coefficient is 0.10. **On the left:** we used the kNN model with k=8. It shows the number of correct predictions for each of the three classes (12 for "without evidence", 32 for "with evidence") and the number of the wrong predictions (4 for "without evidence", 9 for "with evidence"). The kappa coefficient is 0.48.

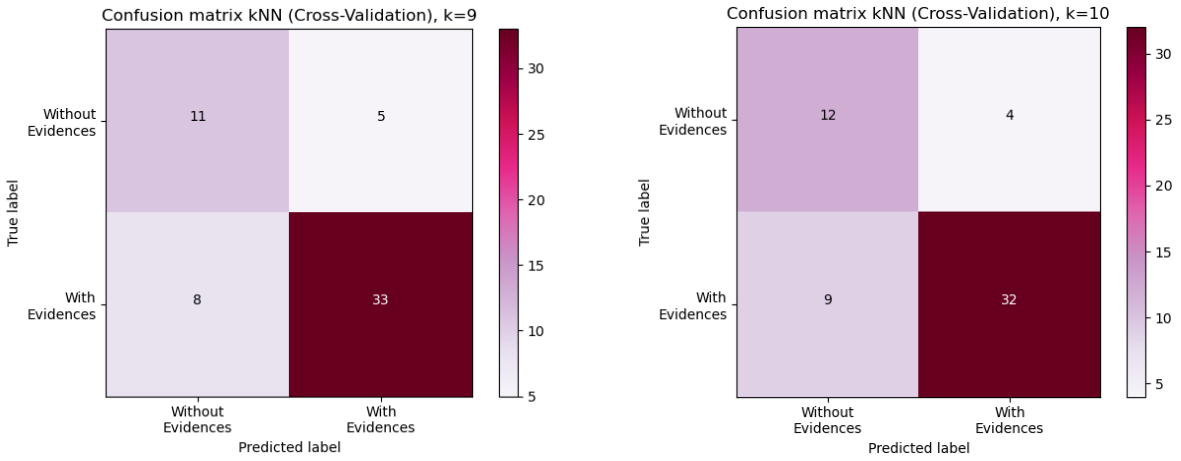


Figure 5.38: Confusion matrix for the two classes of the level 1 of the binary classification ("without evidence", "with evidence") when the kNN model was applied with the cross-validation method. **On the left:** we used k=9. It shows the number of correct predictions for each of the three classes (11 for "without evidence", 33 for "with evidence") and the number of the wrong predictions (5 for "without evidence", 8 for "with evidence"). The kappa coefficient is 0.47. **On the right:** we used k=10. It shows the number of correct predictions for each of the three classes (12 for "without evidence", 32 for "with evidence") and the number of the wrong predictions (4 for "without evidence", 9 for "with evidence"). The kappa coefficient is 0.48.

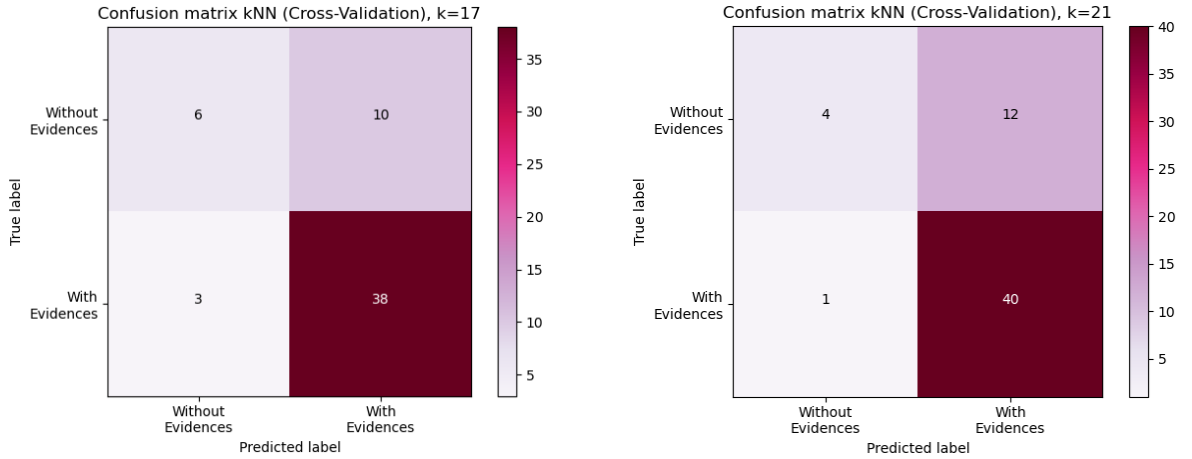


Figure 5.39: Confusion matrix for the two classes of the level 1 of the binary classification ("without evidence", "with evidence") when the kNN model was applied with the cross-validation method. **On the left:** we used k=17. It shows the number of correct predictions for each of the three classes (6 for "without evidence", 38 for "with evidence") and the number of the wrong predictions (10 for "without evidence", 3 for "with evidence"). The kappa coefficient is 0.35. **On the right:** we used k=21. It shows the number of correct predictions for each of the three classes (4 for "without evidence", 40 for "with evidence") and the number of the wrong predictions (12 for "without evidence", 1 for "with evidence"). The kappa coefficient is 0.29.

The LDA model presented the lowest values of the three metrics evaluated in the chart of the figure 5.40. It had the lowest values of accuracy (63.16%), of f1-score of the class "with evidence" (74.07%) and of f1-score of the class "without evidence" (36.36%).

Regarding the five kNN models tested, all the five kNN models had the same accuracy value of 77.19%. The f1-score values of the class "with evidence" ranged between 83.12% and 85.39%, and the f1-score values of the class "without evidence" were between 38.10% and 64.86%. The best f1-score values of the class "without evidence" (64.86%) were obtained when implementing the kNN models with k=8 and k=10. The model that obtained the best f1-score value of the class "with evidence" was the kNN model with k=17 (85.39%), however the f1-score value of the class "without evidence" for this model was only 48.00%. Meanwhile, the kNN models with k=8 and k=10, despite not having the highest f1-score value of the class "with evidence", had a value of 83.12%, in addition to having the best f1-score value of the class "without evidence". These two models had similar values in the three metrics presented in the figure 5.40, as well, both had a kappa coefficient of 0.48.

After evaluating the models of each validation method, we could compare the models that presented the highest results in each method. For the train/test split validation method was the kNN model with k=15, and for the 10-fold cross-validation method it was the kNN model with k=17. We can observe that the first model had higher values in the three metrics analyzed. Approximately, this model had 6% more in the values of accuracy and f1-score of the class "with evidence", and 2% more in the f1-score of the class "without evidence". Although the use of the 10-fold cross-validation method usually makes the classifier more robust, these results can be explained by the small size of the database. Since, with 57 participants, each fold is composed of approximately 6 participants, which can mean a bad distribution of the data.

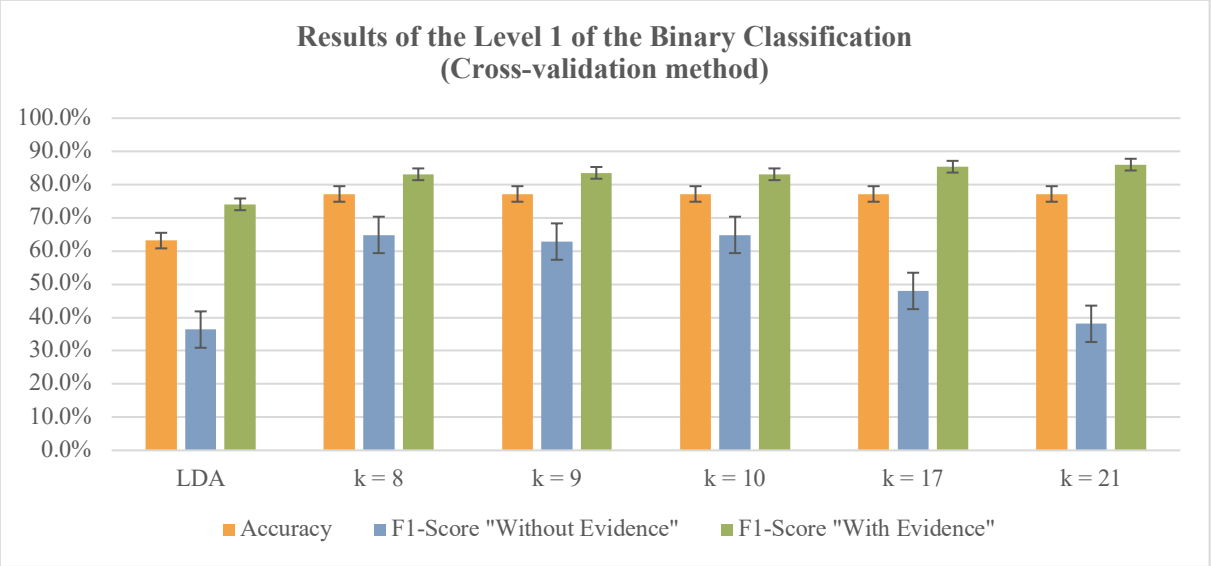


Figure 5.40: Results of the accuracy and f1-scores of each class obtained after the level 1 of the binary classification of all subjects with the LDA model and the five selected kNN models, when the cross-validation method was applied. The maximum accuracy (77.19%) was obtained with all the kNN models, and the minimum accuracy (63.16%) with the LDA model. The highest f1-score for the "without evidence" class (64.86%) was achieved with the kNN models with k=8 and k=10, and the lowest value (36.36%) with the LDA model. The maximum f1-score for the "with evidence" class (86.02%) was reached with the kNN model when k=21, and the minimum value (74.07%) with the LDA model.

5.3.3.3. Binary Classification: Level 2 - Mild or Moderate/Severe Evidence of Postural Problems

Thereafter, we evaluate only the 41 participants that were previously classified as having evidence of postural problems. These subjects were then classified as having mild evidence or as having moderate to severe evidence of postural problems. In this phase, the data was also analyzed with the LDA and kNN models, first using the train/test split validation method and next the 10-fold cross-validation.

Train/Test Split Validation Method

The train/test split method was also applied with 70% of data as train and 30% of data as test. After the data split, we implement the LDA algorithm, and obtain the confusion matrix shown in figure 5.41.

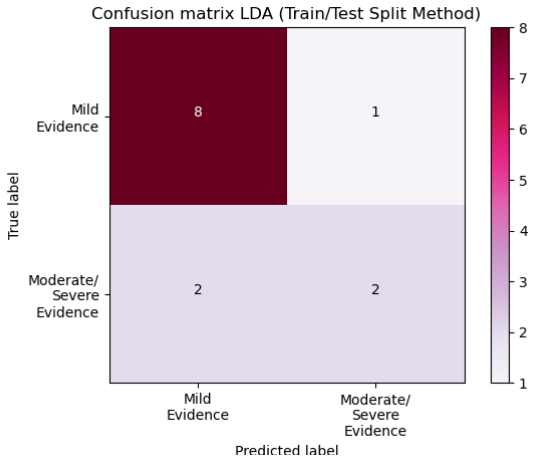


Figure 5.41: Confusion matrix for the two classes of the level 2 of the binary classification ("with mild evidence", "with moderate to severe evidence") when the LDA model was applied with the train/test split validation method. It shows the number of correct predictions for each of the three classes (8 for "with mild evidence", 2 for "with moderate to severe evidence") and the number of the wrong predictions (1 for "with mild evidence", 2 for "with moderate to severe evidence"). The kappa coefficient is 0.42.

For the implementation of the kNN model, we also plotted the accuracy in function of k value to select the k values which corresponded to the first five highest accuracy values. In this level, we range the k value between 2 and 28, since we considered less participants for this step. As can be observed in the figure 5.42, the best values of accuracy were obtained at k = 6, k = 7, k = 9, k = 10 and k = 11.

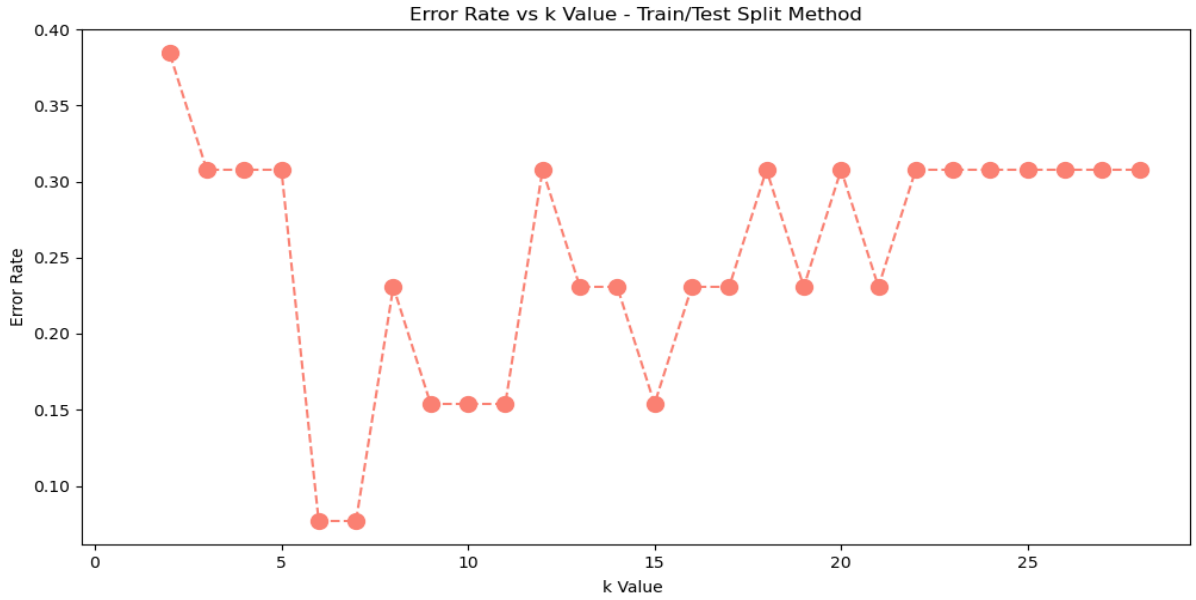


Figure 5.42: Graph of the error rate variation in function of k value when implementing the kNN algorithm with the train/test split validation method to classify the subjects previously defined as having evidence into the two classes of the level 2 of the binary classification: “with mild evidence” and “with moderate to severe evidence”. The five k values that correspond to the first five lowest error rates are k=6, k=7, k=9, k=10 and k=11.

Afterwards, with the five k values selected, we ran the kNN algorithm for each k value and plot the confusion matrix for k=6, k=7 (Figure 5.43), k=9, k=10 (Figure 5.44) and k=11 (Figure 5.45).

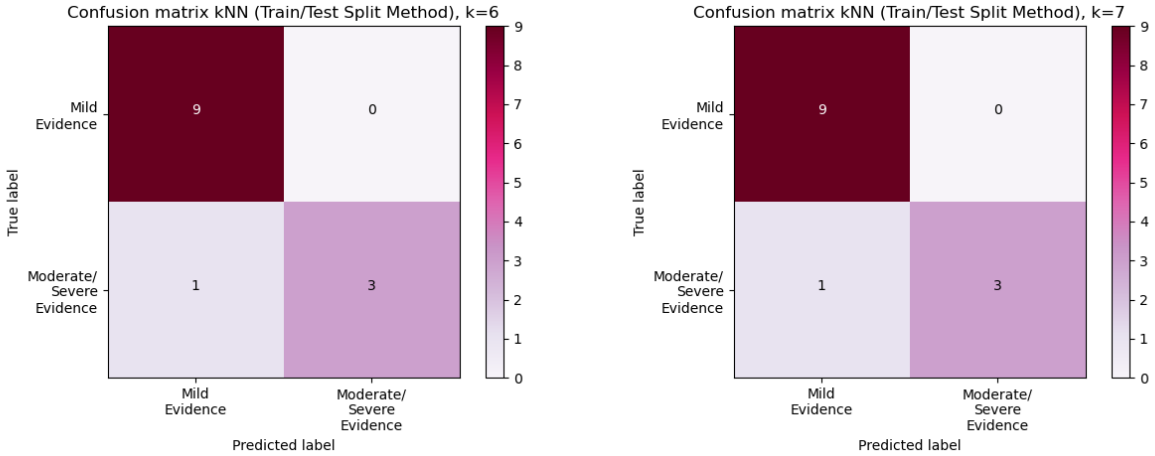


Figure 5.43: Confusion matrix for the two classes of the level 2 of the binary classification (“with mild evidence”, “with moderate to severe evidence”) when the kNN model was applied with the train/test split validation method. It shows the number of correct predictions for each of the three classes (9 for “with mild evidence”, 3 for “with moderate to severe evidence”) and the number of the wrong predictions (0 for “with mild evidence”, 1 for “with moderate to severe evidence”). The kappa coefficient is 0.81. **On the left:** we used k=6. **On the right:** we used k=7.

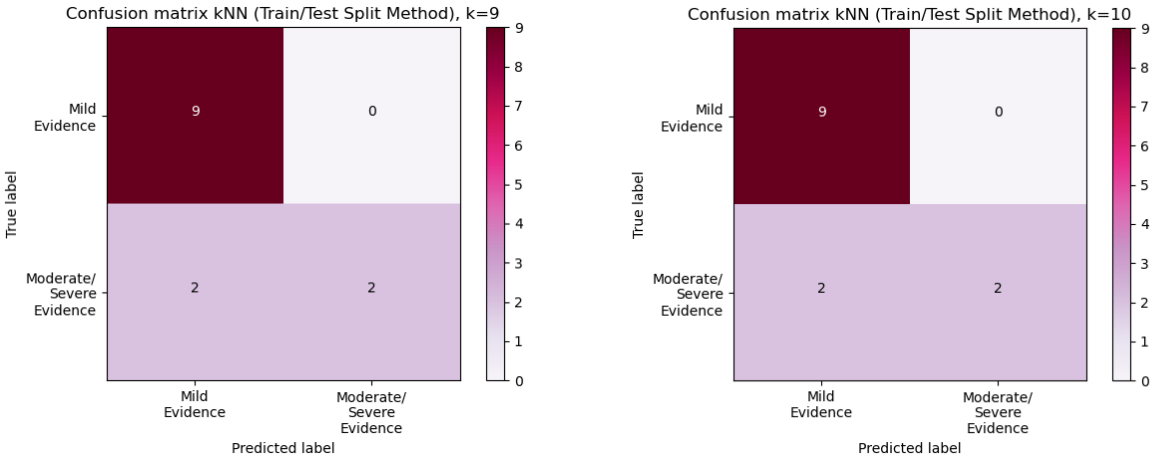


Figure 5.44: Confusion matrix for the two classes of the level 2 of the binary classification ("with mild evidence", "with moderate to severe evidence") when the kNN model was applied with the train/test split validation method. It shows the number of correct predictions for each of the three classes (9 for "with mild evidence", 2 for "with moderate to severe evidence") and the number of the wrong predictions (0 for "with mild evidence", 2 for "with moderate to severe evidence"). The kappa coefficient is 0.58. **On the left:** we used k=9. **On the right:** we used k=10.

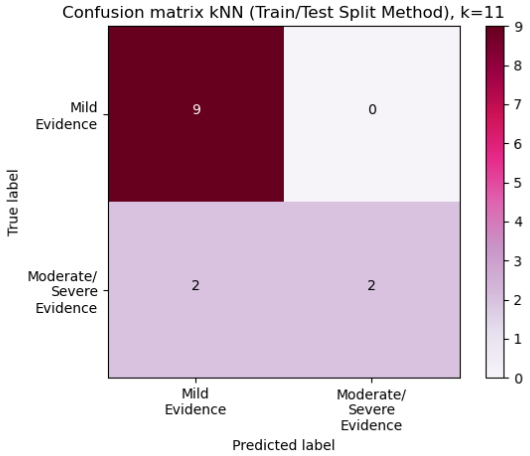


Figure 5.45: Confusion matrix for the two classes of the level 2 of the binary classification ("with mild evidence", "with moderate to severe evidence") when the kNN model with k=11 was applied with the train/test split validation method. It shows the number of correct predictions for each of the three classes (9 for "with mild evidence", 2 for "with moderate to severe evidence") and the number of the wrong predictions (0 for "with mild evidence", 2 for "with moderate to severe evidence"). The kappa coefficient is 0.58.

The calculated results of accuracy and f1-score of each class for the LDA model and for the kNN models with each k value are presented in the chart of figure 5.46, for a better comparison of the values. In the figure 5.46, we can verify that the LDA had lowest values of accuracy (76.92%), of f1-score of the class "with mild evidence" (84.21%) and of f1-score of the class "with moderate to severe evidence" (57.14%).

In opposition, the kNN models with k=6 and k=7 obtained the highest values in three metrics evaluated. They had an accuracy of 92.31%, a f1-score of the class "with mild evidence" of 94.74% and a f1-score of the class "with moderate to severe evidence" of 85.71%. The kappa coefficient of the confusion matrices for both models was 0.81.

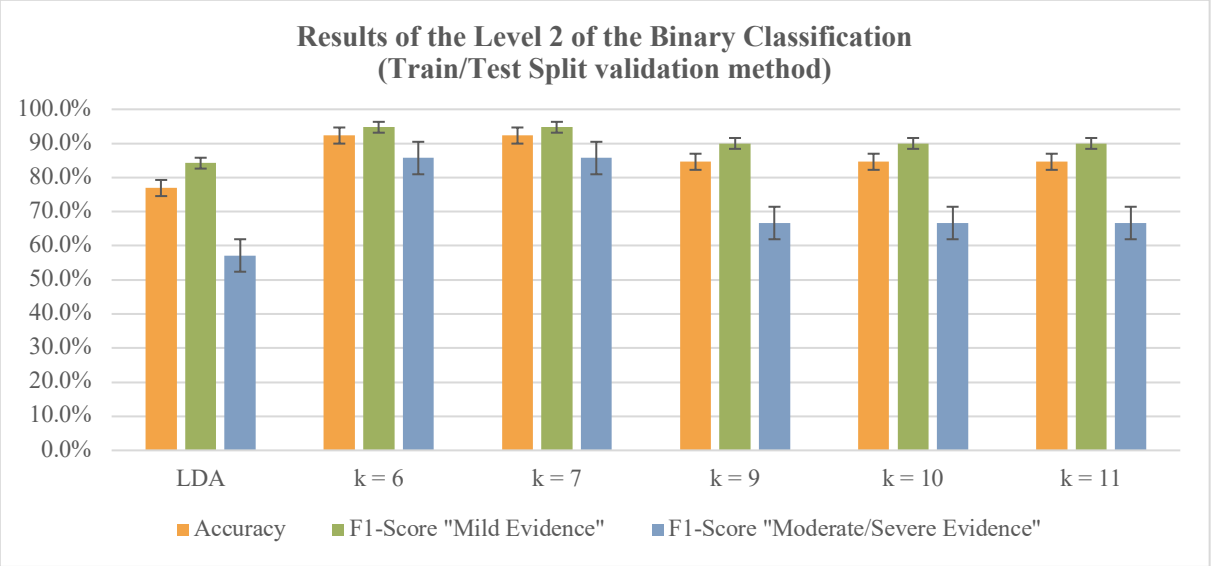


Figure 5.46: Results of the accuracy and f1-scores of each class obtained after the level 2 of the binary classification of the 41 subjects defined as having evidence, with the LDA model and the five selected kNN models when the train/test split validation method was applied. The maximum accuracy (92.31%) was obtained with the kNN models with k=6 and k=7, and the minimum accuracy (76.92%) with the LDA model. The highest f1-score for the "with mild evidence" class (94.74%) was achieved with the kNN models with k=6 and k=7, and the lowest value (84.21%) with the LDA model. The maximum f1-score for the "with moderate to severe evidence" class (85.71%) was reached with the kNN models with k=6 and k=7, and the minimum value (57.14%) with the LDA model.

Cross-Validation Method

With the application of the 10-fold cross-validation method, for the implementation of the kNN algorithm, we also plotted the error rate as a function of the k value, which we can see in figure 5.47. From this graph, we can see which k values had the lowest error rates: k=3, k=4, k=5, k=6 and k=9.

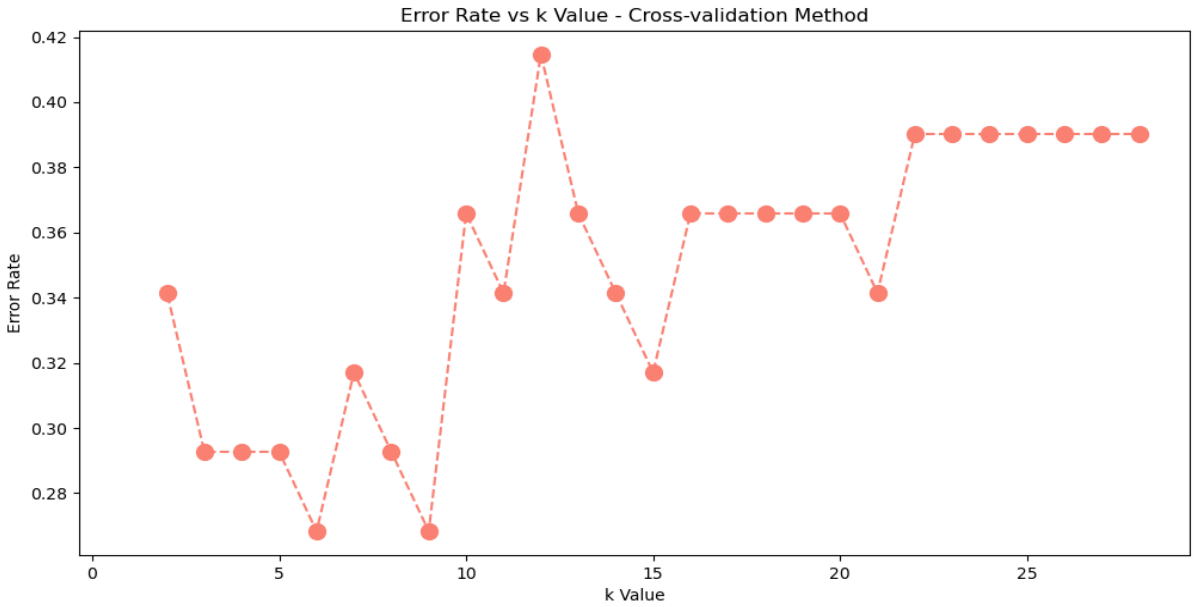


Figure 5.47: Graph of the error rate variation in function of k value when implementing the kNN algorithm with the cross-validation method to classify the subjects previously define as having evidence into the two classes of the level 2 of the binary classification: "with mild evidence" and "with moderate to severe evidence". The five k values that correspond to the first five lowest error rates are k=3, k=4, k=5, k=6 and k=9.

The confusion matrices obtained with these kNN models are represented in figure 5.48 for k=3 and k=4, in figure 5.49 for k=5 and k=6, and in figure 5.50 for k=9. The figure 5.50 also shows the confusion matrix that resulted from the LDA algorithm.

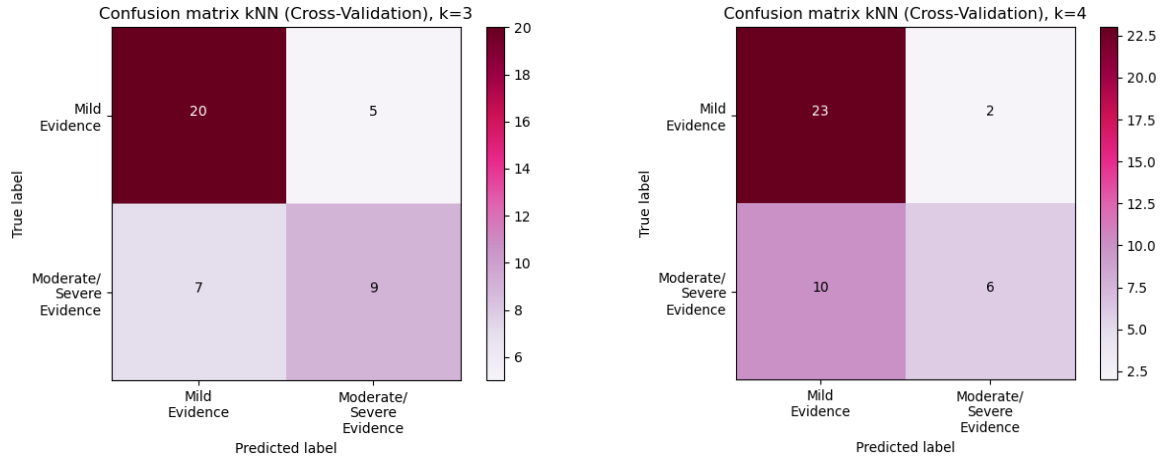


Figure 5.48: Confusion matrix for the two classes of the level 2 of the binary classification ("with mild evidence", "with moderate to severe evidence") when the kNN model was applied with the cross-validation method. **On the left:** we used k=3. It shows the number of correct predictions for each of the three classes (20 for "with mild evidence", 9 for "with moderate to severe evidence") and the number of the wrong predictions (5 for "with mild evidence", 7 for "with moderate to severe evidence"). The kappa coefficient is 0.36. **On the right:** we used k=4. It shows the number of correct predictions for each of the three classes (23 for "with mild evidence", 6 for "with moderate to severe evidence") and the number of the wrong predictions (2 for "with mild evidence", 10 for "with moderate to severe evidence"). The kappa coefficient is 0.32.

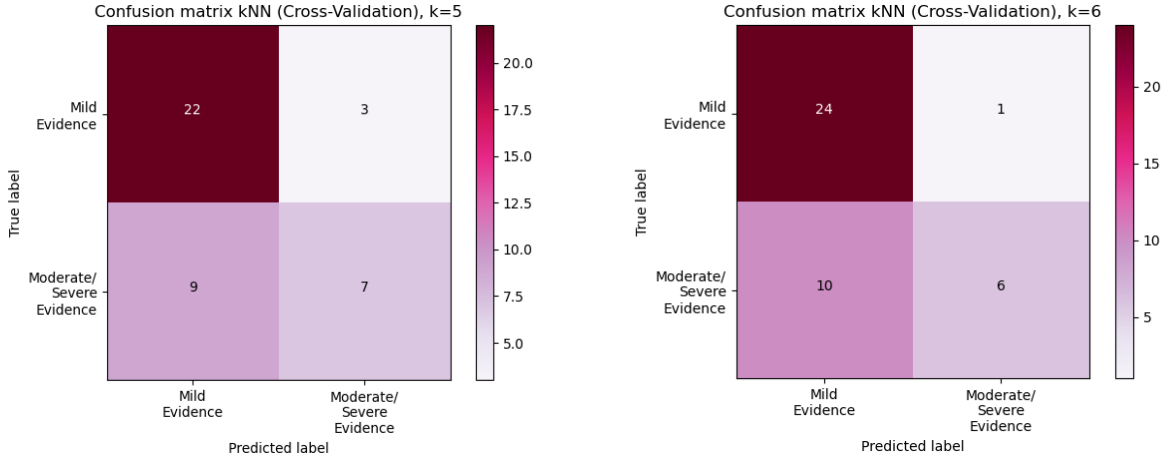


Figure 5.49: Confusion matrix for the two classes of the level 2 of the binary classification ("with mild evidence", "with moderate to severe evidence") when the kNN model was applied with the cross-validation method. **On the left:** we used k=5. It shows the number of correct predictions for each of the three classes (22 for "with mild evidence", 7 for "with moderate to severe evidence") and the number of the wrong predictions (3 for "with mild evidence", 9 for "with moderate to severe evidence"). The kappa coefficient is 0.34. **On the right:** we used k=6. It shows the number of correct predictions for each of the three classes (24 for "with mild evidence", 6 for "with moderate to severe evidence") and the number of the wrong predictions (10 for "with mild evidence", 1 for "with moderate to severe evidence"). The kappa coefficient is 0.37.

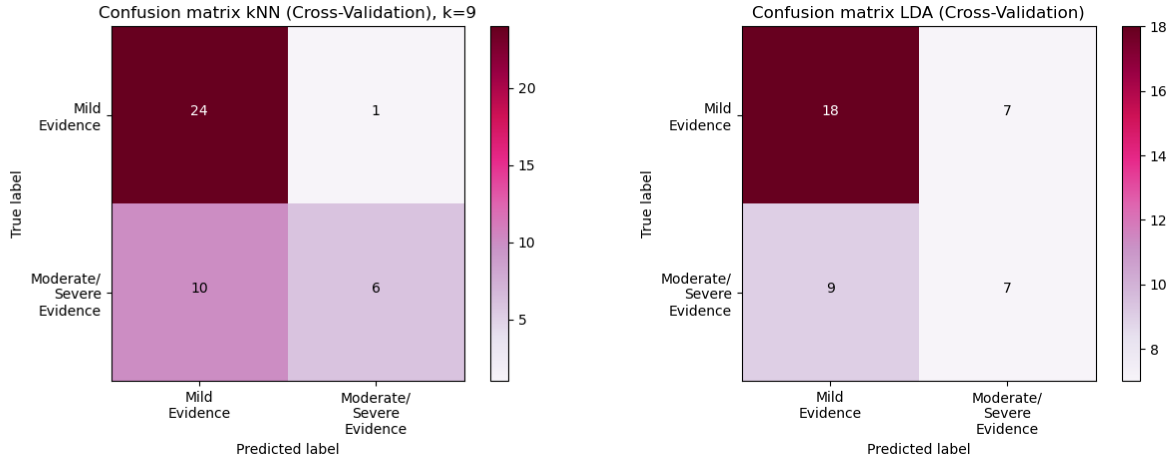


Figure 5.50: Confusion matrix for the two classes of the level 2 of the binary classification ("with mild evidence", "with moderate to severe evidence") when was applied with the cross-validation method. **On the left:** we used the kNN model with k=9. It shows the number of correct predictions for each of the three classes (24 for "with mild evidence", 6 for "with moderate to severe evidence") and the number of the wrong predictions (10 for "with mild evidence", 1 for "with moderate to severe evidence"). The kappa coefficient is 0.37. **On the right:** we used the LDA model. It shows the number of correct predictions for each of the three classes (18 for "with mild evidence", 7 for "with moderate to severe evidence") and the number of the wrong predictions (7 for "with mild evidence", 9 for "with moderate to severe evidence"). The kappa coefficient is 0.16.

The chart in figure 5.51 shows that we acquired the best accuracy while using the kNN algorithm than with the LDA algorithm. As it happened, when we used the previous validation method, also with the 10-fold cross validation method, the LDA model had the lowest values for the three metrics analyzed in the chart of figure 5.51. The running of the LDA algorithm gave an accuracy of 60.98%, a f1-score value of the class "with mild evidence" of 69.23% and a f1-score value of the class "with moderate to severe evidence" of 46.67%.

As for the results obtained with the five kNN models tested, the accuracy values were 70.73% when k=3, k=4 and k=5, and were 73.17% when k=6 and k=9. The f1-score values of the class "with mild evidence" range between 76.92% and 81.36%, and the f1-score values of the class "with moderate to severe evidence" were between 50.00% and 60.00%.

Furthermore, we can observe that the kNN algorithms with k=6 and k=9 also gave the best f1-score value of the class "with mild evidence" (81.36%), although the f1-score value of the class "with moderate to severe evidence" when using these k-values were 52.17%. The best f1-score value of the class "with moderate to severe evidence" obtained occurred when k = 3, however of the five kNN model, this model presented the lowest f1-score value of the class "with mild evidence".

Nevertheless, when calculating the kappa coefficient of the confusion matrices of the confusion matrices, the kNN model with k=6 and k=9 had the highest value of 0.37.

After analyzing which classification models had the best overall results in each one of the validation methods used in this level 2 of the binary classification, we can compare their results. For the train/test split validation method were the kNN models with k=6 and k=7, and for the 10-fold cross validation method were the kNN model with k=6 and k=9. Once again, similarly to what happened in level 1 of the binary classification, the results obtained with the train/test split validation method were higher. Furthermore, the differences between the results of the three metrics using each one of the validation methods were much greater at this level 2 of the binary classification. This situation can be explained once more due to the small size of the database, which at this level is even smaller. A quantity of 41 subjects means that each fold of the 10-fold cross-validation method has approximately 4 participants.

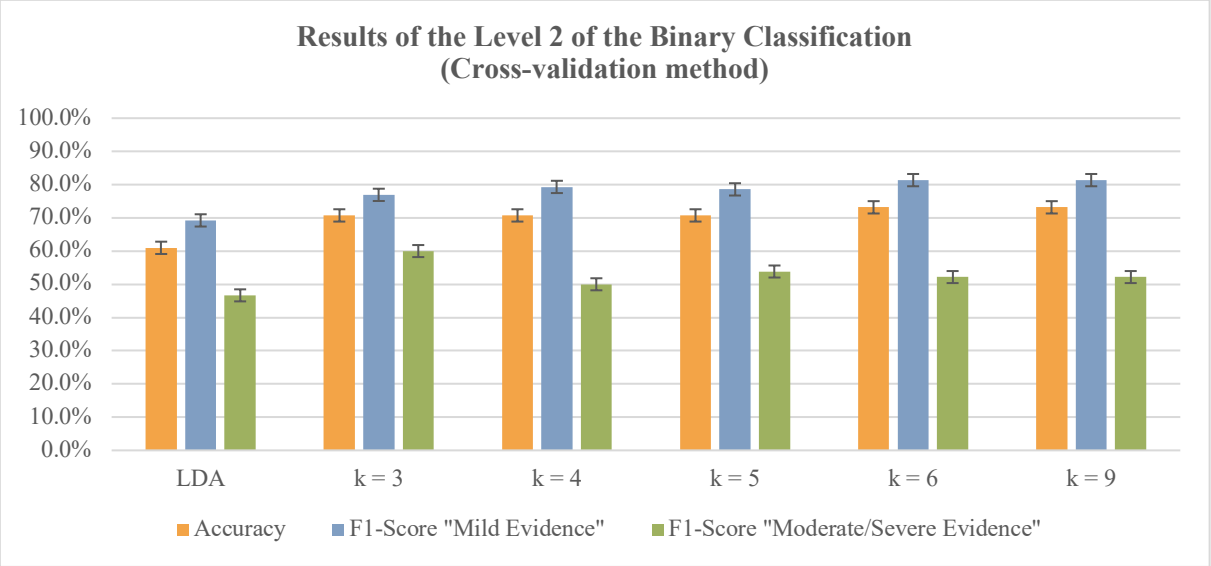


Figure 5.51: Results of the accuracy and f1-scores of each class obtained after the level 2 of the binary classification of the 41 subjects defined as having evidence, with the LDA model and the five selected kNN models when the cross-validation method was applied. The maximum accuracy (73.17%) was obtained with the kNN models with k=6 and k=9, and the minimum accuracy (60.98%) with the LDA model. The highest f1-score for the "with mild evidence" class (81.36%) was reached with the kNN models with k=6 and k=9, and the lowest value (69.23%) with the LDA model. The maximum f1-score for the "with moderate to severe evidence" class (60.00%) was achieved with the kNN model when k=3, and the minimum value (46.67%) with the LDA model.

Comparing the results obtained with the multiclass classification and with the binary classification with 2 levels, we concluded that the binary classification had in general better values and presented less confusion between classes than the multiclass classification.

Accuracy measures the number of all the correctly classified observations. This value is considered good the higher it is. Though, it is difficult to quantify a general value from which we can consider to be a good value, as it will depend on the type of data used and the more or less serious implications of an incorrect classification. Nonetheless, in general, an accuracy greater than 80% can be considered a good value. In the case of imbalance data, in which we want to understand how it is the model performance for each class, it is also important to analyze the f1-score values of each class. The f1-score of each class measures a harmonic mean between the correctly identified cases of all predicted cases and the correctly identified cases of all the actual cases. This value is also better the higher it is. Regarding the classification models with the best accuracy and f1-score values for each of the two levels of the binary classification, we verified that in cases the accuracy was higher than 80%. In the level 1, the kNN model with k=15 with the split train/test validation method had an accuracy of 83.33%. In the level 2, the kNN models with k=6 and k=7 when using the split train/test validation method had both an accuracy of 92.31%.

A summary of the accuracy and f1-scores results obtained when using the LDA and the kNN models with both the 10-fold cross-validation and the train/test split validation methods in the multiclass classification and in the level 1 and level 2 of the binary classification is presented in the table 5.12, below.

Table 5.12: Results obtained from the analysis of the classifiers performance. It shows the accuracy and f1-score values of each class from the multiclass classification and from the level 1 and level 2 of the binary classification.

	Multiclass Classification											
	Train/Test Split validation						**Cross-validation**					
	LDA	kNN, k = 3	kNN, k = 4	kNN, k = 5	kNN, k = 6	kNN, k = 9	LDA	kNN, k = 4	kNN, k = 5	kNN, k = 6	kNN, k = 8	kNN, k = 12
Accuracy	50.00%	55.56%	55.56%	55.56%	55.56%	61.11%	56.14%	59.65%	59.65%	57.89%	59.65%	56.14%
F1-Score "Without Evidence"	36.36%	53.33%	53.33%	53.33%	53.33%	66.67%	43.75%	58.82%	57.89%	57.89%	57.89%	58.82%
F1-Score "Mild Evidence"	57.14%	76.92%	76.92%	76.92%	76.92%	70.59%	65.38%	65.38%	64.00%	62.75%	66.67%	63.33%
F1-Score "Moderate/ Severe Evidence"	54.55%	25.00%	25.00%	25.00%	25.00%	28.57%	53.33%	50.00%	53.85%	48.00%	48.00%	30.00%
	Binary Classification - Level 1											
	Train/Test Split validation						**Cross-validation**					
	LDA	kNN, k = 9	kNN, k = 13	kNN, k = 14	kNN, k = 15	kNN, k = 17	LDA	kNN, k = 8	kNN, k = 9	kNN, k = 10	kNN, k = 17	kNN, k = 21
Accuracy	72.22%	77.78%	77.78%	77.78%	83.33%	77.78%	63.16%	77.19%	77.19%	77.19%	77.19%	77.19%
F1-Score "With Evidence"	80.00%	84.62%	84.62%	83.33%	88.89%	85.71%	74.07%	83.12%	83.54%	83.12%	85.39%	86.02%
F1-Score "Without Evidence"	54.55%	60.00%	60.00%	66.67%	66.67%	50.00%	36.36%	64.86%	62.86%	64.86%	48.00%	38.10%
	Binary Classification - Level 2											
	Train/Test Split validation						**Cross-validation**					
	LDA	kNN, k = 6	kNN, k = 7	kNN, k = 9	kNN, k = 10	kNN, k = 11	LDA	kNN, k = 3	kNN, k = 4	kNN, k = 5	kNN, k = 6	kNN, k = 9
Accuracy	76.92%	92.31%	92.31%	84.62%	84.62%	84.62%	60.98%	70.73%	70.73%	70.73%	73.17%	73.17%
F1-Score "Mild Evidence"	84.21%	94.74%	94.74%	90.00%	90.00%	90.00%	69.23%	76.92%	79.31%	78.57%	81.36%	81.36%
F1-Score "Moderate/ Severe Evidence"	57.14%	85.71%	85.71%	66.67%	66.67%	66.67%	46.67%	60.00%	50.00%	53.85%	52.17%	52.17%

6. Conclusion and Future Perspectives

The musculoskeletal pain and disorders, such as the LBP, NP, and scoliosis, are common in adulthood, especially the LBP. These disturbances affect, on a daily basis, the quality of life of populations in industrialized countries, namely Portugal, and can have consequences on people's professional lives. Additionally, the incidence of these conditions has associated costs due to its treatments, as well as costs due reduction of productivity and absence from work. These last costs tend to be higher than the first ones and are in the hundreds of millions of euros.

Although these problems have a higher incidence in adult life, with early detection during childhood and adolescence, it is possible to reduce the likelihood of LBP, NP and scoliosis in adulthood or they become chronic. During childhood and adolescence, the human being begins to be exposed to stimulus and risk factors that can lead to the development of pain and musculoskeletal disorders. Some of these risks are the being female, the weight and type of the bag used in schools, the intensity and frequency of exercise activities, and the posture adopted during the day, for example having an incorrect spine position when using the mobile phone or when sitting watching television.

Currently, there are several postural analysis software and multidisciplinary pain treatment programs for use by health professionals in a clinical environment, however these are also expensive. Additionally, mHealth apps with an evidence-based multidisciplinary pain treatment are starting to appear for the common user who suffers from low back and neck pain, for which they pay a small fee. These apps allow the user to monitor their posture and learn techniques and exercises to reduce their pain.

However, an early detection of musculoskeletal disorders can decrease the probability of these disorders in adulthood or of progressing to a chronic state, as it allows children and adolescents to obtain help by a health professional.

This dissertation aims to continue the development of a software that was previously started by some colleagues, which will be in the future a mobile phone operable software. The software processes images of different views of the subject’s posture and calculates several angles and distances between body segments using specific anatomical landmarks, to later give a score of the risk indicators for possible problems or disorders.

In this sense, we divided our work in four phases. First, we made improvements to the software that had already been created. We selected which anatomical sites were going to be used as markers to calculate the anatomical metrics that we wanted the software to evaluate. One of the objectives of this dissertation was to improve the software previously created by colleagues of the Master Program in Biomedical Engineering and Biophysics at FCUL. This objective was achieved since we were able to make possible an automatic analysis of images from different cameras and with different scales. The characteristics of the markers used were also modified, increasing the success rate of the software in detecting the different markers.

Second, we assessed the performance of the software. For that, we tested the use of markers with different colors and sizes, in subjects of different heights and volumes, who wore clothes of different colors, to understand which one the software had the least problems detecting. We also evaluated the software’s ability to analyze images with different backgrounds colors and with the use or not of the flash, captured by two different types of cameras. After assessing the best conditions for image acquisition, we analyzed the influence of diverse variables in the results of the calculation of the features. These variables were the use or not of flash, the variation in the distance from the camera to the wall where the subject is, the variation in the camera’s height and the use of different cameras. In this

performance assessment, we decided to choose green colored markers with a diameter of 20 mm as markers. As well, we selected the color blue for the rectangle used to obtain the scale of the image and the color red for the string utilized to know which is the vertical line to the ground. We also, verified that the software was able to detect the green markers and blue rectangle with all color of background tested. The red string was only not detected when the background was red. Moreover, the blue rectangle may have been detected when using the blue background because it was placed on a brown box and not placed directly on the blue wall. The calibration tests showed that even placing the blue rectangle used to obtain the image scale next to the dress form, the parallax errors inherent to the image capture are not fully corrected, because the dress form has volume causing the markers not to be exactly on the same plane as the blue rectangle. These parallax errors were more pronounced in the calculations of the features that used the red string, since it is in a different plane than the dress form. The results of the images captured by the cell phone camera had less variation than the images captured by the photographic camera, and the use or not of flash had the smallest influence on the results obtained. One way to solve the problem of the parallax errors is to find the vanishing point of each image or use the size of the camera sensor, so we can have more exact results using different cameras and have the blue rectangle and red string in different planes of the markers. Furthermore, to have a more simplified setup, we could use a set square with a weight on one of the edges, replacing the blue rectangle and the red string, hence we would use only one object that allows us to obtain the image scale and the vertical to the floor.

Third, we acquired data from a study population composed of the students of the ESSCVP-Lisboa. We analyzed the same images obtained from each participant with our software and with the *Contemplas* software, which is a clinical software for posture and gait analysis. We used both software to calculate the same features in order to compare the differences between the results obtained with statistical hypothesis testing, so we could validate the results given by our software. Most of the features were not significantly different in the statistical tests of comparison of means and in the equivalence tests they presented as equivalent between the boundaries ΔL=-0.05 and ΔU = 0.05, with 95% of confidence. The parameters $ALLL_L$ and PSISHA were equivalent only from the boundaries ΔL=-0.06 and ΔU = 0.06 and were not significantly different in the comparison of means tests only after removing the outliers, with 95% of confidence. With these statistical analyses we were able to validate the results of our software and consider that these results were equivalent to the results obtained with the *Contemplas* software. The reduction of the parallax errors and the data variability during the calibration tests allows us to obtain values with lower uncertainties and reduce the boundaries of the mean equivalence with a confidence of 95%. Furthermore, we also asked to physiotherapy specialists to evaluate the posture of each participant and label them as: without evidence of postural problems; mild evidence of postural problems; and with moderate to severe evidence. Our study population was considered to be composed of 16 participants who did not present any evidence of posture problems, 25 presented mild evidence of posture problems and 16 presented moderate to severe evidence.

In this phase, we had to manually draw green circles on the markers that only appeared half a circle or less because they were covered by body parts, especially in the back area in the lateral view images. For future postural assessments, we need to correct this problem, which we can possibly solve by moving away the markers the same distance from the body while maintaining the angles between them. This solution would mainly be applied with the markers located on the back.

Finally, our fourth phase consisted of assessing the performance of classifiers with the labeled study population. We first performed a multiclass classification, and then a two level binary classification. For both, we implemented the LDA model and the kNN model, with a train/test split validation method and a 10-fold cross-validation. The multiclass classification had a lower accuracy

percentage than each one of the levels of the binary classification. In the binary classification, at level 1, the class “with evidence” achieved better values of f1-score than the class “without evidence”, and at level 2, the class “with mild evidence” obtained higher values of f1-score than the class “moderate to severe evidence”. These higher values can be explained due to both classes having a larger sample size, which allowed the use of more data labeled with these classes to train the classifier. Additionally, the use of the train/test split validation method gave better values of accuracy and f1-score of each class than the 10-fold cross-validation method. Although this second validation method is more robust, these can be justified by the size of the database being 57 participants. This means that each fold of the cross-validation method had only approximately 6 subjects which could lead to a maldistribution of the data. In future projects, we should use a bigger study population so that each fold can be composed of a larger sample size to train the classifiers. As well, a future study population should be composed of children and adolescents with and without musculoskeletal pathologies, in order to start creating a database more adapted to those ages.

Amid the development of this project, we applied to the Born from Knowledge program of the National Innovation Agency. We participated in this program as one of the teams in the first phase, in which we received positive feedback regarding the idea and the overall objectives.

In the future, besides correcting the limitations mentioned throughout this dissertation, we shall start the creation of the prototype kit that will be used by the subjects to assess their posture. This kit is a backpack that has inside an adjustable full body harness with all the markers stitched, the ruler to be used in the images as a reference for the scale and for the vertical to the ground, and signs to be placed on the floor to mark the positions where the subject and the camera must be. The full body harness, besides being adjustable, will be made with black fabric straps with a width slightly greater than the diameter of the markers to reduce the influence of the subject’s clothing on their detection. An example of this harness is presented in appendix E. After the development of the backpack kit, it is necessary to carry out tests with the final version of the backpack kit and the software to assess the quality of the data obtained.

During this thesis, the software was an offline software in Python. In the future, when we develop the software as a mobile app, it will only be able to store and process the subject’s data (the calculated body metrics and the final classification) in a secure server. The images captured must be automatically deleted for data and identity protection purposes.

The final goal is to create protocols with public and private schools and sports associations to be able to test the software in a school-aged population after the development of the software and kit.

Although there are some limitations, this dissertation brought some advances in the development of our software. This work will contribute to the screening of musculoskeletal pain and disorders, such as LBP and scoliosis, in children and adolescents, leading to a decrease in the probability of their appearance in adulthood or of progression to a chronic state.

7. References

[1] P. Galozzi *et al.*, "Prevalence of low back pain and its effect on health-related quality of life in 409 scholar adolescents from the Veneto region.," *Reumatismo*, vol. 71, no. 3, pp. 132–140, Oct. 2019, doi: 10.4081/reumatismo.2019.1229.

[2] T. W. Barreto and K. W. Lin, "Noninvasive Treatments for Low Back Pain.," *Am. Fam. Physician*, vol. 96, no. 5, pp. 324–327, Sep. 2017.

[3] R. Froud *et al.*, "A systematic review and meta-synthesis of the impact of low back pain on people's lives.," *BMC Musculoskelet. Disord.*, vol. 15, no. 1, p. 50, Feb. 2014, doi: 10.1186/1471-2474-15-50.

[4] L. Coelho, V. Almeida, and R. Oliveira, "Lombalgia nos adolescentes: identificação de factores de risco psicossociais . Estudo epidemiológico na Região da Grande Lisboa," vol. 23, pp. 81–90, 2005.

[5] P. O'Sullivan, A. Smith, D. Beales, and L. Straker, "Understanding Adolescent Low Back Pain From a Multidimensional Perspective: Implications for Management," *J. Orthop. Sport. Phys. Ther.*, vol. 47, no. 10, pp. 741–751, Oct. 2017, doi: 10.2519/jospt.2017.7376.

[6] I. Dianat, N. Sorkhi, A. Pourhossein, A. Alipour, and M. Asghari-Jafarabadi, "Neck, shoulder and low back pain in secondary schoolchildren in relation to schoolbag carriage: should the recommended weight limits be gender-specific?," *Appl. Ergon.*, vol. 45, no. 3, pp. 437–42, May 2014, doi: 10.1016/j.apergo.2013.06.003.

[7] T. E. Dorner and R. Crevenna, "Preventive aspects regarding back pain.," *Wien. Med. Wochenschr.*, vol. 166, no. 1–2, pp. 15–21, Feb. 2016, doi: 10.1007/s10354-015-0413-2.

[8] J. MacDonald, E. Stuart, and R. Rodenberg, "Musculoskeletal Low Back Pain in School-aged Children: A Review.," *JAMA Pediatr.*, vol. 171, no. 3, pp. 280–287, Mar. 2017, doi: 10.1001/jamapediatrics.2016.3334.

[9] B. Minghelli, "Musculoskeletal spine pain in adolescents: Epidemiology of non-specific neck and low back pain and risk factors," *J. Orthop. Sci.*, vol. 25, no. 5, pp. 776–780, Sep. 2020, doi: 10.1016/j.jos.2019.10.008.

[10] S. Lee and J. Shim, "The effects of backpack loads and spinal stabilization exercises on the dynamic foot pressure of elementary school children with idiopathic scoliosis.," *J. Phys. Ther. Sci.*, vol. 27, no. 7, pp. 2257–60, Jul. 2015, doi: 10.1589/jpts.27.2257.

[11] S. Bozkurt *et al.*, "Hypermobility Frequency in School Children: Relationship With Idiopathic Scoliosis, Age, Sex and Musculoskeletal Problems.," *Arch. Rheumatol.*, vol. 34, no. 3, pp. 268–273, Sep. 2019, doi: 10.5606/ArchRheumatol.2019.7181.

[12] M. Latalski *et al.*, "Risk factors of postural defects in children at school age.," *Ann. Agric. Environ. Med.*, vol. 20, no. 3, pp. 583–7, 2013.

[13] I. Ghanem and M. Rizkallah, "Adolescent idiopathic scoliosis for the primary care physician: frequently asked questions.," *Curr. Opin. Pediatr.*, vol. 31, no. 1, pp. 48–53, Feb. 2019, doi: 10.1097/MOP.0000000000000705.

[14] E. Z. Gieysztor, L. Sadowska, A. M. Choińska, and M. Paprocka-Borowicz, "Trunk rotation due to persistence of primitive reflexes in early school-age children.," *Adv. Clin. Exp. Med.*, vol. 27, no. 3, pp. 363–366, Mar. 2018, doi: 10.17219/acem/67458.

[15] M. E. R. C. Pinho, M. A. P. Vaz, P. M. F. M. Arezes, J. C. R. Campos, and A. P. B. De Magalhães, "Lesões músculo-esqueléticas relacionadas com as atividades desportivas em crianças e adolescentes: Uma revisão das questões emergentes," *Motricidade*, vol. 9, no. 1, pp. 31–48, Mar. 2013, doi: 10.6063/motricidade.9(1).2461.

[16] Instituto Nacional de Estatística (INE), "Inquérito Nacional de Saúde 2019," 2020.

[17] Instituto Nacional de Estatística (INE), "Portal do INE - Base de Dados Tipo de Doença Crónica." https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0008800&contexto=bd&selTab=tab2&xlang=pt (accessed Dec. 02, 2020).

[18] Instituto Nacional de Estatística (INE), "Portal do INE - Base de Dados População Residente." https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&contecto=pi&indOcorrCod=0008273&selTab=tab0 (accessed Dec. 02, 2020).

[19] EuroStat, "Statistics Persons reporting a chronic disease | Eurostat." https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_CD1E__custom_3979918/default/table?lang=en (accessed Dec. 02, 2020).
[20] Grupo OralMED Saúde, "Programa Educativo Escolas a Sorrir | OralMED Kids," 2020. https://www.oralmedkids.pt/programa-educativo (accessed Dec. 02, 2020).
[21] M. Gouveia and M. Augusto, "Custos indirectos da dor crónica em Portugal," Rev. Port. Saúde Pública, vol. 29, no. 2, pp. 100–107, Jul. 2011, doi: 10.1016/S0870-9025(11)70013-X.
[22] S. Dagenais, J. Caro, and S. Haldeman, "A systematic review of low back pain cost of illness studies in the United States and internationally.," *Spine J.*, vol. 8, no. 1, pp. 8–20, Jan. 2008, doi: 10.1016/j.spinee.2007.10.005.
[23] Instituto Nacional de Estatística (INE), "Acidentes de trabalho e problemas de saúde relacionados com o trabalho – Módulo ad hoc do Inquérito ao Emprego 2.º trimestre de 2020," 2020.
[24] K. S. Saladin, *Human Anatomy*, 2nd ed. New York: McGraw-Hill, 2008.
[25] R. Seeley, C. VanPutte, J. Regan, and A. Russo, *Seeley's Anatomy & Physiology*, 10th ed. New York: McGraw-Hill, 2013.
[26] E. Kenanidis, D. I. Athanasiadis, G. Geropoulos, P. Kakoulidis, M. Potoupnis, and E. Tsiridis, "Does the sternum play a role in the aetiopathogenesis of adolescent idiopathic scoliosis? Preliminary data of a new theory.," *Hippokratia*, vol. 22, no. 4, pp. 173–177, 2018.
[27] B. Minghelli, R. Oliveira, and C. Nunes, "Postural habits and weight of backpacks of Portuguese adolescents: Are they associated with scoliosis and low back pain?," *Work*, vol. 54, no. 1, pp. 197–208, May 2016, doi: 10.3233/WOR-162284.
[28] T. R. Toelle, D. A. Utpadel-Fischler, K.-K. Haas, and J. A. Priebe, "App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial," *npj Digit. Med.*, vol. 2, no. 1, p. 34, Dec. 2019, doi: 10.1038/s41746-019-0109-x.
[29] S. M. Siivola, S. Levoska, K. Latvala, E. Hoskio, H. Vanharanta, and S. Keinänen-Kiukaanniemi, "Predictive Factors for Neck and Shoulder Pain: A Longitudinal Study in Young Adults," *Spine (Phila. Pa. 1976).*, vol. 29, no. 15, pp. 1662–1669, Aug. 2004, doi: 10.1097/01.BRS.0000133644.29390.43.
[30] K. M. Houghton, "Review for the generalist: evaluation of low back pain in children and adolescents.," *Pediatr. Rheumatol. Online J.*, vol. 8, no. 1, p. 28, Nov. 2010, doi: 10.1186/1546-0096-8-28.
[31] G. T. Jones and G. J. Macfarlane, "Epidemiology of low back pain in children and adolescents.," *Arch. Dis. Child.*, vol. 90, no. 3, pp. 312–6, Mar. 2005, doi: 10.1136/adc.2004.056812.
[32] L. Hestbaek, C. Leboeuf-Yde, K. O. Kyvik, and C. Manniche, "The course of low back pain from adolescence to adulthood: eight-year follow-up of 9600 twins.," *Spine (Phila. Pa. 1976).*, vol. 31, no. 4, pp. 468–72, Feb. 2006, doi: 10.1097/01.brs.0000199958.04073.d9.
[33] B. Minghelli, R. Oliveira, and C. Nunes, "Non-specific low back pain in adolescents from the south of Portugal: prevalence and associated factors.," *J. Orthop. Sci.*, vol. 19, no. 6, pp. 883–92, Nov. 2014, doi: 10.1007/s00776-014-0626-z.
[34] H. Breivik, B. Collett, V. Ventafridda, R. Cohen, and D. Gallacher, "Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment.," *Eur. J. Pain*, vol. 10, no. 4, pp. 287–333, May 2006, doi: 10.1016/j.ejpain.2005.06.009.
[35] K. D. Watson *et al.*, "Low back pain in schoolchildren: occurrence and characteristics.," *Pain*, vol. 97, no. 1–2, pp. 87–92, May 2002, doi: 10.1016/s0304-3959(02)00008-8.
[36] J. M. Fritz and S. N. Clifford, "Low back pain in adolescents: a comparison of clinical outcomes in sports participants and nonparticipants.," *J. Athl. Train.*, vol. 45, no. 1, pp. 61–6, Jan. 2010, doi: 10.4085/1062-6050-45.1.61.
[37] M. R. Konieczny, H. Senyurt, and R. Krauspe, "Epidemiology of adolescent idiopathic scoliosis," *J. Child. Orthop.*, vol. 7, no. 1, pp. 3–9, Feb. 2013, doi: 10.1007/s11832-012-0457-4.
[38] J. P. Horne, R. Flannery, and S. Usman, "Adolescent idiopathic scoliosis: diagnosis and management," *Am. Fam. Physician*, vol. 89, no. 3, pp. 193–198, Feb. 2014.
[39] K. A. Greiner, "Adolescent idiopathic scoliosis: radiologic decision-making.," *Am. Fam. Physician*, vol. 65, no. 9, pp. 1817–22, May 2002.

[40] A. Jada *et al.*, “Evaluation and management of adolescent idiopathic scoliosis: a review,” *Neurosurg. Focus*, vol. 43, no. 4, p. E2, Oct. 2017, doi: 10.3171/2017.7.FOCUS17297.
[41] I. D. Gelalis, S. Ristanis, A. Nikolopoulos, A. Politis, C. Rigas, and T. Xenakis, “Loading rate patterns in scoliotic children during gait: the impact of the schoolbag carriage and the importance of its position.,” *Eur. Spine J.*, vol. 21, no. 10, pp. 1936–41, Oct. 2012, doi: 10.1007/s00586-012-2328-6.
[42] P. Mahaudens, C. Detrembleur, M. Mousny, and X. Banse, “Gait in adolescent idiopathic scoliosis: energy cost analysis,” *Eur. Spine J.*, vol. 18, no. 8, pp. 1160–1168, Aug. 2009, doi: 10.1007/s00586-009-1002-0.
[43] K.-D. Kim and P.-N. Hwangbo, “Effects of the Schroth exercise on the Cobb’s angle and vital capacity of patients with idiopathic scoliosis that is an operative indication.,” *J. Phys. Ther. Sci.*, vol. 28, no. 3, pp. 923–6, Mar. 2016, doi: 10.1589/jpts.28.923.
[44] R. Gunzburg *et al.*, “Low back pain in a population of school children,” *Eur. Spine J.*, vol. 8, no. 6, pp. 439–443, Dec. 1999, doi: 10.1007/s005860050202.
[45] S. A. Shah and J. Saller, “Evaluation and Diagnosis of Back Pain in Children and Adolescents,” *J. Am. Acad. Orthop. Surg.*, vol. 24, no. 1, pp. 37–45, Jan. 2016, doi: 10.5435/JAAOS-D-14-00130.
[46] F. Zhu *et al.*, “Rib length asymmetry in thoracic adolescent idiopathic scoliosis: is it primary or secondary?,” *Eur. Spine J.*, vol. 20, no. 2, pp. 254–259, Feb. 2011, doi: 10.1007/s00586-010-1637-x.
[47] S. P. Cohen, “Epidemiology, diagnosis, and treatment of neck pain.,” *Mayo Clin. Proc.*, vol. 90, no. 2, pp. 284–99, Feb. 2015, doi: 10.1016/j.mayocp.2014.09.008.
[48] K. Grimmer and M. Williams, “Gender-age environmental associates of adolescent low back pain.,” *Appl. Ergon.*, vol. 31, no. 4, pp. 343–60, Aug. 2000, doi: 10.1016/s0003-6870(00)00002-8.
[49] I. Dianat, Z. Javadivala, and H. Allahverdipour, “School Bag Weight and the Occurrence of Shoulder, Hand/Wrist and Low Back Symptoms among Iranian Elementary Schoolchildren.,” *Heal. Promot. Perspect.*, vol. 1, no. 1, pp. 76–85, 2011, doi: 10.5681/hpp.2011.008.
[50] S. Dockrell, C. Simms, and C. Blake, “Schoolbag carriage and schoolbag-related musculoskeletal discomfort among primary school children.,” *Appl. Ergon.*, vol. 51, pp. 281–90, Nov. 2015, doi: 10.1016/j.apergo.2015.05.009.
[51] G. I. Sheir-Neiss, R. W. Kruse, T. Rahman, L. P. Jacobson, and J. A. Pelli, “The Association of Backpack Use and Back Pain in Adolescents,” *Spine (Phila. Pa. 1976).*, vol. 28, no. 9, pp. 922–930, May 2003, doi: 10.1097/01.BRS.0000058725.18067.F7.
[52] S. Negrini and R. Carabalona, “Backpacks on! Schoolchildren’s Perceptions of Load, Associations With Back Pain and Factors Determining the Load,” *Spine (Phila. Pa. 1976).*, vol. 27, no. 2, pp. 187–195, Jan. 2002, doi: 10.1097/00007632-200201150-00014.
[53] K. Watanabe *et al.*, “Physical Activities and Lifestyle Factors Related to Adolescent Idiopathic Scoliosis,” *J. Bone Jt. Surg.*, vol. 99, no. 4, pp. 284–294, Feb. 2017, doi: 10.2106/JBJS.16.00459.
[54] P. Korovessis, G. Koureas, S. Zacharatos, and Z. Papazisis, “Backpacks, Back Pain, Sagittal Spinal Curves and Trunk Alignment in Adolescents,” *Spine (Phila. Pa. 1976).*, vol. 30, no. 2, pp. 247–255, Jan. 2005, doi: 10.1097/01.brs.0000150505.59690.1b.
[55] S. Negrini and A. Negrini, “Postural effects of symmetrical and asymmetrical loads on the spines of schoolchildren.,” *Scoliosis*, vol. 2, no. 1, p. 8, Jul. 2007, doi: 10.1186/1748-7161-2-8.
[56] J. T. F. Lopes, “O Transporte de Cargas em Mochilas Escolares e o Desenvolvimento Motor Harmonioso das Crianças,” Universidade do Porto - Faculdade de Ciências do Desporto e Educação Física, 2002.
[57] Direcção-Geral da Saúde and DGS, “Programa Nacional de Saúde Escolar 2015,” 2015.
[58] Ministério da Educação, “Mochila Leve - Sugestões para levar às costas,” 2019.
[59] A. G. Silva, P. Sa-Couto, A. Queirós, M. Neto, and N. P. Rocha, “Pain, pain intensity and pain disability in high school students are differently associated with physical activity, screening hours and sleep.,” *BMC Musculoskelet. Disord.*, vol. 18, no. 1, p. 194, Dec. 2017, doi: 10.1186/s12891-017-1557-6.
[60] K. Hébert-Losier and F. Abd Rahman, “Reliability of postural measures in elite badminton players using Posture Pro 8.,” *Physiother. Theory Pract.*, vol. 34, no. 6, pp. 483–494, Jun. 2018,

doi: 10.1080/09593985.2017.1420117.
[61] S. Huber, J. A. Priebe, K.-M. Baumann, A. Plidschun, C. Schiessl, and T. R. Tölle, “Treatment of Low Back Pain with a Digital Multidisciplinary Pain Treatment App: Short-Term Results.,” *JMIR Rehabil. Assist. Technol.*, vol. 4, no. 2, p. e11, Dec. 2017, doi: 10.2196/rehab.9032.
[62] X. B. Petermann and E. C. W. Meereis, “Postura corporal: Uma revisão sistemática sobre métodos de avaliação,” *Man. Ther. Posturology Rehabil. J.*, vol. 14, no. February, p. 273, Feb. 2016, doi: 10.17784/mtprehabjournal.2016.14.273.
[63] E. A. G. Ferreira, M. Duarte, E. P. Maldonado, T. N. Burke, and A. P. Marques, “Postural assessment software (PAS/SAPO): validation and reliabiliy,” *Clinics*, vol. 65, no. 7, pp. 675–681, Jul. 2010, doi: 10.1590/S1807-59322010000700005.
[64] E. A. Ferreira, M. Duarte, E. P. Maldonado, A. A. Bersanetti, and A. P. Marques, “Quantitative assessment of postural alignment in young adults based on photographs of anterior, posterior, and lateral views.,” *J. Manipulative Physiol. Ther.*, vol. 34, no. 6, pp. 371–80, Jul. 2011, doi: 10.1016/j.jmpt.2011.05.018.
[65] CONTEMPLAS GmbH, “CONTEMPLAS - professional motion analysis software,” 2022. https://contemplas.com/en/motion-analysis/ (accessed Mar. 05, 2022).
[66] D. M. Boland, E. V. Neufeld, J. Ruddell, B. A. Dolezal, and C. B. Cooper, “Inter- and intra-rater agreement of static posture analysis using a mobile application.,” *J. Phys. Ther. Sci.*, vol. 28, no. 12, pp. 3398–3402, Dec. 2016, doi: 10.1589/jpts.28.3398.
[67] K. A. Szucs and E. V. D. Brown, “Rater reliability and construct validity of a mobile application for posture analysis,” *J. Phys. Ther. Sci.*, vol. 30, no. 1, pp. 31–36, Jan. 2018, doi: 10.1589/jpts.30.31.
[68] F. P. Kendall, E. K. McCreary, and P. G. Provance, *Muscles: Testing and Function with Posture and Pain*, 5th ed. Lippincott Williams & Wilkins, 2005.
[69] T. Silveira Furlanetto, J. Adami Sedrez, C. Tarragô Candotti, and J. Fagundes Loss, “Photogrammetry as a tool for the postural evaluation of the spine: A systematic review,” *World J. Orthop.*, vol. 7, no. 2, pp. 136–148, 2016, doi: 10.5312/wjo.v7.i2.136.
[70] G. A. Meliscki, L. Z. Monteiro, and C. A. Giglio, “Avaliação postural de nadadores e sua relação com o tipo de respiração,” *Fisioter. em Mov.*, vol. 24, no. 4, pp. 721–728, Dec. 2011, doi: 10.1590/S0103-51502011000400017.
[71] U.-V. Albrecht and U. von Jan, “Safe, sound and desirable: development of mHealth apps under the stress of rapid life cycles.,” *mHealth*, vol. 3, no. 5, p. 27, Jul. 2017, doi: 10.21037/mhealth.2017.06.05.
[72] Kaia Health, “Kaia Health | Digital Therapeutics to Manage Chronic Diseases.” https://www.kaiahealth.com/ (accessed Dec. 27, 2020).
[73] Kiio Inc., “Kiio for Digital MSK Solutions.” https://www.kiio.com/ (accessed Jan. 07, 2021).
[74] A. Ming and H. Ma, “A blob detector in color images,” in *Proceedings of the 6th ACM international conference on Image and video retrieval - CIVR ’07*, 2007, pp. 364–370, doi: 10.1145/1282280.1282335.
[75] T. Lindeberg, “Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention,” *Int. J. Comput. Vis.*, vol. 11, no. 3, pp. 283–318, Dec. 1993, doi: 10.1007/BF01469346.
[76] G. Bradski, “The OpenCV Library: SimpleBlobDetector Class Reference.” https://docs.opencv.org/4.5.1/d0/d7a/classcv_1_1SimpleBlobDetector.html (accessed Jan. 07, 2021).
[77] A. Baíllo and J. E. Chacón, “Statistical outline of animal home ranges: An application of set estimation,” *Handb. Stat.*, vol. 44, pp. 3–37, 2021, doi: 10.1016/bs.host.2020.10.002.
[78] R. M. Stark and C. C. Morris, *Finite Mathematics: Models and Applications*. Wiley, 2015.
[79] Y. Bee Wah and N. Mohd Razali, “Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests,” *J. Stat. Model. Anal.*, vol. 2, no. November, pp. 21–33, 2011.
[80] B. Rosner, *Fundamentals of Biostatistics*, 8th ed. Boston: Cengage Learning, 2015.
[81] D. Lakens, “Equivalence Tests,” *Soc. Psychol. Personal. Sci.*, vol. 8, no. 4, pp. 355–362, May 2017, doi: 10.1177/1948550617697177.
[82] D. L. Streiner, “Unicorns Do Exist: A Tutorial on ‘Proving’ the Null Hypothesis,” *Can. J.*

Psychiatry, vol. 48, no. 11, pp. 756–761, Dec. 2003, doi: 10.1177/070674370304801108.
[83] O. Altay and M. Ulas, "Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children," *6th International Symposium on Digital Forensic and Security, ISDFS 2018 - Proceeding*, vol. 2018-Janua. pp. 1–4, 2018, doi: 10.1109/ISDFS.2018.8355354.
[84] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*. CRC press, 2012.
[85] I. Stanimirova, M. Daszykowski, and B. Walczak, *Robust Methods in Analysis of Multivariate Food Chemistry Data*, 1st ed., vol. 28. Copyright © 2013 Elsevier B.V. All rights reserved., 2013.
[86] N. Mohanty, A. L. S. John, R. Manmatha, and T. M. Rath, *Shape-based image classification and retrieval*, vol. 31. Elsevier B.V., 2013.
[87] S. Balakrishnama, A. Ganapathiraju, and I. Processing, "Linear discriminant analysis-a brief tutorial," *Inst. Signal Inf. Process.*, vol. 18, no. 1998, pp. 1–8, 1998.
[88] J. Ye, R. Janardan, and Q. Li, *Two-Dimensional Linear Discriminant Analysis.*, vol. 17. 2004.
[89] T. Hastie, J. Friedman, and R. Tibshirani, *The Elements of Statistical Learning*, 2nd ed., no. 2. New York, NY: Springer New York, 2001.
[90] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2888, pp. 986–996, 2003, doi: 10.1007/978-3-540-39964-3_62.
[91] A. Burkov, *The Hundred-Page Machine Learning Book*. Andriy Burkov, 2019.
[92] K. S. Kim, H. H. Choi, C. S. Moon, and C. W. Mun, "Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions," *Curr. Appl. Phys.*, vol. 11, no. 3, pp. 740–745, 2011, doi: 10.1016/j.cap.2010.11.051.
[93] C. Castaldello *et al.*, *A Model-Based Support for Diagnosing von Willebrand Disease*, vol. 40. Elsevier Masson SAS, 2017.
[94] Vaibhaw, J. Sarraf, and P. K. Pattnaik, *Brain-computer interfaces and their applications*. INC, 2020.
[95] S. B. Imandoust and M. Bolandraftar, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events : Theoretical Background," *Int. J. Eng. Res. Appl.*, vol. 3, no. 5, pp. 605–610, 2013.
[96] D. Berrar, "Cross-Validation," in *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, 2019, pp. 542–545.
[97] A. G. Davide Anguita, Luca Ghelardoni and L. O. and S. Ridella, "The 'K' in K-fold Cross Validation," no. April, pp. 25–27, 2012.

A. Questionnaire given to the study participants to be fulfilled

Inquérito

Estudo: “PosturAll: A Posture Assessment Software for Children”

1. Dados Pessoais

1.1. Idade: _____

1.2. Sexo: Masculino ☐
Feminino ☐

1.3. Altura (cm): _____

1.4. Peso (kg): _____

2. Dados Gerais

Teve no último ano dores na coluna: Sim ☐ Não ☐

Se respondeu sim à anterior, assinale todas as regiões da coluna onde se localizaram as dores:

Cervical ☐ Dorsal ☐ Lombar ☐

Já recorreu a consultas médicas devido a dores nas costas?

Sim ☐ Não ☐

Se sim, foi diagnosticado com alguma evidência de problema postural?

Sim ☐ Qual? _____ Não ☐

B. Images to Select the Color of the Markers

White Ping Pong Markers:

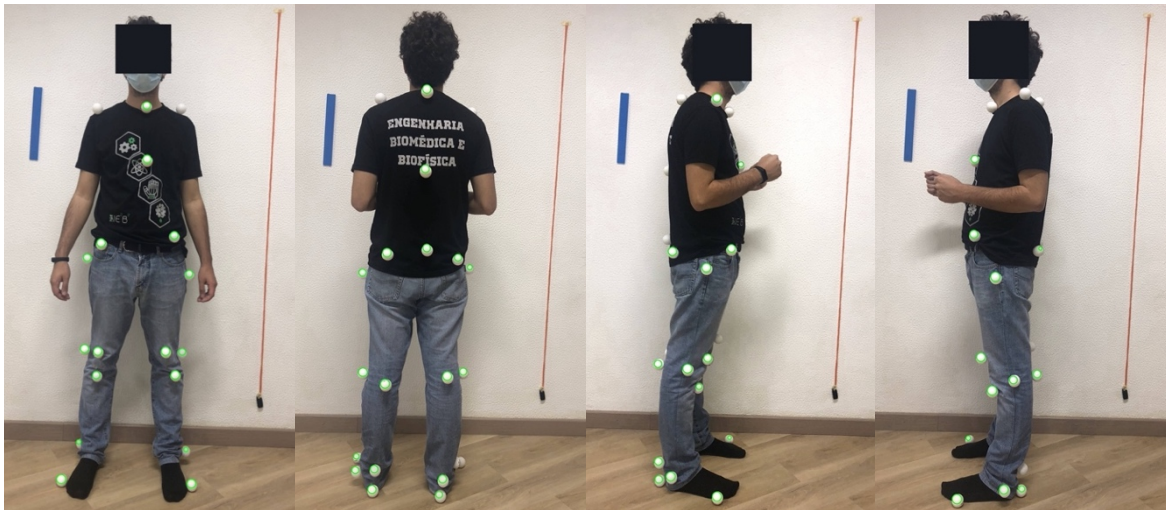


Figure B.1: Images captured with flash of the participant wearing a black shirt and jeans while using white ping pong balls as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

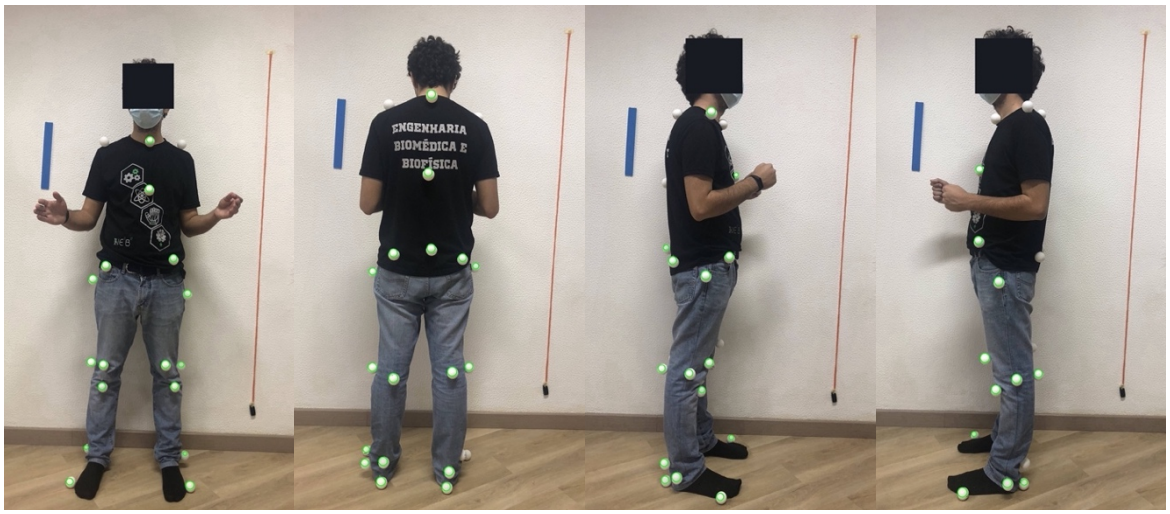


Figure B.2: Images captured without flash of the participant wearing a black shirt and jeans while using white ping pong balls as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.3: Images captured with flash of the participant wearing a yellow shirt and gray pants while using white ping pong balls as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.4: Images captured without flash of the participant wearing a yellow shirt and gray pants while using white ping pong balls as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.5: Images captured with flash of the participant wearing a brown shirt and blue pants while using white ping pong balls as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.6: Images captured without flash of the participant wearing a brown shirt and blue pants while using white ping pong balls as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

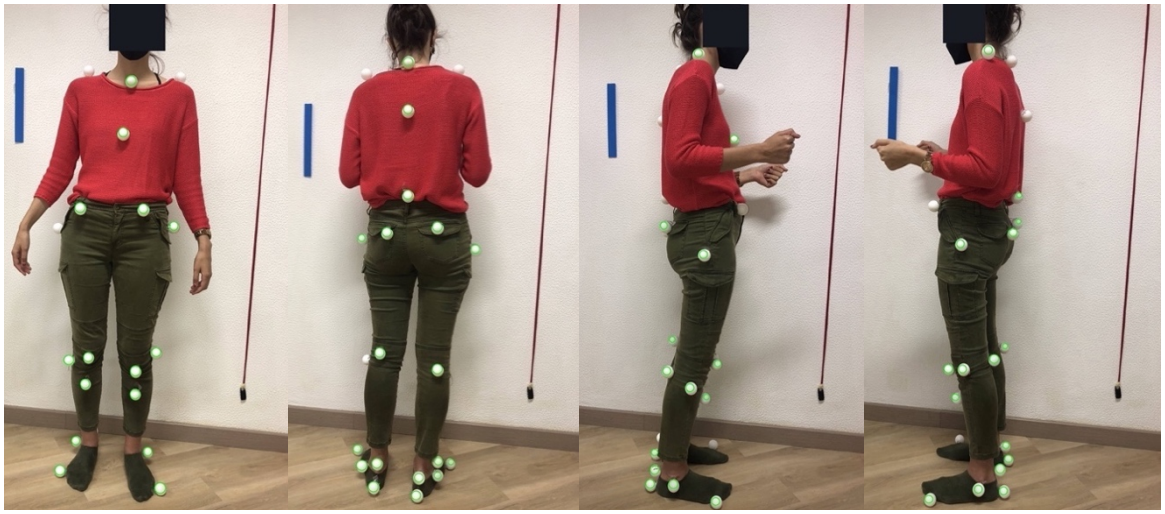


Figure B.7: Images captured with flash of the participant wearing a red shirt and green pants while using white ping pong balls as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

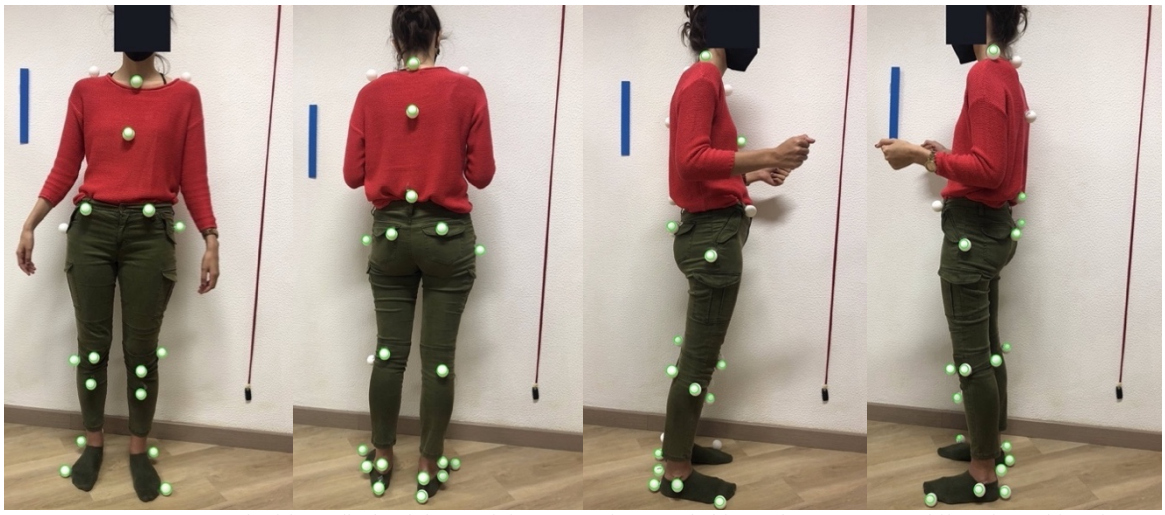


Figure B.8: Images captured without flash of the participant wearing a red shirt and green pants while using white ping pong balls as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

White Styrofoam Markers:

Figure B.9: Images captured with flash of the participant wearing a black shirt and jeans while using white styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.10: Images captured without flash of the participant wearing a black shirt and jeans while using white styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.11: Images captured with flash of the participant wearing a yellow shirt and gray pants while using white styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.12: Images captured without flash of the participant wearing a yellow shirt and gray pants while using white styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.13: Images captured with flash of the participant wearing a brown shirt and blue pants while using white styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.14: Images captured without flash of the participant wearing a brown shirt and blue pants while using white styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.15: Images captured with flash of the participant wearing a red shirt and green pants while using white styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.16: Images captured without flash of the participant wearing a red shirt and green pants while using white styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Yellow Styrofoam Markers:

Figure B.17: Images captured with flash of the participant wearing a black shirt and jeans while using yellow styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.18: Images captured without flash of the participant wearing a black shirt and jeans while using yellow styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.19: Images captured with flash of the participant wearing a yellow shirt and gray pants while using yellow styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.20: Images captured without flash of the participant wearing a yellow shirt and gray pants while using yellow styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.21: Images captured with flash of the participant wearing a brown shirt and blue pants while using yellow styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.22: Images captured without flash of the participant wearing a brown shirt and blue pants while using yellow styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.23: Images captured with flash of the participant wearing a red shirt and green pants while using yellow styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.24: Images captured without flash of the participant wearing a red shirt and green pants while using yellow styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Orange Styrofoam Markers:

Figure B.25: Images captured with flash of the participant wearing a black shirt and jeans while using orange styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

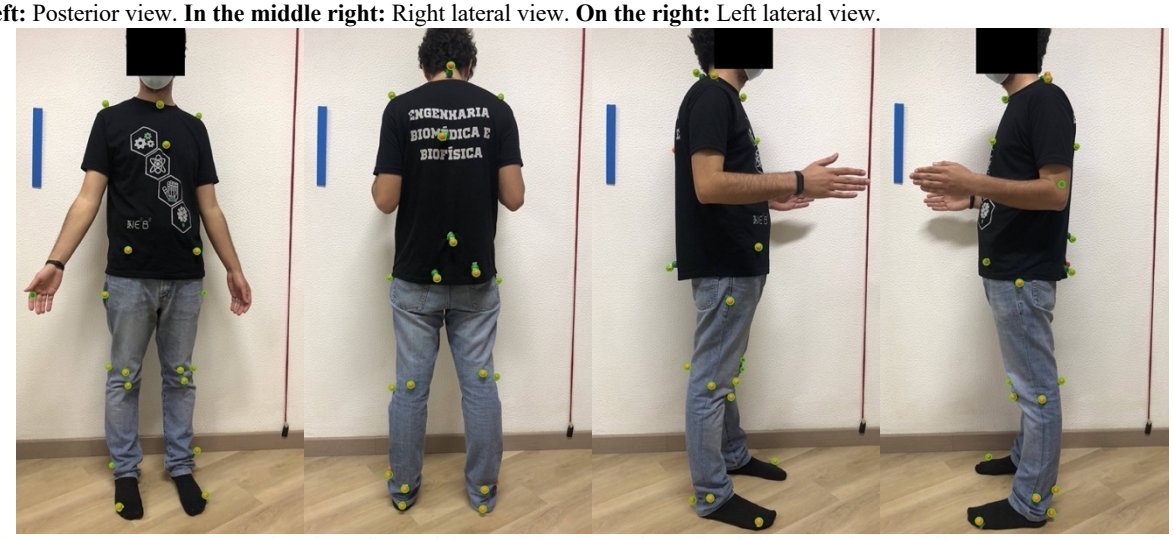


Figure B.26: Images captured without flash of the participant wearing a black shirt and jeans while using orange styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.27: Images captured with flash of the participant wearing a yellow shirt and gray pants while using orange styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.28: Images captured without flash of the participant wearing a yellow shirt and gray pants while using orange styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.29: Images captured with flash of the participant wearing a brown shirt and blue pants while using orange styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.30: Images captured without flash of the participant wearing a brown shirt and blue pants while using orange styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.31: Images captured with flash of the participant wearing a red shirt and green pants while using orange styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.32: Images captured without flash of the participant wearing a red shirt and green pants while using orange styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Red Styrofoam Markers:

Figure B.33: Images captured with flash of the participant wearing a black shirt and jeans while using red styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.34: Images captured without flash of the participant wearing a black shirt and jeans while using red styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.35: Images captured with flash of the participant wearing a yellow shirt and gray pants while using red styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.36: Images captured without flash of the participant wearing a yellow shirt and gray pants while using red styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.37: Images captured with flash of the participant wearing a brown shirt and blue pants while using orange styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.38: Images captured without flash of the participant wearing a brown shirt and blue pants while using orange styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.39: Images captured with flash of the participant wearing a red shirt and green pants while using red styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.40: Images captured without flash of the participant wearing a red shirt and green pants while using red styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Blue Styrofoam Markers:

Figure B.41: Images captured with flash of the participant wearing a black shirt and jeans while using blue styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

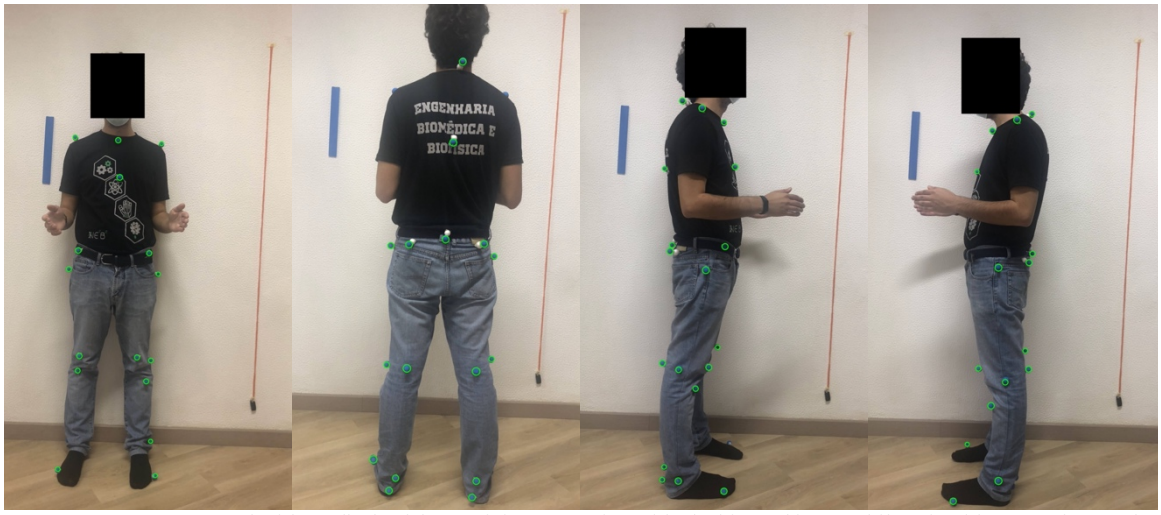


Figure B.42: Images captured without flash of the participant wearing a black shirt and jeans while using blue styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.43: Images captured with flash of the participant wearing a yellow shirt and gray pants while using blue styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.44: Images captured without flash of the participant wearing a yellow shirt and gray pants while using blue styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.45: Images captured with flash of the participant wearing a brown shirt and blue pants while using blue styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.46: Images captured without flash of the participant wearing a brown shirt and blue pants while using blue styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.47: Images captured with flash of the participant wearing a red shirt and green pants while using blue styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.48: Images captured without flash of the participant wearing a red shirt and green pants while using blue styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Green Styrofoam Markers:

Figure B.49: Images captured with flash of the participant wearing a black shirt and jeans while using green styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.50: Images captured without flash of the participant wearing a black shirt and jeans while using green styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.51: Images captured with flash of the participant wearing a yellow shirt and gray pants while using green styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.52: Images captured without flash of the participant wearing a yellow shirt and gray pants while using green styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.53: Images captured with flash of the participant wearing a brown shirt and blue pants while using green styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.54: Images captured without flash of the participant wearing a brown shirt and blue pants while using green styrofoam spheres as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.55: Images captured with flash of the participant wearing a red shirt and green pants while using green styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

Figure B.56: Images captured without flash of the participant wearing a red shirt and green pants while using green styrofoam as markers. The markers detected by the software are surrounded with green circles. **On the left:** Anterior view. **In the middle left:** Posterior view. **In the middle right:** Right lateral view. **On the right:** Left lateral view.

C. Images Captured for the Software Calibration Tests

Figure C.1: Images captured using flash with a mobile phone camera at a height of 120 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

Figure C.2: Images captured using flash with a mobile phone camera at a height of 110 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

Figure C.3: Images captured using flash with a mobile phone camera at a height of 100 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

Figure C.4: Images captured without using flash with a mobile phone camera at a height of 120 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

Figure C.5: Images captured without using flash with a mobile phone camera at a height of 110 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

Figure C.6: Images captured without using flash with a mobile phone camera at a height of 100 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

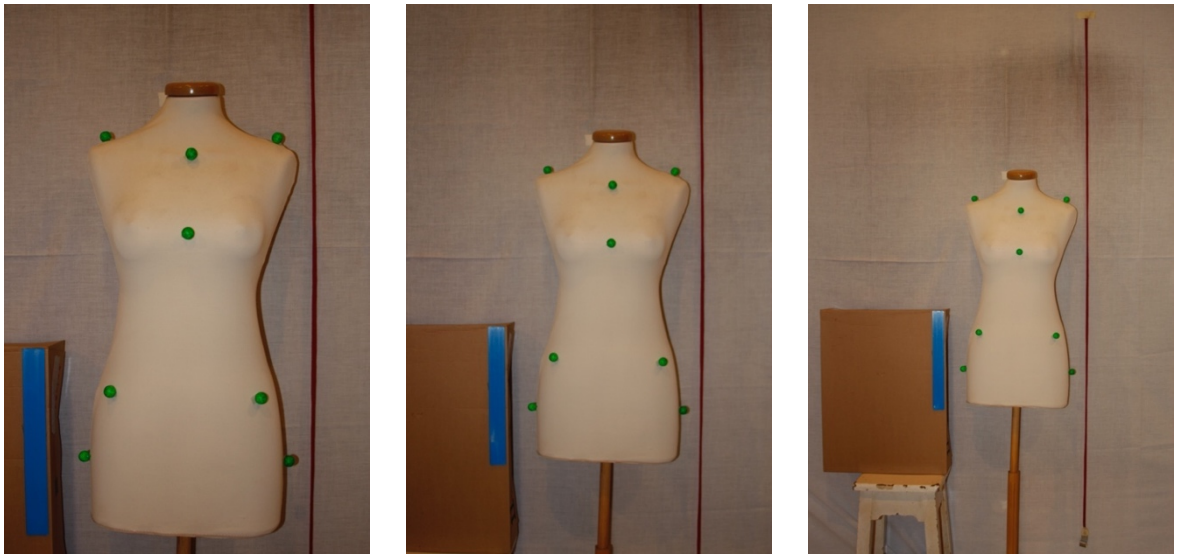


Figure C.7: Images captured using flash with a photographic camera at a height of 120 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

Figure C.8: Images captured using flash with a photographic camera at a height of 110 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

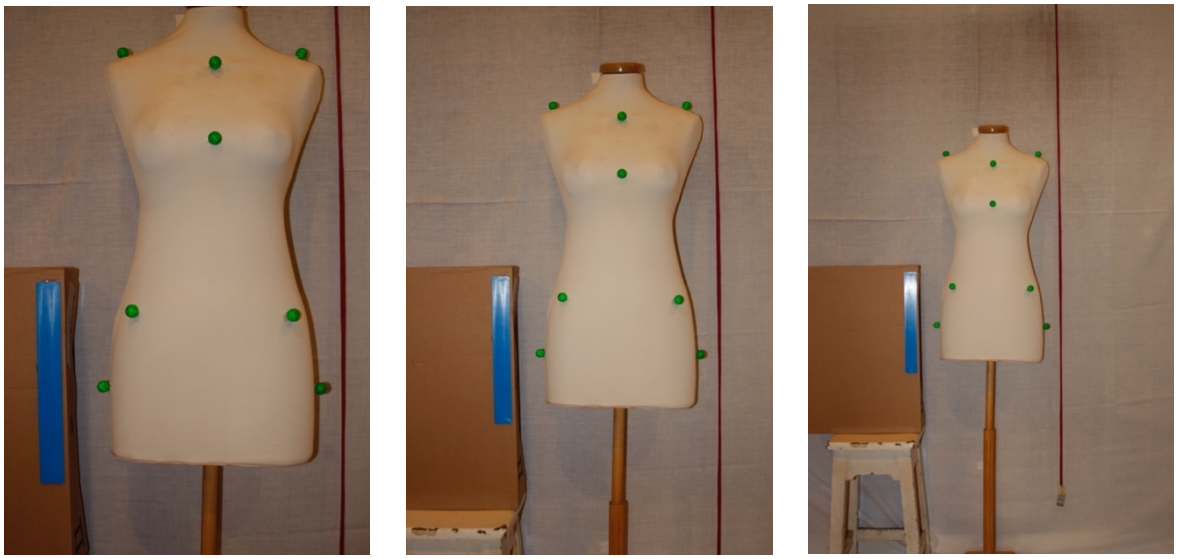


Figure C.9: Images captured using flash with a photographic camera at a height of 100 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

Figure C.10: Images captured without using flash with a photographic camera at a height of 120 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

Figure C.11: Images captured without using flash with a photographic camera at a height of 110 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

Figure C.12: Images captured without using flash with a photographic camera at a height of 100 cm. We varied the distance of the camera to the markers. The camera is positioned on the left at 125 cm, in the middle at 180 cm and on the right at 250 cm.

D. Results of the features calculated in the Software Calibration tests

Table D.1: Results of the distance and angles obtained in the analysis of the images captured with the mobile phone camera, varying the height and distance of the camera to the markers.

	Cell Phone Camera																	
	With Flash									Without Flash								
Camera Height	**100 cm**			**110 cm**			**120 cm**			**100 cm**			**110 cm**			**120 cm**		
Distance to the Markers	**125 cm**	**180 cm**	**250 cm**	**125 cm**	**180 cm**	**250 cm**	**125 cm**	**180 cm**	**250 cm**	**125 cm**	**180 cm**	**250 cm**	**125 cm**	**180 cm**	**250 cm**	**125 cm**	**180 cm**	**250 cm**
AHA_A (º)	-0.22	-0.25	-0.30	-0.23	-0.26	-0.31	-0.24	-0.27	-0.31	-0.23	-0.25	-0.30	-0.24	-0.25	-0.32	-0.25	-0.26	-0.32
ASA (º)	88.91	88.90	88.87	88.90	88.89	88.87	88.89	88.88	88.87	88.91	88.89	88.87	88.89	88.88	88.87	88.89	88.87	88.86
$AJND_L$ (cm)	14.00	14.02	14.04	14.01	14.03	14.05	14.03	14.04	14.04	14.01	14.03	14.04	14.02	14.04	14.05	14.04	14.04	14.05
$AJND_R$ (cm)	14.63	14.65	14.66	14.62	14.63	14.65	14.63	14.64	14.65	14.64	14.65	14.65	14.62	14.63	14.66	14.63	14.64	14.65
AAD_L (cm)	41.43	41.46	41.47	41.42	41.44	41.46	41.43	41.45	41.46	41.44	41.46	41.47	41.43	41.45	41.47	41.44	41.45	41.46
AAD_R (cm)	40.73	40.74	40.78	40.75	40.76	40.77	40.74	40.75	40.77	40.74	40.75	40.77	40.74	40.77	40.78	40.75	40.75	40.77
ASISHA (º)	-1.69	-1.73	-1.77	-1.66	-1.71	-1.75	-1.71	-1.74	-1.76	-1.70	-1.73	-1.78	-1.67	-1.72	-1.76	-1.71	-1.73	-1.77

Table D.2: Results of the distance and angles obtained in the analysis of the images captured with the photographic camera, varying the height and distance of the camera to the markers

	Photographic Camera																	
	With Flash									Without Flash								
Camera Height	**100 cm**			**110 cm**			**120 cm**			**100 cm**			**110 cm**			**120 cm**		
Distance to the Markers	**125 cm**	**180 cm**	**250 cm**	**125 cm**	**180 cm**	**250 cm**	**125 cm**	**180 cm**	**250 cm**	**125 cm**	**180 cm**	**250 cm**	**125 cm**	**180 cm**	**250 cm**	**125 cm**	**180 cm**	**250 cm**
AHA_A (º)	-0.20	-0.23	-0.30	-0.17	-0.24	-0.32	-0.18	-0.21	-0.34	-0.21	-0.27	-0.29	-0.18	-0.24	-0.31	-0.20	-0.23	-0.34
ASA (º)	88.90	88.86	88.84	88.89	88.87	88.85	88.92	88.87	88.84	88.90	88.86	88.84	88.89	88.87	88.84	88.91	88.88	88.85
$AJND_L$ (cm)	14.00	14.01	14.05	14.02	14.03	14.04	14.02	14.03	14.06	13.99	14.01	14.05	14.00	14.04	14.04	14.03	14.03	14.06
$AJND_R$ (cm)	14.60	14.64	14.66	14.62	14.65	14.67	14.63	14.64	14.68	14.61	14.64	14.67	14.63	14.64	14.67	14.62	14.65	14.68
AAD_L (cm)	41.45	41.47	41.49	41.44	41.47	41.48	41.45	41.49	41.50	41.43	41.47	41.48	41.44	41.48	41.49	41.42	41.47	41.48
AAD_R (cm)	40.72	40.73	40.75	40.73	40.74	40.78	40.74	40.75	40.77	40.72	40.73	40.75	40.74	40.74	40.77	40.75	40.76	40.77
ASISHA (º)	-1.65	-1.73	-1.79	-1.66	-1.70	-1.77	-1.66	-1.71	-1.76	-1.64	-1.71	-1.78	-1.65	-1.72	-1.75	-1.66	-1.72	-1.76

E. Prototype Kit

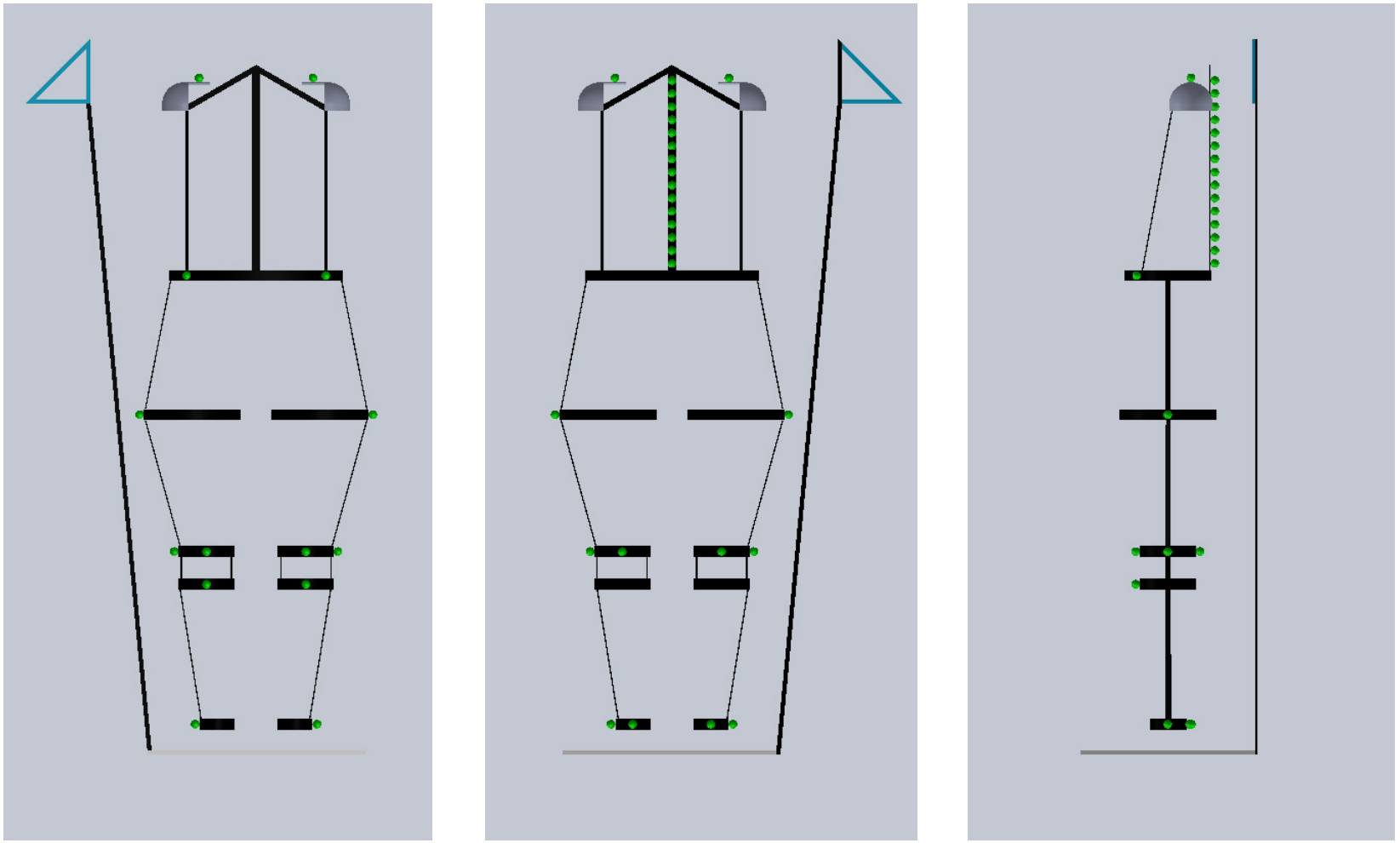


Figure E.1: Representation of the prototype kit to be used in the postural analysis. This consists of an adjustable full body harness with the green markers stitched, a mat to mark the place where the subject should be placed and the blue set square to obtain the scale of the image and the vertical line to the floor. **On the left:** Anterior view. **In the middle** Posterior view. **On the right:** Left lateral view.