Empirical Study of Dimensionality Reduction Methodologies for Pose Comparison problems using Computer Vision

Daniel Gerald Orbegoso Barrantes

Master en Ciencia de Datos
Computer Vision

Tutor/a de TF
Raúl Parada Medina(UOC)
Thomas Bohné (Cambridge)
Profesor/a responsable de la asignatura
Jordi Casas Roma

28.06.2023

Universitat Oberta
de Catalunya

Ficha del Trabajo Final

Título del trabajo:	Empirical Study of Dimensionality Reduction Methodologies for Pose Comparison using Computer Vision
Nombre del autor/a:	Daniel Gerald Orbegoso Barrantes
Nombre del Tutor/a de TF:	Raúl Parada Medina (UOC) Thomas Bohné (Cambridge)
Nombre del/de la PRA:	Jordi Casas Roma
Fecha de entrega:	25/06/2023
Titulación o programa:	Master Ciencia de Datos
Área del Trabajo Final:	Area 2
Idioma del trabajo:	Inglés
Palabras clave	Computer Vision, Dimensionality Reduction, Pose Comparison

Abstract (Spanish)

Esta investigación investiga la viabilidad de incorporar técnicas de reducción de dimensionalidad en problemas de comparación de posturas dentro del campo de visión por computadora. Esta investigación aborda la brecha en la literatura existente al investigar la viabilidad de incorporar técnicas de reducción de dimensiones en problemas de comparación de posturas dentro del campo de visión por computador y determinar las dimensiones clave que definen la correctitud de la postura. Para abordar estas limitaciones, esta investigación emplea técnicas como PCA, Random Forest y ANOVA. Se utiliza un conjunto de datos de Esqueleto 3D que contiene movimientos realizados por varios sujetos para comprender la correctitud de los movimientos entre los sujetos.

Se emplea la metodología CRISP-DM para llevar a cabo las diferentes fases de esta investigación. Los resultados demuestran una reducción exitosa del conjunto de datos de más del 80% al tiempo que conservan información significativa sobre la correctitud de los movimientos. Además, se entrenan dos modelos CNN distintos para clasificar la correctitud de los movimientos. El modelo entrenado con el conjunto de datos reducido muestra un tiempo de procesamiento un 45% más rápido y una mejora del 5% en precisión en comparación con el modelo CNN entrenado con el conjunto de datos completo.

Estos hallazgos resaltan la eficacia de incorporar técnicas de reducción de dimensiones en problemas de comparación de posturas dentro del campo de visión por computadora, ayudando a crear sistemas más eficientes, reduciendo el uso de recursos y producción de CO2. Las aplicaciones prácticas de esta investigación se extienden a campos como deportes, rehabilitación e interacción humano-computadora. La identificación de dimensiones clave que definen la correctitud de la postura puede mejorar la precisión de los modelos de estimación de postura existentes. El trabajo futuro implica la implementación de diferentes técnicas de selección y extracción de características como Kendall, LDA, t-SNE, el desarrollo de modelos Transformer o LSTM para mejorar la clasificación y la expansión de conjuntos de datos para incluir una gama más amplia de datos.

Abstract (English)

This research investigates the feasibility of incorporating dimensionality reduction techniques in pose comparison problems within the computer vision field. This research addresses the gap in existing literature by investigating the feasibility of incorporating dimensionality reduction techniques in pose comparison problems within the computer vision field and determining the key dimensions that define pose correctness. To address these limitations, this investigation employs techniques such as PCA, Random Forest, and ANOVA. A 3D Skeleton dataset containing movements performed by various subjects is utilized to understand the correctness of movements across subjects.

The CRISP-DM methodology is employed to carry out the different phases of the research. The results demonstrate successful reduction of the dataset by over 80% while retaining meaningful information regarding movement correctness. Additionally, two distinct CNN models are trained to classify movement correctness. The model trained with the reduced dataset exhibits a 45% faster processing time and a 5% improvement in accuracy compared to the CNN model trained with the complete dataset.

These findings highlight the effectiveness of incorporating dimensionality reduction techniques in pose comparison problems within computer vision, resulting in more efficient systems that reduce resource usage and CO2 production. The practical applications of this research extend to fields such as sports, rehabilitation, and human-computer interaction. The identification of key dimensions defining pose correctness can enhance the accuracy of existing pose estimation models. Future work entails implementing different feature selection and extraction techniques such as Kendall, LDA, t-SNE, developing Transformer or LSTM models to enhance classification, and expanding datasets to include a broader range of data.

Index

Table of Contents

List of Figures

1. Introduction

Pose estimation enables the identification of object positions and orientations through joint angle measurements. Pose comparison, on the other hand, involves assessing the similarity or difference between object poses. We can find several environments where the movement of subjects can be analysed, but we have selected the field of sports because the movements are usually explosive and dynamic. From the other side, depending in the specific sport movement, several parts of the body get involved, incrementing the resources necessary to process the huge amount of data. This research aims to investigate the definition of pose correctness by applying dimensionality reduction techniques to extract key attributes and improve accuracy and efficiency in pose estimation models. But the complexity when working with this kind of data is one of the biggest challenges. The subjects used to study can have different sizes, the movements used to compare can have different velocities or lengths, therefore different strategies should be defined and developed to prepare the data, before implementing the different dimensionality reduction algorithms.

The main motivation of this project is based on the desire to understand how we can define the correctness of sport movements, which parts of the body or attributes are the most important when trying to understand correctness, and how dimensionality reduction can be implemented on 3D Skeleton data to improve the results obtained in pose comparison. In Figure 1, we can appreciate the comparison between landmarks for two estimation models. How can we define the correctness of these movements? [1].

Figure 1. Pose comparison example. 3D Skeleton landmarks and connections on figures comparing two pose estimation model projections on the body of different sport practitioners.

1.1. Context and justification of the work

Pose estimation is an important part of Computer Vision [2]. We are able to identify the position and orientation of objects in an image or video by measuring the angles of joints. On the other hand, pose comparison involves comparing the poses of two or more objects in order to determine the similarity or difference of those objects. During the last years we had significant advancements in the field, but our models still have some limitations in terms of accuracy and efficiency [3]. The question we are trying to answer is what defines pose correctness or equality, as for different tasks the factors which are of critical importance vary.

Through the application of different dimensionality reduction techniques, we want to try to find in the data the attributes that give us the most important information about the pose, keeping them in our dataset. For this, it is important to explore and understand the data we are working with in order to apply the correct dimensionality reduction techniques.

Implementing dimensionality reduction algorithms in pose estimation and comparison tasks offers significant advantages. It enhances computational efficiency, improves accuracy by focusing on informative features, prevents overfitting, enables data visualization, and aids in interpretability and insight generation. By reducing the complexity of high-dimensional data, these techniques lead to faster processing times, reduced resource requirements, as well reducing the CO2 footprint, and improved model performance. Additionally, the visualization of reduced-dimensional data provides valuable insights, while uncovering underlying structures helps in understanding key factors influencing pose correctness or similarity.

1.1.1. Personal motivation

My personal motivation comes from a personal experience and passion for sports. I have practiced for many years different sports, from karate, to boxing or football and I have often wondered if computer vision could be used to help people to improve their movements and techniques without relying 100% on a trainer or coach [4]. By reducing the dependence on trainers or coaches, we could empower the people to learn and improve on their own movements and techniques, reducing the risk of injury during the physical activities. I hope we can contribute to the field of sports and computer vision, reducing injuries, which can seriously reduce people's quality life.

1.2. Objectives of the study

With this research we would like to achieve the next objectives:

General Objective:

The implementation of dimensionality reduction techniques into comparison problems is the main goal of our investigation. The implementation of these techniques allows us to reduce the complexity of the data and to keep in our dataset only the most important information, enabling the detection of the key dimension and improving the accuracy and efficiency of the existing pose estimation models. The approach we are presenting here has practical applications in several fields, like sports or rehabilitation [5], where accuracy and efficiency are crucial. By proving that dimensionality reduction has positive results when estimating a person's pose in terms of accuracy and efficiency, we can make great contributions to the field of computer vision.

Specific Objectives:

- Investigate the feasibility of implementing dimensionality reduction techniques in pose comparison problems in computer vision.
- Develop a deep learning-based model for pose comparison and evaluate its performance and accuracy.
- Identify the key dimensions in the dataset that enable accurate and efficient pose comparison.
- Identify certain specific conditions and correlations of a pose, within the selected body part.
- Explore and understand the data to be able to apply the correct dimensionality reduction technique.
- Contribute to the improvement and understanding of best techniques for pose comparison in computer vision.
- While this research focuses on the techniques to improve pose comparison, it does not seek to develop the best possible pose recognition model.

1.3. Impact on sustainability, ethical-social and diversity

This section aims to analyse the project's ethical and global impact in alignment with the United Nations' Sustainable Development Goals for 2030. The University of the Open University

(UOC) upholds a strong commitment to sustainability, social responsibility, ethical behaviour, and the promotion of human rights and diversity in both academia and professional practice.

The UOC recognizes the significance of sustainability across all its operations and endeavours. This entails minimizing the university's environmental footprint, advocating for renewable energy usage, and supporting initiatives that contribute to a sustainable future.

Furthermore, the UOC places great importance on social responsibility and ethical conduct. This involves considering the potential impacts of its actions on the local and global communities and adopting a responsible approach to research and decision-making.

In line with its dedication to human rights and diversity, the UOC ensures that every member of its community is treated with respect and dignity, while promoting equal access to opportunities and resources. The university takes an active role in addressing issues of inequality and discrimination, striving to foster an inclusive and diverse environment.

In summary, the UOC is committed to being a responsible and ethical global entity, leveraging its knowledge and expertise to contribute meaningfully and sustainably towards improving the world we live in.

Sustainability:

In addition to the commitments, the UOC acknowledges the importance of sustainability in machine learning projects. Such projects can potentially have adverse environmental impacts, and it is crucial to address these concerns.

Some of the key sustainability considerations in machine learning projects include:

- **Energy consumption:** Training and running machine learning models can require substantial amounts of electricity, contributing to greenhouse gas emissions.
- **Data storage**: The collection and storage of large datasets used in machine learning projects can consume significant energy and resources, potentially leading to e-waste and other environmental challenges if not managed responsibly.
- **Algorithmic bias**: If the training data used in machine learning models is biased, it can result in biased predictions and decisions, negatively affecting marginalized groups and the environment.
- **Unintended consequences:** Machine learning models may produce unintended outcomes that can be detrimental to the environment.

To address these concerns, it is essential to consider the environmental impacts of machine learning projects and take proactive measures to minimize any negative consequences. This includes adopting energy-efficient practices, implementing responsible data management strategies, addressing algorithmic biases, and actively monitoring and mitigating unintended consequences. By integrating sustainability considerations into machine learning projects, the UOC aims to ensure a more environmentally conscious and responsible approach to research and development.

In this project, we have discovered the potential of machine learning in assessing the correctness of poses during training or exercise, which can significantly reduce the risk of injuries and have wide-ranging economic and health implications. Moreover, machine learning techniques can positively impact not only the sports domain but also the medical field. By providing posture correction during rehabilitation exercises, machine learning can aid patients in their recovery from physical injuries.

Furthermore, we have conducted an analysis of the carbon emissions associated with our project. Based on the use of the library CODECARBON, our project's estimated energy and carbon footprint amounts is:

- CO2 footprint: 0.00022 kg CO2e.
- RAM Energy consumed: 0.000100 kWh.
- CPUs Energy consumed: 0.000744 kWh.
- Energy consumed since the start of the execution: 0.000844 kWh.

Although this is considered a minimum footprint, we are committed to further reducing it by utilizing energy-efficient equipment, incorporating renewable energy sources, and optimizing resource usage during training to minimize waste.

It is crucial to recognize the potential long-term impacts of a machine learning project. While the emissions linked to training and deploying a model may constitute only a fraction of the total emissions incurred throughout its lifespan, we acknowledge the importance of considering and mitigating these impacts. Understand as well that the dataset used for this research is small and other results should be expected with a bigger dataset, model or scenario.

In addition to addressing environmental concerns, we also emphasize the significance of social responsibility and ethical behaviour within the project. It is imperative to ensure that the development and deployment of machine learning models prioritize fairness, inclusivity, and respect for individual rights and privacy.

By integrating environmental sustainability, social responsibility, and ethical considerations, we strive to foster a comprehensive approach to machine learning research and development, benefiting both society and the environment.

Social responsibility and ethical behaviour:

Social responsibility and ethical behaviour are integral aspects of our research project. We conducted a thorough assessment of the potential societal impacts, carefully weighing the benefits in terms of advancements in fields like sports, medicine, rehabilitation among others. By considering the potential implications, we ensured that our project aligned with responsible and ethical practices. We were transparent in our approach, clearly outlining the goals, methods, and limitations of the project. Throughout the project, we have strived to align our actions with the principles of social responsibility and ethical correctness. However, we recognize that this is an ongoing process, and we remain dedicated to continuously evaluating and addressing the potential impacts of our research. Our unwavering commitment extends beyond the completion of this project, as we are resolved to always uphold responsible and ethical practices in our future endeavours.

Human rights and diversity:

In this project, our analysis focused solely on publicly available 3D Skeleton data, which ensures that the project is accessible to everyone and respects the principles of openness and inclusivity. It is crucial to acknowledge the existence of biases in the field of machine learning. Machine learning algorithms heavily rely on the data they learn from, and if the data is not representative of the target population, biases may arise based on factors such as gender, age, race, or religion. Therefore, when working with datasets, it is essential to exercise caution and critically evaluate their suitability to avoid potential misleading results. By being aware of these biases, we can work towards developing more equitable and unbiased machine learning models in the future.

1.4. Approach and method to follow

In this study, we aimed to understand the feasibility of implementing dimensionality reduction technique in pose comparison problems, using a combination of feature selection and feature extraction techniques and a deep learning model using a dataset with 3D Skeleton data. Introduction. The approach and methodology followed are going to be explained in this section.

The Fit3D dataset, which was generated by fit3D IMAR, was the dataset used. The dataset comprises frames capturing diverse movements performed by multiple subjects. Each frame provides the 3D skeleton coordinates of 25 joints, including the 17 Human3.6m joints.

During the last years, deep learning solutions have demonstrated to outperform classical computer vision methods in various tasks such as image segmentation and object detection [6].

Therefore, in this investigation, the CNN model is going to be the approach to use combined with dimensionality reduction techniques. Convolutional neural networks are the base of the state-of-the-art methods when designing architectures for object or human pose inference applications. Its local receptive fields and convolutional layers enable it to capture fine-grained details and recognize patterns and relationships between joints.

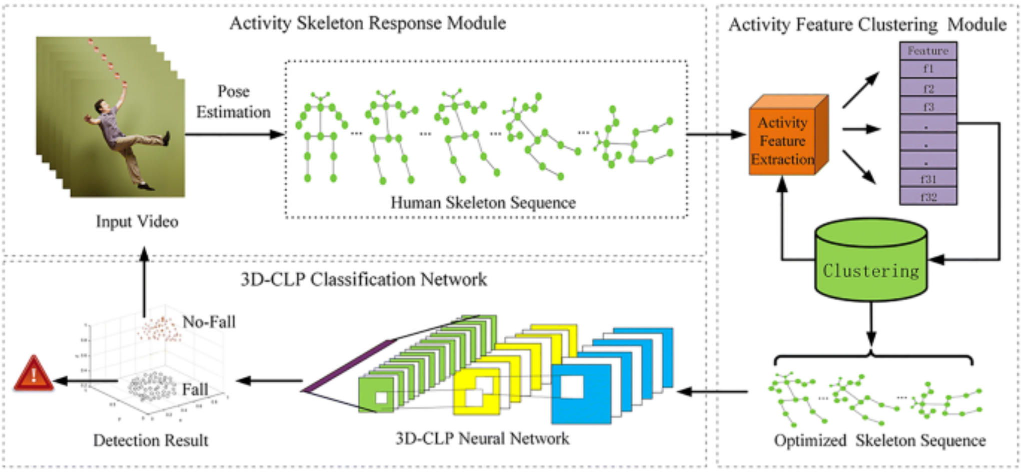


Figure 2. CNN model used in combination with s3CD [7].

Our specific focus lies in analysing the squat exercise among the various exercises available in the dataset. In order to understand the correctness of the squat, we will train two separate models for each type of dataset, one with all the dimensions and a second one with the reduced dimensions. These models will not only identify the relevant body parts crucial for categorizing a pose as correct but also discern specific conditions, correlations, or artifacts within the selected body part data that correspond to accurate execution of the movement.

The language program will be Python; therefore, we are going to set up the environment with the necessary libraries and tools.

1.4.1. Methodology

The methodology selected to carry this investigation is the CRISP-DM methodology. CRISP-DM is a data mining process model used commonly by data mining experts to tackle problems [8]. This methodology provides a structured approach for carrying out a data mining project. It is a robust, flexible and useful methodology, using analysis to solve problems. The model is a sequence of idealized phases but in truth, many of the phases can be solved in any order and very often it will be necessary to repeat many of them or backtrack to previous tasks in order to contrast improvements in the analysis process. Phases such as modelling, preparation of the data or understanding of the data will be usually repeated more than once as we have experience in this project. This methodology can be applied to any of the projects used regardless of the area to which it belongs, business, medicine, automotive, etc. It is possible to find several guides explaining this methodology online [9]. It breaks the process of data mining into six major phases:

Figure 3. Process diagram showing the relationship between the different phases of CRISP-DM

Business Understanding

In this initial phase, we are going to understand and define which ones are the business problems that need to be solved and the objectives, everything from a business perspective. With all this information, a data mining problem will be defined, and a plan designed to achieve the objective.

This first phase is important because if you do not understand the objectives of the research, you could invest a lot of time and effort in producing the correct answers for the incorrect questions.

Some of the tasks in this phase are:

- **Determine Business** Objectives: It seeks to describe the main objective and plan required from a business perspective, as well as addressing other types of questions or objectives. We are going to set the criteria used to determine if the project has succeeded or not, answering the questions specified in the beginning. Tasks of this phase:

    - Background.
    - Business Objectives.

- Business success criteria.

- **Assess Situations:** We will make an inventory of the necessary resources, as well as people, data, computer resources, and software. Here we list all the requirements of the project, including the complete schedule, the comprehensibility and quality of the required results. The limitations are listed too, and it can be the availability of resources, technological limitations such as the size of the dataset to model when being small, etc. Here we list also the risks that may delay the project or cause it to fail. Finally, in the case of having costs, a small analysis could be done to understand if it would be worth with the benefits that would be obtained. Tasks of this phase:

    - Inventory of Resources
    - Requirements
    - Assumptions and Constraints
    - Risks and Contingencies
    - Terminology
    - Costs and Benefits.

- **Determine Data Mining Goal**: We will determine the criteria to be able to say that the project has been successfully completed, and the results obtained have been as expected. Tasks of this phase:

    - Data Mining Goals
    - Data Mining Success Criteria.

- **Produce Project Plans:** The planned plan for achieving the objectives of the project is described. You must specify the tools considered in the project and the initial techniques to be implemented. The plan made initially for each stage of the project can be shown in a diagram. Tasks of this phase:

    - Project plan
    - Initial Assessment of Tools and Techniques.

Data Understanding

In this phase, we are going to start collecting data, which means we will get the data from internal or external sources and with specific characteristics including data volume, variety, formats and so on, as well as whether the data is in data files, in the cloud, it is life or stream.

We will identify data quality problems, discover insights into the data, detect if we have inconsistent in the data or duplicate values, the degree to which data is missing, etc.

Some of the tasks in this phase are:

- **Collect Initial Data**: Here we are going to perform the "Initial Data Collection Report" It will indicate where the data has been acquired, the source, the problems obtained to acquire them and the solution found to acquire them. This will help both the future execution of the project and the creation of similar projects.

- **Describe Data:** Here we are going to perform the "Data Description Report" The properties of the acquired data will be examined. Describing its format, the quantity, identities of the fields and evaluating if the data acquired satisfies our requirements.

- **Explore Data:** Here we are going to perform the "Data Exploration Report". In this step, we will address data mining questions, directly addressing project objectives, refining the description of data or quality reports and even transforming and preparing data for further analysis.

- **Verify Data Quality:** Here we are going to perform the "Data Quality Report". We are going to examine the quality of the data, if the data is complete or if we are missing values in the data.

Data Preparation

It seeks to prepare the data to be used by machine-learning algorithms. This process involves a number of tasks including cleaning, filtering, transformation, feature selection or dimensionality reduction. The objective is to improve the set of features eliminating variables which are not necessary and could produce noise or contribute to a poor precision of the algorithms. Feature Engineering is going to be performed here. Some of the tasks in this phase are:

- **Select Data:** Here we are going to perform the "Rationale for inclusion/exclusion criteria". We decide which data is going to be used for the analysis. These decisions are made based on the goals of our project, the type of data, etc. The selection of the attributes can be both rows and columns. If it is possible, it can be explained why this data is excluded and the reasons.

- **Clean Data:** Here we are going to perform the "Data Cleaning Report". We will raise the quality of the data to the technical level of the analysis that we have selected. To do this we will select subsets of the data, insert new values or apply techniques estimating missing values through modelling.

- **Construct Data:** By this, we mean construction operations such as the production of derived attributes or transformation of values to existing attributes. The derived attributes will be generated through one or more existing values. On the other hand, new attributes are going to be generated that may not be in our dataset, but after studying it we prove that they would be necessary to achieve our objectives. Tasks of this phase:
  - Derived attributes
  - Generated Records

- **Integrate Data:** Here we are going to perform the "Merged Data". When integrating data, it is sought to mix or merge information from different datasets or different databases. In this process, the aggregations are made, calculating new values summarizing information from different registers or tables.

- **Format Data:** Here we are going to perform the "Reformatted Data". We are going to seek the correct format to work with the machine learning algorithms.

Modelling

The development of a machine-learning model is based on the calibration and evaluation of different techniques to obtain the best performance.

These models are used to predict, analyse, or search for patterns, associations, or groups in our dataset. Different techniques can be used to solve a problem, but it depends on the type of data we are working, so it is recommended to study the dataset first to understand the data and the algorithms that can work properly with it. In case it is necessary we could return to the data preparation phase, to improve our dataset. We will encounter two types of machine-learning algorithms: supervised and unsupervised.

The tasks in this phase are the next:

- **Select Modelling Techniques:** In this section, we will select the techniques of modelling, and it is possible that many of them have changed with the techniques we were supposed to use in the beginning. We must explain the chosen modelling technique and the assumptions we have made to carry them out, for example, all elements must be numbered or that the elements must be normalized. Tasks of this phase:

- Modelling Technique
- Modelling Assumptions

- **Generate Test Design:** Here we are going to generate the Test Design. Before creating a model, we will have to prove the quality and the validity of the model.

In a situation like ours, where data mining is supervised to classify different models, certain patterns are used as measures of quality to indicate that a model works correctly or not. So, the dataset is divided into two parts, "test" which is the set to estimate the quality of the final algorithm and "Train", where we train and build our algorithm. We must describe the techniques we used and especially how we have performed this testing process.

- **Build Model:** We are going to run the modelling tools, creating one or more models. We will adjust the parameters for each algorithm, performing small tests checking which parameters work best in which case. It would be helpful to make a description of the resulting models, informing about the interpretation of the models and the problems that we had in order to implement them. Tasks of this phase:

  - Parameter Settings
  - Models
  - Models Description

- **Assess Model:** We must interpret the models with the knowledge that we have of that field and the criteria we have about data mining. After this, we have to discuss with experts in the field such as mathematicians or business analysts the results obtained.

We must classify the models and evaluate them according to the evaluation criteria. We must consider the goals and understand that it is usual to apply a technique more than once, obtaining results generated by different techniques. The results obtained will be evaluated, enumerating the qualities of the models generated and with relation to each other. Check that the algorithms have been tuned correctly and that the chosen parameters are indeed correct for that model. Tasks of this phase:

- Model Assessment
- Revised Parameter Settings

Evaluation

After developing a model, we will evaluate if the performance is the expected and we are predicting the classes correctly. For them, it is recommended to review the steps followed from the beginning. For example, in the case of predictions, we try to evaluate the number of elements that were predicted correctly. After this phase, we must express if the results are satisfactory and determine which models we will use. The tasks in this phase are the next:

- **Evaluate Results:** We will evaluate the results, checking the accuracy obtained from the models. We are going to summarize the results of the evaluation, stating whether our model finally met the specified objectives in principle or not. In case we agree with the results obtained, we will give the models as good, and we will approve them to use them in a real situation. In case the budget or time allows other test models with real data can also be evaluated. Tasks of this phase:

    - Assessment of Data Mining Results
    - Business Success Criteria
    - Approved Models.

- **Review Process:** We are going to do a review of the models to verify that we have not forgotten some important tasks. Questions will be asked to check if the process has been performed correctly. If it is possible, we have to remark which activities should be repeated or probably have been missed.

- **Determine Next Steps:** In this part, we can decide how to proceed with the results obtained. In case we have possible potential actions to implement, we should put pros and cons to implement them. Tasks of this phase:

    - List of Possible Action
    - Decisions.

Deployment

In the final step, if we are happy with the model we created, now we are going to deploy the models to run in different environments It does not mean that this one is the end of the project, even if we were just looking to improve the performance of the model, the knowledge we gained have to be organized and presented in a way the customers can understand and use. In many cases will be the customer the one which will carry out the deployment steps, and not the analyst. Tasks in this phase:

- **Plan Deployment:** At this point, it is where we will describe the procedure followed to create the model so that we can use it later.
- **Plan monitoring and maintenance:** Here we are going to deploy the monitoring and maintenance plan. It would be helpful to detail a data mining plan, with the process we are following.
- **Produce Final Report:** Finally, we are going to redact a report, explaining the experiences obtained and making a report. It could be a presentation of the data mining results to Tasks of this phase:

    - Final Report
    - Final Presentation.

- **Review Project:** In the Experience Documentation, we are going to explain what went right and what went wrong, summarizing the experience obtained in this project. At the same time would be helpful to explain any pitfalls we encountered, misleading approaches, etc. for future research.

With this introduction, I wanted to clarify some doubts with the terminologies in this Area and give an idea about the direction of this Research for those Students interested in this Field.

1.5. Planning

When it comes to a large piece of work like dissertation, managing the time well is essential to stay on track. For this reason, we define a list of milestones, tasks and risk analysis because is a great way to see what we need to do at each stage.

Milestones:

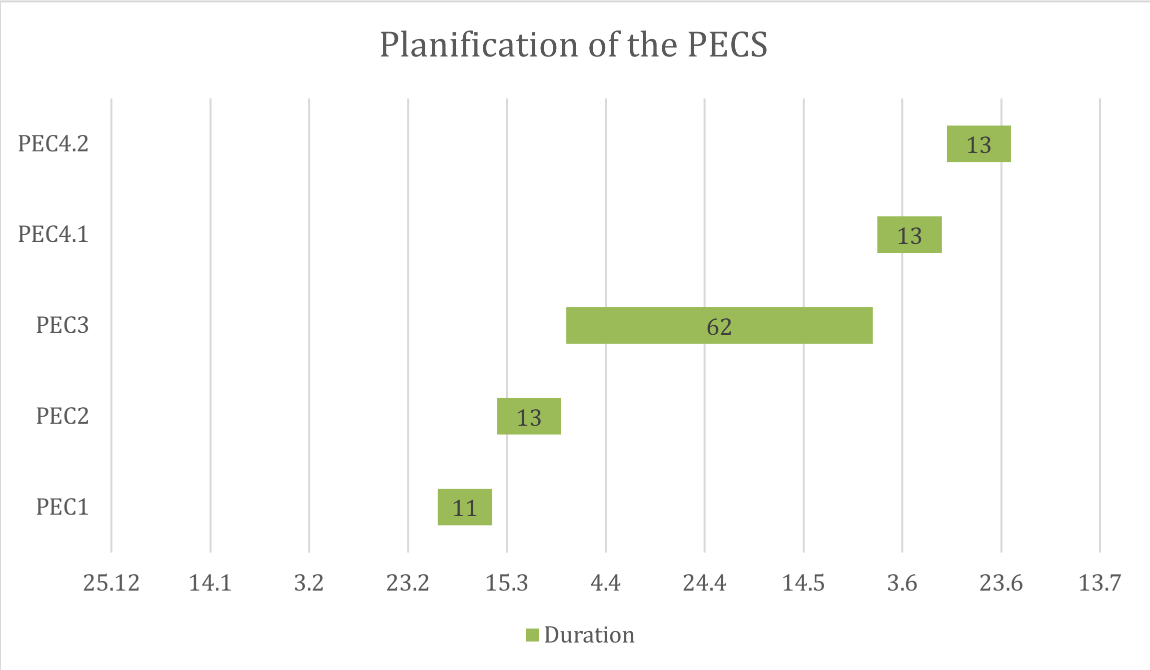


Figure 4. Planification of the pecs. The boxes represent the number of days during the time span of the project.

In figure 16 can see the number of days for each PEC and the start date and end date for each PEC.

Tasks:

Next, we are going to show the Gantt chart in figure 17 with the tasks that we plan to carry out to carry out this investigation. Dates may vary as we develop the project. The important tasks of the project are:

- Search for resources and read articles related to the investigation.
- Identifying an appropriate dataset for the research (labelled dataset)
- Creation of a deep learning-based model for pose-correctness evaluation
- Applying dimensionality reduction techniques to the selected dataset and train the model.
- Evaluating the performance of the proposed model using benchmark datasets.
- Development of the memory and preparation of the defence.

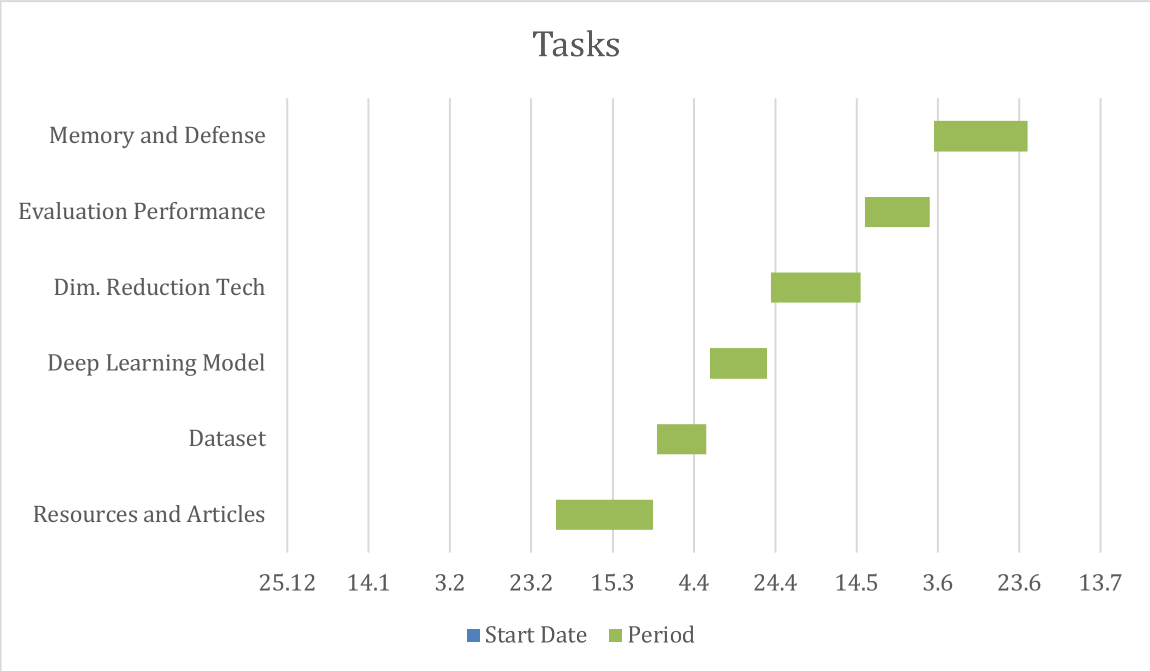


Figure 5. Tasks to be developed. The boxes represent the effort implemented for each part of the project.

Risk analysis:

A list of the risk that could affect the project:

DESCRIPTION OF THE RISK	SEVERITY	PROBABILITY	MIGRATION
Large scope of the project	High	Low	Although the project is large in scope, there are multiple stopping points. Once we achieve to implement dimensionality reduction in our data and train a model, it is sufficient to compare the pose results.
Inconclusive or poor results	High	Medium	We hope to contribute to this field by addressing this problem. The objective can still be achieved even if the model does not perform well. It is possible to find

			another dimensionality reduction techniques and train our model, improving the results
Time	High	Low	We want to use public available data, so we do not need to collect the data on our own. The team in Cambridge is going to support us to find a suitable dataset.
Lack of supervision	High	Low	Working closely with the supervisor of this project is key to answer doubts and questions. Weekly meetings and constant feedback from the supervisor.

1.6. Brief summary of products obtained

- **Product**: several functions were created in this code. These different functions could be implemented in other projects in the future. The code will be described in deep inside of the Annexes:

    - **process__json_files**: This function iterates over the folders in the data directory and extracts pose data from the JSON files. It calls the function **calculates_angles** and organizes the extracted data in separate lists.
    - **calculate_angles**: Calculate the angle (in degrees) between three points represented as vectors.
    - **interpolate_frames:** This function takes a DataFrame 'df' containing sequential frames and performs interpolation to generate a specified number of intermediate frames between each pair of consecutive frames.
    - **synth_data_process**: Process synthetic data from a dataframe of joint coordinates and calculate angles between consecutive joints.
    - **sequence_extractor**: Extracts sequences from a DataFrame, using the angles formed between the left hip, left knee and left ankle that represents the movement of the squats.

- **alignment_sequence:** Aligns and extracts sequences from a DataFrame based on provided indices.
- **get_sorted_indices:** Returns the sorted indices of a given list.
- **get_equally_spaced_elements**: Returns the indices of equally spaced elements from a given list
- **calculate_errors**: Calculates the errors between trainee and trainer coordinates and angles for each frame in the given DataFrame.
- **random_forest_calculator:** Calculates the feature importances using a Random Forest classifier.
- **select_features_by_cumulative_importance:** Selects features based on their cumulative importance scores.
- **anova_calculator:** Calculates the Analysis of Variance (ANOVA) for each column in the provided DataFrame.

- **Repository**: a repository with the code is created with the code, documents and it will be the main source of information for this research https://github.com/danielGOB/TFM/edit/main/README.md

- **Pipeline:** The strategy developed to implement dimensionality reduction successfully can be found on the document "Strategy implemented" in the github repository and it is recommended to read it while reading this research, to have a better understanding of the code. This pipeline can be implemented again in other projects.

1.7. Brief description of the chapters of the report

- **State of the art:** This section describes the various steps taken to carry out the analysis of this dissertation.
- **Experiment design:** This section contains all the phases followed during the development of this research.
- **Results:** This section shows the results obtained after the analysis as well as their discussions.
- **Discussion and future work:** This section describes the conclusions drawn from the research, critical reflection of the objectives achieved, as well as the planning and methodology implemented.
- **Glossary:** Alphabetized catalogue of words and expressions from one or more texts that are difficult to understand, together with their meaning or some comment.

- **Bibliography:** detailed list of all the sources consulted and cited in a research paper or project.
- **Annexes:** extra or complementary information that is included at the end of the thesis and that provides data whose purpose is the information about this research.

2. State of the Art

We will conduct a comprehensive analysis of the current state of the art in the field of computer vision and pose comparison, examining four areas I think are key to understand the background literature.

In this first part of the analysis, we are going to make a review of the current state of knowledge in the field of Computer Vision, with a specific focus on the next four areas: digital human avatars, biomechanics of the human body, correctness of a movement – pose and dimensionality reduction methodologies for pose comparison problems. Pose comparison is an important task in computer vision, that has significant applications in areas like sports, rehabilitation or human computer interaction. When we speak about pose comparison, we speak about the task of identifying the difference or similarity between two observed positions of bodies or objects. In particular, the comparison of human skeletal poses is useful for analysing movements.

In this context, dimensionality reduction methodologies have emerged as a promising method to reduce the complexity of pose comparison problems by identifying the key dimensions that define pose correctness.

2.1. Digital human avatars, representation and comparison

In recent years, digital human avatars helped us to understand better human movement through the creation of virtual representations of the human body. This provides us with a tool potentially accurate when representing the human body and its movements, which can be used to better understand and analyse poses and movements.

Several studies highlighted the potential of using human avatars to improve human movement analysis. For instance, researchers at Meta's Artificial Intelligence Research Lab and the University of Twente, have co-developed an open-source framework called "MyoSuite", which combines advanced musculoskeletal models with advance artificial intelligence to create "digital humans" that mimic complex movements. With this tool they can analyse the biomechanics of movement and evaluate the effectiveness of a robotic therapy beforehand, optimizing it for a patient and delivering a truly personalized and cost-effective treatment [10]. In sport science, digital human avatars have been used to analyse the movement in athletes and understand better potential injuries [11].

Several challenges and limitations must be faced at the moment of using digital human avatars. How we obtain accurately the movements and gestures is one of the main challenges. To achieve that, we need to use high-quality motion tracking techniques, as well as data analysis [12].

The use of digital human avatars for human analysis is an early non-invasive technology, that has the potential to help us understand the human movement better, with a lower cost and fully capturing its complexity. However, it is necessary to solve first the current limitations of this technology.

2.2. Biomechanics of the human body

Biomechanics of the human body is a critical area in the study of human movement. Biomechanics is the study of the mechanical principles of living organisms, especially the human body. It involves understanding how the various systems and structures in the body, such as muscles, bones, and joints, work together to produce movement. Biomechanical analysis can be used to understand and improve movement patterns in sports, physical therapy, and other areas.

The description of the biomechanics of the human body can be achieved through different approaches. For example, understand and capture the joint constraints of the human body using markerless motion capture techniques, that rely on computer vision algorithms, allow us to understand better the biomechanics of the anatomy of a person, improving the accuracy at the moment of stating a pose [13]. This first approach has the potential to be the best cost-effective and accessible option, but currently faces challenges in accurately capture the subtle nuances of human movement.

Another approach uses motion capture technologies, which help us to understand more accurately the biomechanics of the body but increments the effort when processing and analysing the data because of its complexity. Both approaches help us to understand better how to reduce the physical overload of the body during certain activities. Understanding which joints and muscles have a bigger impact is key, when developing better guidelines for certain activities and movements, helping us to prevent potential injuries [14].

There are several challenges and limitations in the field of biomechanics and movement analysis, including the need for more robust and accurate algorithms, better understanding of the complex interactions between the different components of the human body during movement or the lack of datasets with training data [15].

Understanding the biomechanics of the human body is crucial for developing accurate and robust computer vision systems for pose comparison, and ongoing investigations in this area have a strong potential to achieve promising results in the field.

2.3. Correctness of a movement - pose

When we analyse human movement, one of the toughest parts is to define a movement or a pose as correct. This can vary depending on the use case, such as sports performance, physical therapy or biomechanical research. We can find several investigations that attempt to define what constitutes a correct movement or pose based on biomechanical principles and the opinion from experts.

For example, one study by David W Meister and Amy L Ladd was trying to determine the biomechanical factors that may influence golf swing power generation [16]. This study identified biomechanical factors highly correlated to golf swing power generation. Similarly, a review by Kyle R Barnes and Andrew E Kilding try to identify and determine the biomechanical factors that determine or influence the Running Economy [17].

However, determining what constitutes a correct movement or pose can also involve subjective factors, such as aesthetics and cultural norms. A study done by Lurlynn Maharaj-Pariagsingh and Phaedra S. Mohammed presented a prototype intelligent dance tutoring system, DanceTutor, for coaching students. Evaluation done by three experienced professors from different countries of the prototype revealed the highly subjective nature and cultural biases of evaluating the quality of a dancer’s technique [18].

Ultimately, determining what constitutes a correct movement or pose requires a combination of objective biomechanical analysis and subjective expert opinion. Advances in computer vision and pose comparison techniques have enabled more precise and objective analysis of movement, but subjective factors cannot be ignored.

2.4. Dimensionality reduction techniques

Dimensionality reduction methods help us to analyse large datasets by reducing their dimensionality while retaining the most important information. For example, dimensionality reduction methods with motion data allow us to extract key features and patterns in the movement, retaining the most important information and providing the models with cleaner data. This data can improve the results of our models, giving more accurate results when comparing different movements, as the data used to feed them has less dimensions, therefore lower noise.

One of the most common dimensionality reduction techniques that have been used in the analysis of motion data for example is principal component analysis (PCA). PCA has been widely used for motion analysis, for example at the moment of estimate hand pose. Experiments demonstrate that the implementation of PCA methods achieves better accuracy in hand pose recovery compared to state-of-the-art baseline methods [19].

Other dimensionality reduction techniques that have been used in the analysis of motion data include linear discriminant analysis (LDA). Depending on the technique we use, the strengths and weaknesses of each method must be considered. For example, in the investigation of Nojun Kwak and Sang-II Choi, a modified version of LDA is used with regression problems to estimate head pose and compare the performance with other conventional extraction methods [20].

The use of dimensionality reduction methods aims at finding the transformation from the original feature space to a low dimensional subspace that retains most of the discriminative information. In this research we can find how different dimensionality reduction algorithms were applied when inferring 3D human poses from monocular videos, obtaining better results, proving the effectiveness of the proposed methods [21].

Overall, dimensionality reduction techniques are a powerful tool for analysing motion data and skeletal movement, allowing us more efficient processing and analysis of large datasets. While there are many different techniques available, each with its own strengths and weaknesses, researchers must carefully consider the specific requirements of their analysis when choosing the most appropriate technique.

2.5. Conclusion

In this state-of-the-art review, we have explored four areas: digital human avatars, biomechanics of the human body, what defines a movement - pose as correct on a conceptual level, and dimensionality reduction techniques applied to skeletons or motion data.

Digital human avatars have the potential to be powerful tools for analysing and improving human movement in fields such as sports science and physical therapy. However, there are still challenges and limitations to be addressed in the field, such as accurately capturing subtle movements and gestures.

The biomechanics of the human body provide a framework for understanding how different poses and movements are produced and can be described. This knowledge allows us to more accurately track and analyse movements for a wide range of applications. Defining what constitutes a "correct" movement or pose on a conceptual level is a challenging task, and one that may depend on specific use cases.

Finally, dimensionality reduction techniques can be applied to skeletal and motion data to reduce complexity, providing us with more efficient processing and analysis. This research has important implications for fields such as sports science and physical therapy, where large amounts of movement data must be processed and analysed. We expect to be able to understand how to implement dimensionality reduction algorithms on 3D Skeleton data and in the case, we obtain positive results, provide a strategy to be able to reduce the dimensions successfully.

Overall, the intersection between computer vision and human movement analysis is an evolving and exciting area of research, with a wide range of potential applications across multiple fields. Exploration and refine of these four areas are key to unlock even more possibilities for improving human movement and understanding the complexities of the human body.

With this knowledge, we are going to delve further in the experiment design section, implementing different techniques and tools, to successfully accomplish our objectives.

3. Experiment design

Two different experiments were performed, one for the dataset with the reduced dataset and one for the dataset with all the dimensions.

Furthermore, the obtained results for the accuracy and loss of each classification are displayed, along with a corresponding graphical representation. Additionally, a confusion matrix is provided for each pair of "Classification Model-Reduction algorithm."

Firstly, let's begin by describing the tools utilized in this research:

3.1. CRISP-DM

The next diagram in Figure 18 [22] explains briefly the procedure followed, according to the CRISP-DM methodology:

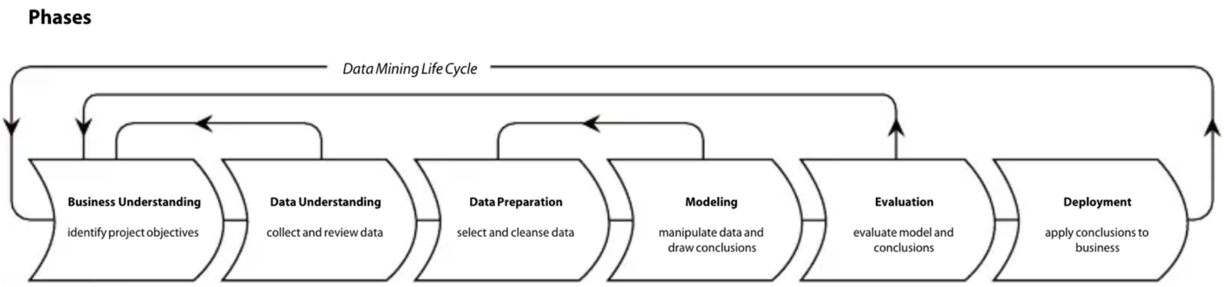


Figure 6. CRISP-DM phases followed to understand the experiment.

The upcoming phases to be implemented in this research are outlined, with a more detailed explanation provided in section 1.4.1 regarding the methodology.

3.2. Feature Engineering

A feature is an attribute that we use for modelling. A table is to be composed of columns and rows, where the rows are usually the observations of our study, and the columns are the features. A feature is a part of the attributes of an observation that will have a meaning for our objectives. They will be an essential part of solving our problem.

When we speak about feature engineering, we speak about the process to create features that makes machine algorithms work. It is quite important and will determine the success of our project, without matter your skills in statistical or computer techniques. A quotation I like is the next one:

"At the end of the day, some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used. If you have many independent features that each correlate well with the class, learning is easy. On the other hand, if the class is a very complex function of the features, you may not be able to learn it. Often, the raw data is not in a form that is amenable to learning, but you can construct features from it. This is typically where most of the effort in a machine learning project goes." [23]

It is the core of a data science project, and comes to life in the Data Preparation phase, covering the phases of selecting the data, cleaning the data, constructing the data and attribute selection.

This phase can be performed several times taking in some cases almost 70% of the time in a data mining project.

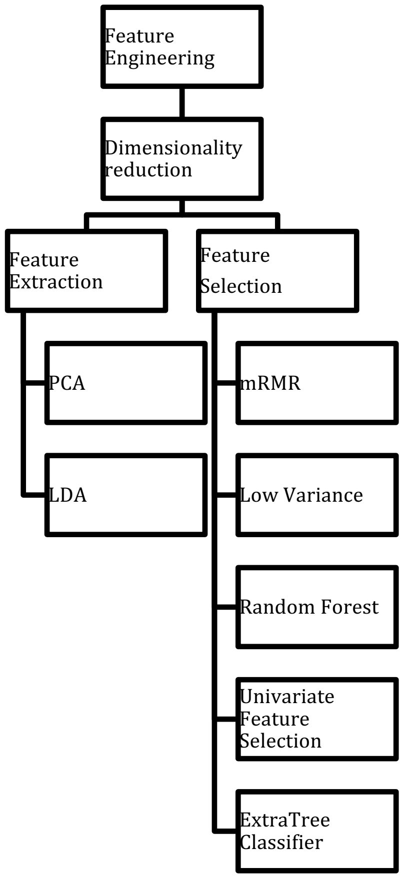


Figure 7. Feature Engineering diagram. It has some dimensionality reduction examples.

The features will directly influence the obtained results, so the better we choose and prepare the features, the results will be more accurate and effective.

3.2.1. Dimensionality Reduction

When we speak about Feature Engineering, we speak as well about Dimensionality Reduction as we can see in the figure 4. The core of this project focuses on the dimensionality reduction implementation. After applying the proper feature engineering techniques to the raw data, we use dimensionality reduction algorithms to reduce the number of input features or variables in the dataset. The high dimensionality of data can often lead to challenges such as increased computational complexity, overfitting, and difficulty in visualization.

Dimensionality reduction methods aim to transform the original dataset into a lower-dimensional representation while preserving important information or patterns present in the data.

We can appreciate some of these algorithms in Figure 4. The results obtained will depend on many factors, among which we have the metrics used to evaluate our results, the features used, and how we have prepared them. Therefore, we are going to need quality features, which correctly describe the inherent structures of our dataset.

Having quality features will give us flexibility, allowing us to even choose the wrong models and still get decent results. By having more flexibility, we will be allowed to use easier models, which allow us to run faster, being more understandable and maintainable. We will not even need to tune our models with the best parameters, but we will still get good results.

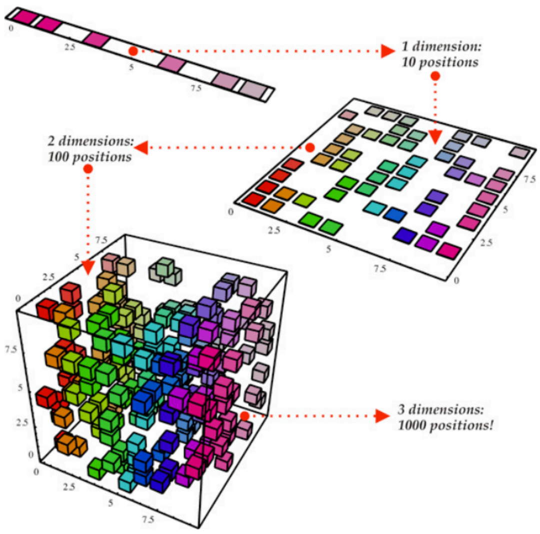


Figure 8. Image showing a briefly explanation of Dimensionality Reduction [24]. The data is reduced from 3 dimensions to 1 dimension.

In the figure 5, we can appreciate how several points in a 3D space is converted into a one-dimensional space. Even though we can see how we lose information in the process, dimensionality reduction algorithms try to maintain the maximum information as possible. When we speak about Dimensionality Reduction techniques, usually we speak about Feature Selection and Feature Extraction.

3.2.1.1. Feature Selection Algorithms

Feature selection is the process of reducing the number of input variables when developing a predictive model [25]. Statistical-based feature selection methods involve evaluating the relationship between each input variable and the target variable using statistics and selecting those input variables that have the strongest relationship with the target variable. The more that is known about the data type of a variable, the easier it is to choose an appropriate statistical measure for a filter-based feature selection method.

In figure 6 we can appreciate different feature selection methods that can be implemented depending in the input and output data we are working with. For example, a numerical output variable indicates a regression predictive modelling problem, and a categorical output variable indicates a classification predictive modelling problem.

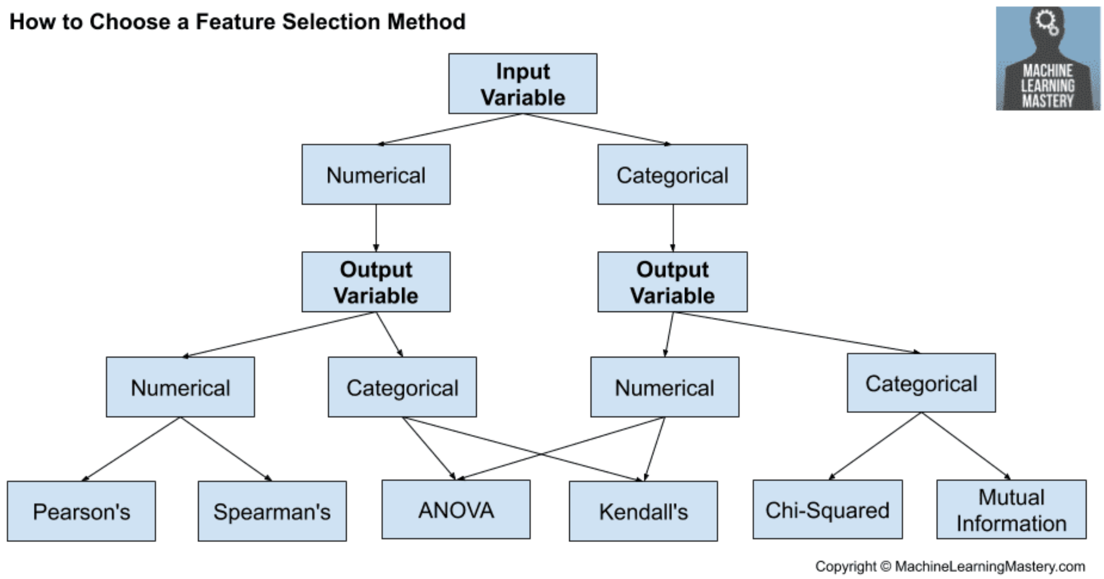


Figure 9. How to choose feature selection methods[25].

We are going to perform two different feature selection algorithms. They are going to select a subset of relevant features from the original dataset and create a new one with the selected features. We are going to delete redundant features trying to keep as much information as possible from our original dataset. Those algorithms are going to be quite effective when we are trying to understand the data.

- **Random Forest**: Random Forests are often used for feature selection [26]in a data science project and the reason is the tree-based strategies performed by random forests. Random Forests are a collection of decision trees classifiers which are fitted on various subsamples of the dataset and use averaging to improve the prediction accuracy. A Decision Tree is basically a set of decisions on whether or not classify something. Random Forest ranks how well the nodes improves the purity[27]. It means, that the Nodes with the greatest decrease in impurity happen at the beginning and the nodes with the least decrease in impurity at the end (Gini impurity).

The outcome of the Random Forest can be visualized by the “Gini Importance” and be used as an indicator of relevance between the features in the dataset. The feature importance scores can be normalized to sum up to 1 or scaled to a specific range, depending on the implementation. These scores help you understand which dimensions have the most significant impact on the model's predictions. Features with higher importance scores are considered more influential in determining the correct class, as they contribute more to the overall predictive power of the model.

By analysing the feature importance, you can gain insights into the underlying relationships between the input dimensions and the target variable. It allows you to focus on the most relevant features and potentially improve the model's performance by selecting a subset of informative dimensions or by prioritizing feature engineering efforts.

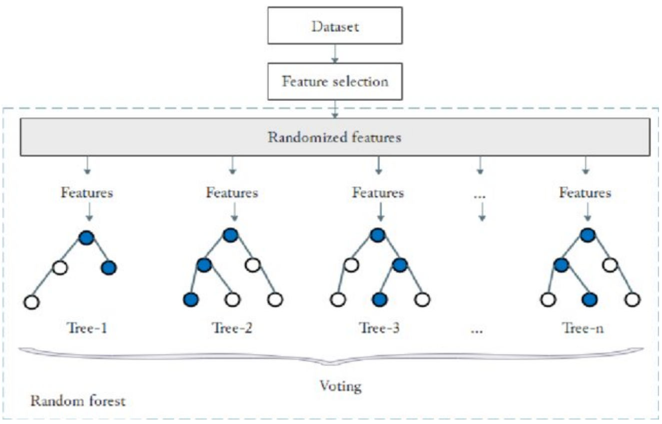


Figure 10. Random Forest selection of top features[28].

However, it's important to note that feature importance scores from tree-based models should be interpreted with caution. They reflect the model's internal decision-making process and may not capture complex interactions or dependencies between features.

- **ANOVA**: ANOVA (Analysis of Variance) is a statistical technique used to analyse the differences between group means. It determines whether the means of two or more groups are significantly different from each other by examining the variances within and between groups. ANOVA will help us identify which dimensions have a strong relationship with the target variable, indicating their importance for predicting correctness.

ANOVA is commonly used in experimental studies and can help identify the factors that contribute significantly to the variability in the data [29].

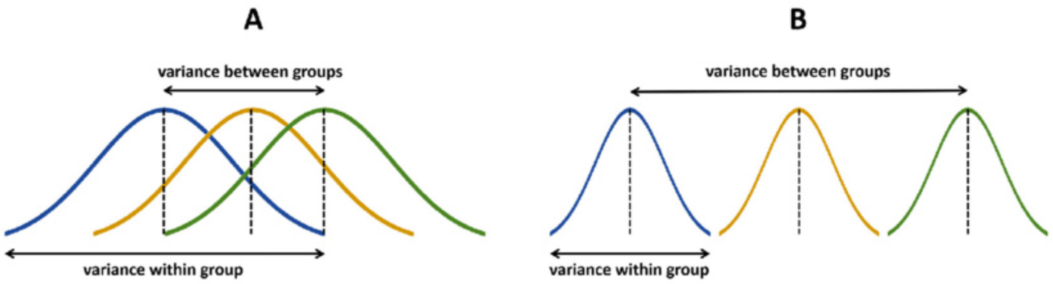


Figure 11. The graphic A shows us an example of a dimension with poor dependence on the target variable. The graphic B shows us a dimension highly dependent on the target variable. We can see the variance in the graph B allows us to identify clearly the groups.

Using the ANOVA test, we can compare more than two groups at once to see whether there is a correlation between them. The F statistic, or F-ratio, which is the outcome of the ANOVA formula, enables the examination of several sets of data to ascertain the variability within and across samples. By using this F-statistic score, each feature of the data can be ranked accordingly, and the features with higher ranks can be considered as the optimal set of features.

3.2.1.2. Feature Extraction Algorithms

Feature Extraction algorithms are a little bit more complex in comparison with feature selection, but at the same time are the proper ones if we are trying to discriminate values in a dataset with the intention of classifying. Feature Extraction involves reducing the number of amount of data used to describe a large dataset, which is one of the major problems performing analysis of complex data. Usually, as much bigger is the dimension of my dataset, the computation complexity and amount of memory increase. With Feature Extraction, we build combinations of the variables,

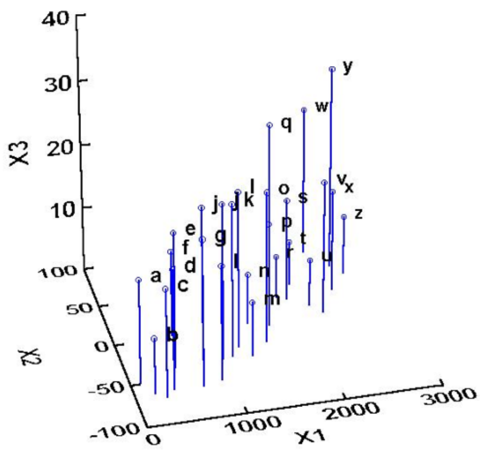


Figure 12. Some values in a 3-dimension graph. Blue lines represent distance of X3 axis.

solving the dimensionality problem, and still describing the data with sufficient accuracy.

- **PCA**: Principal component analysis is a method of extracting Principal component analysis important variables from a data set, identifying the different patterns in data. PCA aims to find the correlation between the variables in the data, searching the principal components of the dataset and transforming the data into a lower dimensional subspace. Most of the variance in your dataset would be explained by a small number of components, usually in 10% of the data can we find 95% of the variance.

  Generally speaking, we are going to search the bigger variance between our points. We try to project our data on a simple line, visualizing it like the shadow of the points. As we can see in the figure 9, when we don't cover properly the area, we are not going to maximize the variance on the projected dimension, therefore we have to look for the maximum variance and at the same time minimize the mean squared distance between the data and the projections, resulting in a line which is closest to the data.

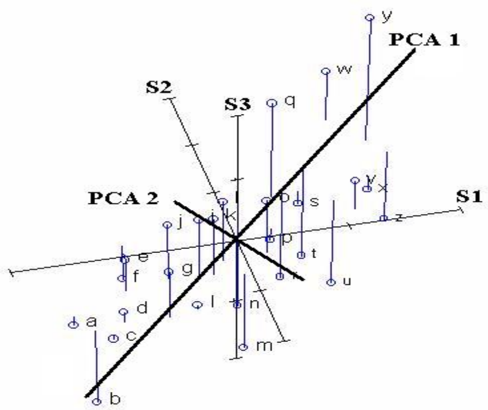


Figure 13. The same data is standardized and centered. We can appreciate the principal components PCA 1 and PCA 2. The axes S1, S2 and S3 represent the new axes.

  We would obtain like this what is called the first principal component, reducing everything in one dimension capturing the largest variance of the data. No other principal component can have a higher variability than the first principal component.

  The second principal component is also a linear combination of the predictors or points in our graph, taking the remaining bigger variance in our dataset, having a correlation between the first and the second principal component of 0. If the correlation is 0, then our PC have to be orthogonal.

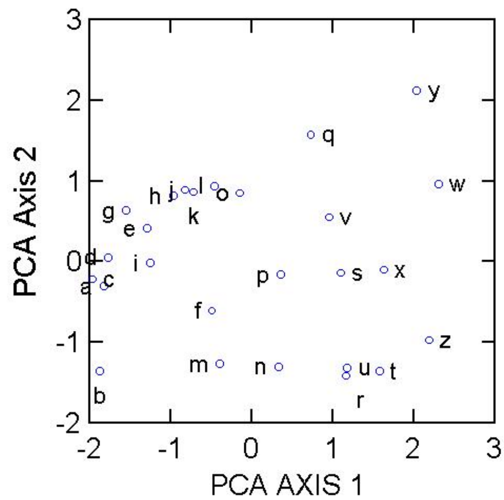


Figure 14. Data in the new subspace represented by the PCA axes.

  We can appreciate in the figures 9, 10 and 11 the process in a graphical way. These figures can help us to understand how do we convert a 3d dimensional problem in a 2-dimensional problem. The final axes are going to be represented by the PCA axes [30].

3.3. Dataset

The dataset utilized in this study is the Fit3D dataset [31], generated by fit3D IMAR. It consists of frames capturing a variety of sport movements performed by multiple subjects. Each frame includes the 3D coordinates of 25 joints, including the 17 Human3.6m joints.

Figure 15. Representation of a 3D Skeleton data over a subject from 4 different camera perspectives. We can see the segments connecting the different landmarks and each one of these segments starting or ending in one of the 25 different landmarks.

The dataset created from the fit3d Dataset treats mainly sport movements, focusing on the squats, but in the original dataset, we can also find other movements and types of data with information from other movements, videos from the movements, camaras information and types of data to store movement.

The original dataset has a json format and consists of frames capturing different movements performed by 8 different subjects. One trainer labelled as s08 and 7 trainees are represented in the dataset. Each frame provides the coordinates for 3D skeletons with 25 joints, including the 17 Human3.6m joints. The landmarks included in the dataset can be found in the figure 13:

0 - Hip	1 - Left Hip	2 - Left Knee	3 - Left Ankle	4 - Right Hip	5 - Right Knee
6 -Right Ankle	7 - Spine	8 - Neck / Upper Torso	9 - Head	10 - Head Top	11 - Left Shoulder

12 - Left Elbow	13 - Left Wrist	14 - Right Shoulder	15 - Right Elbow	16 - Right Wrist	17 - Left Toe
18 - Tip of Left Foot	19 - Right Toe	20 - Tip of Right Foot	21 - Left Thumb	22 - Tip of Left Hand	23 - Right Thumb
24 - Tip of Right Hand					

Figure 16. Joints in the json file, where the landmarks are labelled with numbers.

A representation of the 3D Skeleton landmarks and connections between landmarks are represented in the figure 14.

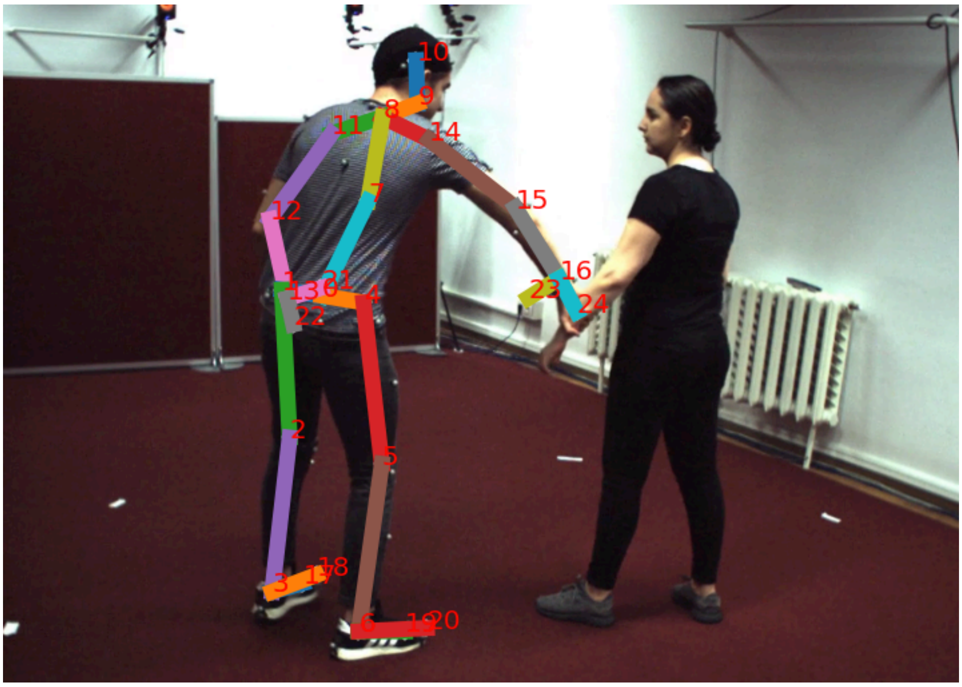


Figure 17. Representation of the 3D Skeleton landmarks and connection between landmarks on subjects [31].

3.4. Classification Model

In recent years, deep learning techniques such as recurrent neural networks (RNNs) and one-dimensional convolutional neural networks (1D-CNNs) have emerged as powerful tools for activity recognition tasks.

Originally designed for image classification, CNN models excel at learning intricate representations from two-dimensional inputs. However, the same concept can be applied to one-dimensional data sequences, such as acceleration and gyroscopic measurements used in human activity recognition. By employing 1D-CNNs, the model can effectively extract features from sequences of observations and map these internal representations to different activity types. Recently, a one-dimensional Convolutional Neural Network (1D CNN) has been suggested and carried out at the best performance levels in numerous applications, such as the classification of personalized biomedical data and time series classification [32].

In our specific case, we are focusing on classifying the correctness of frames within a sequence of movements. The CNN model can take advantage of the spatial information [33] encoded in the 3D coordinates, angles, and angle errors of the landmarks. The model can learn to recognize patterns and features that differentiate between correct and incorrect movements by analyzing the spatial relationships between the landmarks.

By training the CNN model on a labeled dataset where the correctness of each frame is specified, the model can learn to classify new frames based on their similarity to correct or incorrect examples. The CNN model can extract relevant features from the input data and make predictions based on those features, allowing it to determine the correctness of a given frame. The CNN model is designed to categorize each movement as correct, almost correct, or incorrect.

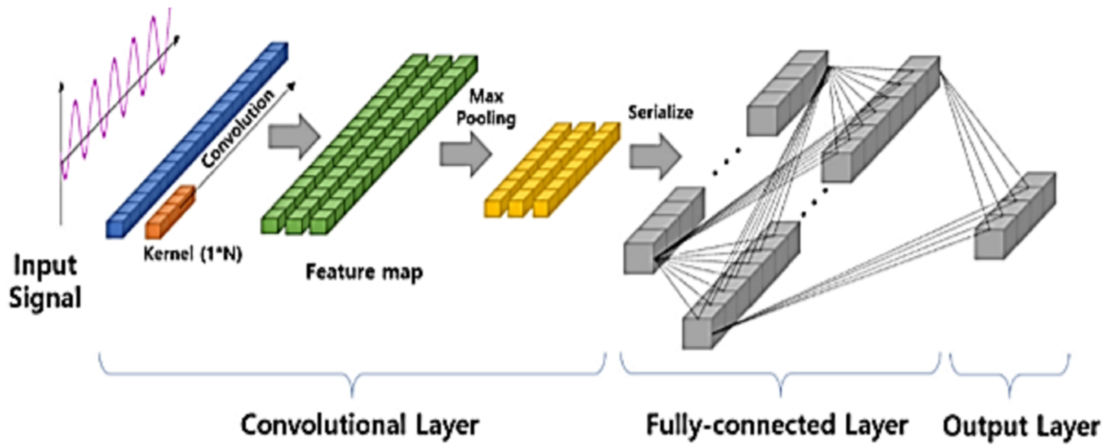


Figure 18. Architecture of a 1D CNN. We can appreciate that a numerical values is used as input [33].

In figure 15 we can appreciate that as input in the convolutional layer, a wave that could be represented as a numerical input is being feed into the CNN.

3.5. Implementation of the methodology

The methodology used is going to be implemented and all phases are going to be described in the upcoming sections.

3.5.1. Business Understanding and Data Understanding

In the phase Business Understanding, we determined as requirements would be needed a student with knowledge of the subject and a PhD student as cotutor of the Cyberhuman Lab from the University of Cambridge to carry out this project. We will also need a computer to work with a minimum of 16 Gb Ram and a core processor M2 Pro working on a macOs system.

As a programming language, Python 2.7.13 has been chosen and the necessary software is:

- Analytics platform: Anaconda 23.3.1 for 64-bit.
- Web development environment: Jupyter Lab.
- Libraries: scikit-learn, tensorflow, scipy, sklearn, codecarbon.
- Terminal: iTerm.

CRISP-DM was chosen as the methodology to be followed in this project, because it is a robust, flexible and useful methodology, using analysis to solve problems.

Speaking about the data, a dataset would be needed with the intention of applying dimensionality reduction techniques on the subject for classification pose problem for a given sport movement. There are three types of models for human body modelling [34]:

- **Kinematic model**: also called the skeleton-based model, is used for 2D and 3D pose estimation. This flexible and intuitive human body model includes a set of joint positions and limb orientations to represent the human body structure. Therefore, skeleton pose estimation models are used to capture the relations between different body parts. For this research a dataset with 3D Skeleton data is going to be used.
- **Planar model:** is used for 2D pose estimation. The planar models are used to represent the appearance and shape of a human body. Usually, body parts are represented by multiple rectangles approximating the human body contours.
- **Volumetric model:** which is used for 3D pose estimation. There exist multiple popular 3D human body models used for deep learning-based 3D human pose estimation for recovering 3D human mesh.

Since this kind of datasets are not common, it will be difficult to acquire it, spending a lot of time searching one with the quality and information necessary to analyse the data. Speaking of quality, we refer to the information of the data and the facility to understand what is expressed in the dataset. This at first delayed us quite a lot, since it was not easy for us to find the right one with the movements required from our supervisor, and when found, the information was not clear, and the creators of the dataset had to be contacted.

Another problem we encountered was the amount of data to analyse since it was not large enough, we had to use data mining techniques to perform acceptable algorithm training. The lack of larger datasets with this specific data was a big challenge and delayed the research critically.

At the beginning of this project, a diagram was given where we can find the different stages proposed for this project. It is expected that after finishing this project, a better understanding of the different dimensional reduction algorithms will be obtained.

During the first few weeks, the experiment was anchored in this point, since other datasets were initially considered but did not meet the objectives described in phase 1. Datasets like the SportsPose dataset [35], UCF101 dataset [36] or the MPII Dataset [37] were select at the beginning for consideration. Reason like the complexity to understand the content of the dataset or the lack of information, incomplete codebook, lack of 3D Skeleton data and low quality of the dataset were decisive to reject them and use the fit3D dataset. Therefore, it is recommended to have an agreement with an institution, which provides us the data, or if possible, in any other case, obtain such data by oneself.

Marker-based motion capture systems like Optitrack [38] accurately capture body motion by tracking markers attached to the body with millimetre precision [39]. However, they require initial setup and calibration, making them time-consuming and requiring expertise [40]. Additionally, markers can shift or detach during the capture process, adding complexity to the system setup and operation. On the other hand, markerless systems like OpenPose extract joint trajectories without the need for markers or sensors. They offer easier setup as they only require video input of the subject. However, markerless systems have limitations in terms of centimetre-level accuracy in landmark extraction. For instance, studies have reported approximately 3 centimetres of error when detecting joints of a running person [41]. Given the emphasis on precision in this research, an Optitrack system was used to generate the required data. However, due to the complexity and setup requirements of the Optitrack system, the data generated was not implemented in this study but reserved for future work.

3.5.2. Data Preparation and Modelling

For this research, a dataset is created and pre-processed from the original json data. A total of 105 dimensions dataset is created, containing the next dimensions:

- **Landmark coordinates**: landmarks described by the coordinates XYZ. In total 75 Dimensions, three for each one of the 25 landmarks.
- **Landmark angles**: angles between the joints connected in the human body. 25 dimensions are obtained between the landmarks.
- **Subject**: descriptor of the subject in our dataset. The subject number 08 is the trainer. The data from the trainer is used as the standard to compare in means of correctness. We have in total 8 subjects from our raw data. After processing the data, we obtain a total of 12 subjects.
- **Correctness**: the level of correctness related to that specific frame. Originally two level of correctness are defined. Finally, three levels are defined.
- **Sequence**: the number of the sequence for the whole squat movement. 5 sequences are extracted per subject.
- **Part**: The first part or second part of the sequence. Each sequence is divided in two parts.
- **Cord_errors**: the coordinates error for a specific frame in compare with the same frame for the trainer in the same sequence and part.
- **Angle_errors:** the angle error for a specific frame in compare with the same frame for the trainer in the same sequence and part.

After a detailed study of which one of the features would give us greater information and the implementation of the different dimensionality reduction techniques, the landmark angles features were selected to create the reduced dataset. The correctness dimension is going to be used as the target dimension, as it is the topic we would like to understand. We selected the squat movement because its complexity, as it involves upper and lower body parts in the movement.

To achieve this, a great deal of time was invested in applying different combinations of techniques to obtain all the necessary dimensions:

- **Angle calculation:** a function to calculate the angles between the joints of the 3D Skeleton for each frame in our dataset is developed. This function plays a pivotal role in our analysis as we leverage these angles to gain insights into the correctness of the movements and it is going to be used as base for the correctness calculation.

- **Sequence extractor:** a function is designed to extract sequences from a Dataframe column using the angles formed between the left hip, left knee and left ankle, which represent the main movement of the squat. It identifies the indices of the smallest local minima in the whole dataframe (index where the subject performs the lowest part of the squat), which are going to be 5 indices and taking this middle index as reference, returns the indices of the first local maxima before each middle index, and the index of the first local maxima after each middle index. This function is useful for detecting patterns or sequences in angle data, such as squats or other repetitive movements. It provides valuable insights from the movement data as it allows us to identify the start and end of a whole movement or sequence.

- **Aligning sequences:** In our dataset, we have multiple subjects performing movements at different velocities, in our dataframe we can appreciate it with the number of the rows or frames. To accurately compare and evaluate the correctness of these movements, we need to align the frames between subjects. To align the sequences, we determine the highest angle before and after the lowest angle frame for each sequence. These points mark the start and end frames of the sequence.

  To achieve consistent sequence lengths, we extract a fixed number of rows for each half sequence. By using the function **linspace**, we calculate the jump size, which determines the frames to be extracted. This approach guarantees that each half sequence will have the same length, regardless of the original number of frames. We repeat this alignment process for all subjects, with each sequence comprising 32 frames (16 frames per half sequence).

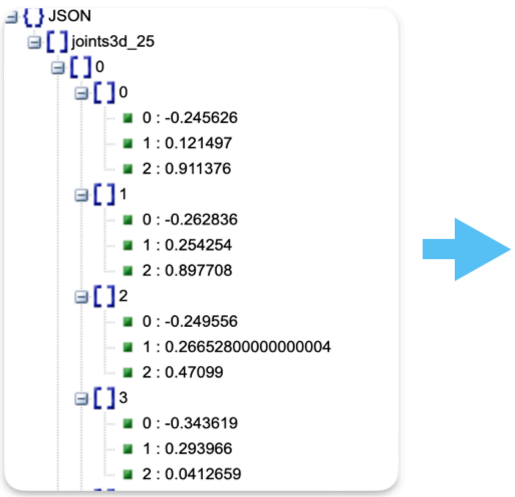
- **Angle and coordinate errors calculation**: to assess the accuracy of movements, we employ error metrics such as **Mean Joint Percentage Error (MJPE)**and **Mean Angle Percentage Error (MAPE)**. These metrics allow us to calculate the discrepancies between the trainer's movements and those of each trainee on a frame-by-frame basis. By comparing the errors, we can categorize each frame as correct, nearly correct, or incorrect based on predefined thresholds**.**

The whole strategy implemented to achieve successfully the goals described at the beginning, therefore this process is shown below.

3.5.2.1. Preliminary Steps

We need to transform the json file into a dataframe, using the functions ***process_json_files*** and ***calculate_angles.*** The raw data is composed by the frames of the movements, and each frame is composed by the 25 landmarks of the body. Each landmark has the three coordinates XYZ represented by the number 0 1 2. We assume that all the different elements are normalized and are numbers. In the process of the initial transformation, we classify the subject s08 (trainer) as correct and the rest of subjects as incorrect. It is the first approach, and we will deep in the next steps into the correct calculation of the correctness for the different subjects.

STEP 1



	Correctness	Subject	Hip_x	Hip_y	Hip_z	LHip_x
0	incorrect	s05	-0.165455	0.146322	0.980330	-0.163326
1	incorrect	s05	-0.165206	0.146329	0.980305	-0.163012
2	incorrect	s05	-0.164474	0.146397	0.980226	-0.162208
3	incorrect	s05	-0.163903	0.146443	0.980143	-0.161694
4	incorrect	s05	-0.163367	0.146452	0.980077	-0.161147
...	...	...	...	...	...	...
10821	incorrect	s07	-0.189601	0.146696	0.969931	-0.189115
10822	incorrect	s07	-0.189798	0.146373	0.969955	-0.189227
10823	incorrect	s07	-0.190014	0.145796	0.969952	-0.189380
10824	incorrect	s07	-0.190275	0.145322	0.969923	-0.189569
10825	incorrect	s07	-0.190376	0.144963	0.969926	-0.189633

Figure 19. Raw data transformed into a dataframe with own defined functions in the step1. The functions used to create synthetic data is mentioned between the dataframes.

With the dataframe obtained, we can verify whether the dataset is imbalanced.

3.5.2.1.1. Imbalanced dataset

The distribution of the subjects in the data is composed by 6 men and 2 women. While there is not information about the subjects in the dataset, we can appreciate watching the videos that the hight of the women is smaller in compare with the men subjects. The dataset exhibits an imbalance due to having only one trainer among of the eight subjects. This observation highlights the need to address the issue of class imbalance. Considering the trainer's movements as the reference for correctness, it is noteworthy that only 11.51% of the total rows are classified as correct in this initial analysis. In figure 29 a representation of the imbalance dataset can be appreciated.

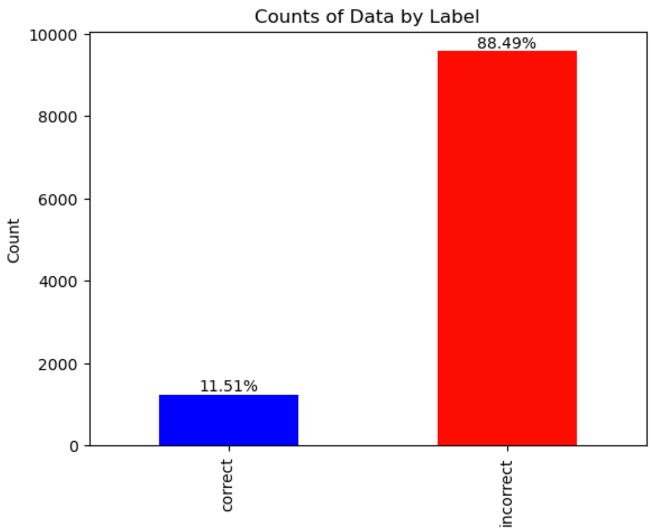
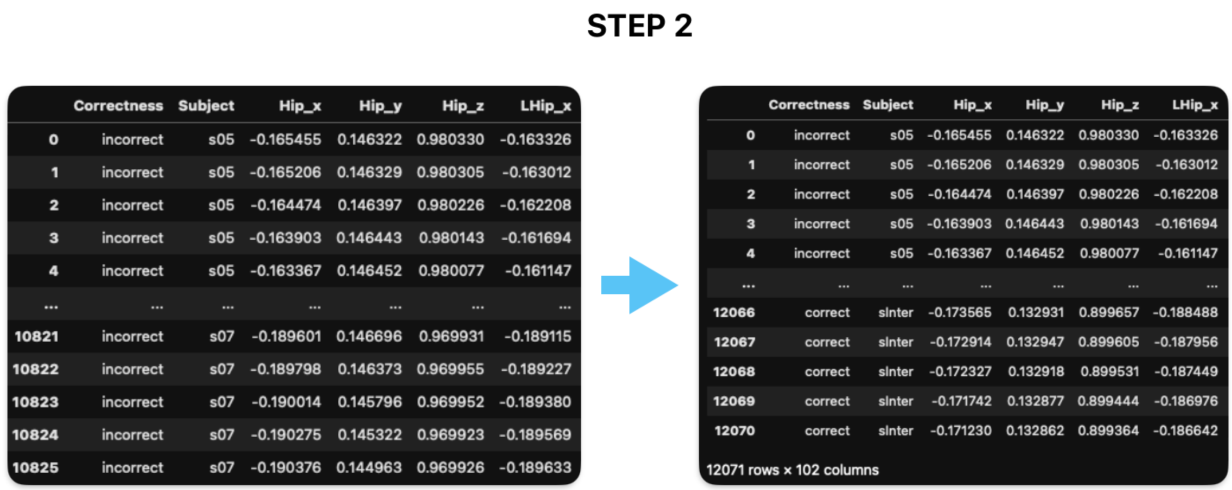


Figure 20. Imbalanced dataset. The Y axis contains the count of elements and the X axis contains the correctness label. In our first approach we considered only the labels correct and incorrect.

To tackle this challenge, first we are going to use interpolation, oversampling the minority class. These approaches aim to mitigate the impact of class imbalance and enhance the effectiveness of the subsequent analysis. In Figure 21 we can appreciate the transformation of the original dataframe into a second dataframe with synthetic data, labelled with the label “sInter”. For this process the functions ***interpolate_frames*** and ***synth_data_process*** where implemented.



	Correctness	Subject	Hip_x	Hip_y	Hip_z	LHip_x
0	incorrect	s05	-0.165455	0.146322	0.980330	-0.163326
1	incorrect	s05	-0.165206	0.146329	0.980305	-0.163012
2	incorrect	s05	-0.164474	0.146397	0.980226	-0.162208
3	incorrect	s05	-0.163903	0.146443	0.980143	-0.161694
4	incorrect	s05	-0.163367	0.146452	0.980077	-0.161147
...	...	...	...	...	...	...
10821	incorrect	s07	-0.189601	0.146696	0.969931	-0.189115
10822	incorrect	s07	-0.189798	0.146373	0.969955	-0.189227
10823	incorrect	s07	-0.190014	0.145796	0.969952	-0.189380
10824	incorrect	s07	-0.190275	0.145322	0.969923	-0.189569
10825	incorrect	s07	-0.190376	0.144963	0.969926	-0.189633

	Correctness	Subject	Hip_x	Hip_y	Hip_z	LHip_x
0	incorrect	s05	-0.165455	0.146322	0.980330	-0.163326
1	incorrect	s05	-0.165206	0.146329	0.980305	-0.163012
2	incorrect	s05	-0.164474	0.146397	0.980226	-0.162208
3	incorrect	s05	-0.163903	0.146443	0.980143	-0.161694
4	incorrect	s05	-0.163367	0.146452	0.980077	-0.161147
...	...	...	...	...	...	...
12066	correct	sInter	-0.173565	0.132931	0.899657	-0.188488
12067	correct	sInter	-0.172914	0.132947	0.899605	-0.187956
12068	correct	sInter	-0.172327	0.132918	0.899531	-0.187449
12069	correct	sInter	-0.171742	0.132877	0.899444	-0.186976
12070	correct	sInter	-0.171230	0.132862	0.899364	-0.186642

12071 rows × 102 columns

Figure 21. Dataframe treated with imbalanced data using own defined functions in the step 2. The functions used to create synthetic data is mentioned between the dataframes.

After the creation of synthetic data, the discriminated class was oversampled, obtaining a 20.64% of data for the correct data. Even though the dataset is still imbalanced, it is considered slightly imbalanced. We will continue working on this problem when we are able to understand the error, using the MJPE (Mean Joint Percentage Error) and MAPE (Mean Angle Percentage

Error). To understand how to calculate the error, first we need to understand the problem with the lengths of the videos. The problem with the sequence alignment and the errors will be explained in the next sections.

3.5.2.1.2. Sequence alignment

In our dataset, we have multiple subjects performing movements at different velocities, in our dataframe we can appreciate it with the number of the rows or frames. To accurately compare and evaluate the correctness of these movements, we need to align the frames between subjects. This alignment is crucial for calculating error metrics such as Mean Joint Percentage Error (MJPE) and Mean Angle Percentage Error (MAPE). Our strategy focuses on pre-processing the data, detecting sequences, and aligning them to ensure consistent frame comparisons. We can appreciate in this article from the Lund University other techniques that as well try to understand and extract segments of videos that define a specific movement [42].

We detect the sequences by identifying the moments when a subject performs a squat. By analysing the angles, we locate the frames with the lowest angles for the connections between the hip, knee, and ankle. These frames represent the lowest points of each squat sequence. As we can see in the figure 22, all the videos start with a subject standing straight, followed by a preparation phase and the start of the different sequences. After finishing the sequence, the subject leaves the bar on the floor and stands straight again.

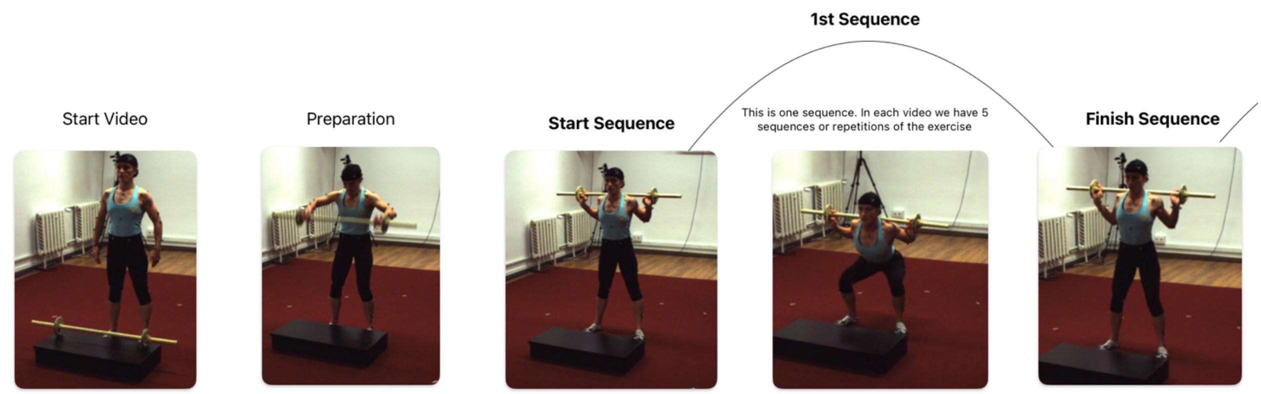
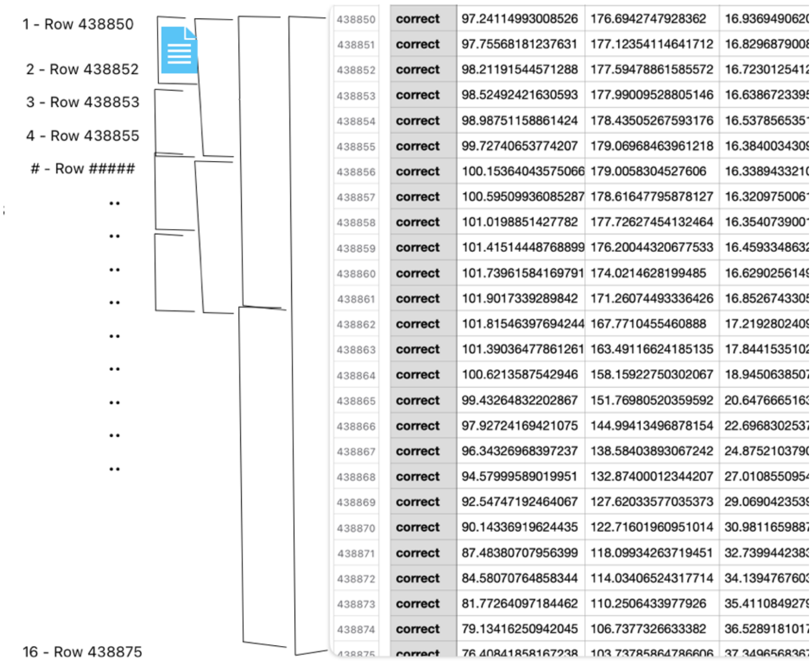


Figure 22. Frames of the video representing the start, the preparation, the start of the sequence, the lowest angle or lowest part of the squat and the end frame of the sequence [31].

To align the sequences, we determine the highest angle before and after the lowest angle frame for each sequence. These points mark the start and end frames of the sequence. However, since sequences may vary in length due to different subject velocities, we standardize the sequence length to ensure fair comparisons.

To achieve consistent sequence lengths, we extract a fixed number of rows for each half sequence. In our first approach we were extracting the frames automatically dividing each half sequence by 2 and dividing each division again until a total of three times. It means, we used to have 16 frames in total taking the maximum and minimum element for each division as we can see in the figure 23.



438850	correct	97.24114993008526	176.6942747928362	16.936949062
438851	correct	97.75568181237631	177.12354114641712	16.829687900
438852	correct	98.21191544571288	177.59478861585572	16.723012541
438853	correct	98.52492421630593	177.99009528805146	16.638672339
438854	correct	98.98751158861424	178.43505267593176	16.537856535
438855	correct	99.72740653774207	179.06968463961218	16.384003430
438856	correct	100.15364043575066	179.0058304527606	16.338943321
438857	correct	100.59509936085287	178.61647795878127	16.320975006
438858	correct	101.0198851427782	177.72627454132464	16.354073900
438859	correct	101.41514448768899	176.20044320677533	16.459334863
438860	correct	101.73961584169791	174.0214628199485	16.629025614
438861	correct	101.9017339289842	171.26074493336426	16.852674330
438862	correct	101.81546397694244	167.7710455460888	17.219280240
438863	correct	101.39036477861261	163.49116624185135	17.844153510
438864	correct	100.6213587542946	158.15922750302067	18.945063850
438865	correct	99.43264832202867	151.76980520359592	20.647666516
438866	correct	97.92724169421075	144.99413496878154	22.696830253
438867	correct	96.34326968397237	138.58403893067242	24.875210379
438868	correct	94.57999589019951	132.87400012344207	27.010855095
438869	correct	92.54747192464067	127.62033577035373	29.069042353
438870	correct	90.14336919624435	122.71601960951014	30.981165988
438871	correct	87.48380707956399	118.09934263719451	32.739944238
438872	correct	84.58070764858344	114.03406524317714	34.139476760
438873	correct	81.77264097184462	110.2506433977926	35.411084927
438874	correct	79.13416250942045	106.7377326633382	36.528918101
438875	correct	76.40841858167238	103.73785864786606	37.349656836

Figure 23. Example how the half sequence used to be divided three times, obtaining a total of 8 subsets of rows. We used to take in this first approach the maximum and minimum row for each subset, obtaining a total of 16 rows.

A second approach was developed using the function linspace, that calculates automatically the jump size, giving as entrance the number of frames, we would like to obtain. This approach guarantees that each half sequence will have the same length, regardless of the original number of frames. We repeat this alignment process for all subjects, including trainers and trainees. We aim to obtain five aligned sequences per subject, with each sequence comprising 32 frames (16 frames per half sequence as it was defined for the first approach). In figure 24 we can appreciate how the dataframe with the synthetic data is transformed into a new dataframe containing as well a numerical identifier for the sequence that a subject performs and whether is the firs of second part of that sequence (check figure 22). The function used for this process are ***sequence_extractor, alignment_sequence, get_sorted_indices*** and ***get_equally_spaced_elements***.

STEP 3

	Correctness	Subject	Hip_x	Hip_y	Hip_z	LHip_x
0	incorrect	s05	-0.165455	0.146322	0.980330	-0.163326
1	incorrect	s05	-0.165206	0.146329	0.980305	-0.163012
2	incorrect	s05	-0.164474	0.146397	0.980226	-0.162208
3	incorrect	s05	-0.163903	0.146443	0.980143	-0.161694
4	incorrect	s05	-0.163367	0.146452	0.980077	-0.161147
...	...	...	...	...	...	...
12066	correct	sInter	-0.173565	0.132931	0.899657	-0.188488
12067	correct	sInter	-0.172914	0.132947	0.899605	-0.187956
12068	correct	sInter	-0.172327	0.132918	0.899531	-0.187449
12069	correct	sInter	-0.171742	0.132877	0.899444	-0.186976
12070	correct	sInter	-0.171230	0.132862	0.899364	-0.186642

12071 rows × 102 columns

Spine_Neck_RShoulder	LShoulder_Neck_Head	RShoulder_Neck_Head	Sequence	Part
74.928603	113.058004	111.673278	5	first
75.041432	112.254176	111.874575	5	first
75.202370	111.839880	111.486567	5	first
75.358129	111.712705	111.135588	5	first
75.614261	111.292650	111.148646	5	first
...	...	...	...	...
42.678119	123.172558	108.592250	2	second
44.430849	118.687763	107.953147	2	second
46.160583	114.750802	106.138479	2	second

Figure 24. Aligned sequences calculated from the balanced dataframe using our owned defined fucnitons in the step 3. The functions used to perform the alignment are mentioned between the dataframes.

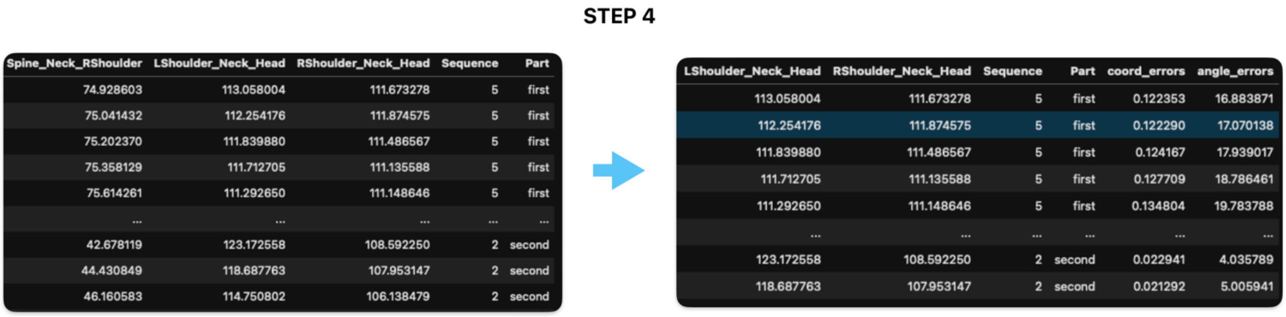
With the aligned sequences, we can now calculate the MJPE and MAPE by matching frames on a 1-to-1 basis between the trainer and each trainee at corresponding moments. These error metrics provide quantitative measures of the correctness of movements, allowing us to assess the accuracy of trainees compared to the trainer and to define different level of correctness of a movement based on the error.

3.5.2.1.3. Calculating Errors

To assess the accuracy of movements, we employ error metrics such as Mean Joint Percentage Error (MJPE) and Mean Angle Percentage Error (MAPE). These two terminologies are going to be explained next:

- **Mean Joint Percentage Error:** MJPE measures the average percentage error between the predicted joint coordinates and the coordinates from the trainer (subject s08) for each frame. It is calculated as the mean of the percentage errors for each joint.
- **Mean Angle Percentage Error:** MAPE measures the average percentage error between the predicted joint angles and the angles calculated for the trainer (subject s08) for each frame. It is calculated as the mean of the percentage errors for each joint angle.

These metrics allow us to calculate the discrepancies between the trainer's movements and those of each trainee on a frame-by-frame basis. By comparing the errors, we can categorize each frame as correct, nearly correct, or incorrect based on predefined thresholds. This classification provides insights into the accuracy of trainees' movements in relation to the trainer's, enabling us to evaluate their performance and identify areas for improvement.


STEP 4

Spine_Neck_RShoulder	LShoulder_Neck_Head	RShoulder_Neck_Head	Sequence	Part
74.928603	113.058004	111.673278	5	first
75.041432	112.254176	111.874575	5	first
75.202370	111.839880	111.486567	5	first
75.358129	111.712705	111.135588	5	first
75.614261	111.292650	111.148646	5	first
...	...	...	...	...
42.678119	123.172558	108.592250	2	second
44.430849	118.687763	107.953147	2	second
46.160583	114.750802	106.138479	2	second

LShoulder_Neck_Head	RShoulder_Neck_Head	Sequence	Part	coord_errors	angle_errors
113.058004	111.673278	5	first	0.122353	16.883871
112.254176	111.874575	5	first	0.122290	17.070138
111.839880	111.486567	5	first	0.124167	17.939017
111.712705	111.135588	5	first	0.127709	18.786461
111.292650	111.148646	5	first	0.134804	19.783788
...	...	...	...	...	...
123.172558	108.592250	2	second	0.022941	4.035789
118.687763	107.953147	2	second	0.021292	5.005941

Figure 25. Calculate errors and merge dataframes using our own defined functions in the step 4.

Now that we have obtained the necessary attributes in our updated dataframe, we can proceed with the classification of each frame based on the errors. In figure 25 we can appreciate how the dimensions cord_errors and angle_errors are calculated. These dimensions are key to calculate the correctness of the different frames based on this error.

To establish meaningful thresholds for classification, we will plot the errors per subject.

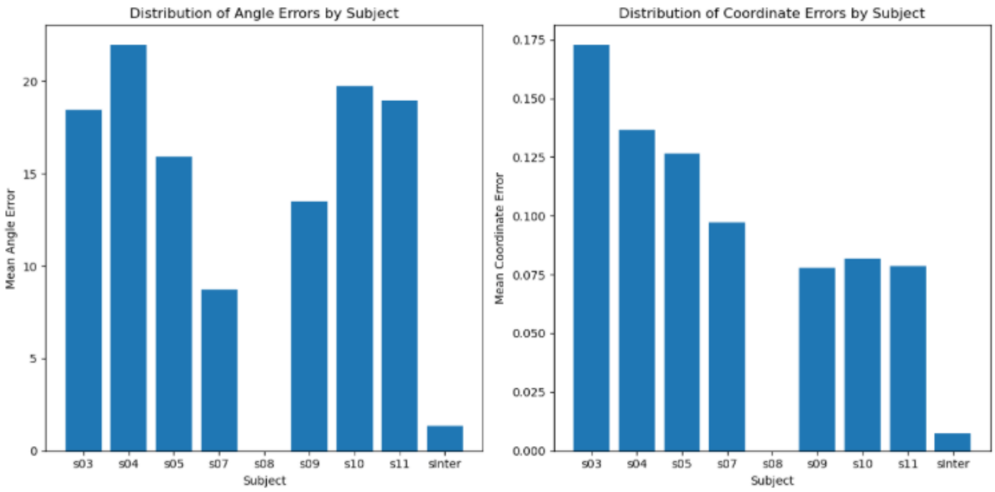


Figure 26. Distribution of Angle and Coordinate Errors per subject. We can see in the left side the mean angle error in degrees for all the subjects and in the right side the mean coordinate error expressed between 0 and 1.

After carefully evaluating the errors and their relevance to our analysis, we have come to the realization that relying solely on the coordinates may not be the most effective approach for determining the correctness of a movement in this research. Different approaches could be implemented normalising the bones for the 3D Skeleton, using the same scale for all the bones. But for this research that approach was not used as it was time consuming, therefore we decided to use the angles error as the reference used for the correctness in this study. Upon closer examination, we have observed that the synthetic subject created "Sinter", exhibits remarkably low angle errors in compared to the trainer (s08). Additionally, we have identified Subject S07 as the closest match to the correctness defined by the trainer, suggesting that their movements can also be classified as correct.

3.5.2.1.4. Production extra synthetic data

As we are lacking data and we can realize that the production of synthetic data is working properly, we are going to generate three more synthetic elements. Observing the graphic, we are going to take the s04, s05 and s07 because the difference in the error between them can be easily appreciated. This will help us, at the moment of testing our model.

LShoulder_Neck_Head	RShoulder_Neck_Head	Sequence	Part	coord_errors	angle_errors
113.058004	111.673278	5	first	0.122353	16.883871
112.254176	111.874575	5	first	0.122290	17.070138
111.839880	111.486567	5	first	0.124167	17.939017
111.712705	111.135588	5	first	0.127709	18.786461
111.292650	111.148646	5	first	0.134804	19.783788
...	...	...	...	...	...
123.172558	108.592250	2	second	0.022941	4.035789
118.687763	107.953147	2	second	0.021292	5.005941

	Correctness	Subject	Hip_x	Hip_y	Hip_z	LHip_x
0	incorrect	s05	-0.158593	0.150440	0.987652	-0.161762
1	incorrect	s05	-0.162997	0.150301	0.988584	-0.167643
2	incorrect	s05	-0.169160	0.148634	0.989643	-0.175345
3	incorrect	s05	-0.179233	0.144897	0.991757	-0.186243
4	incorrect	s05	-0.191349	0.140473	0.991918	-0.199481
...	...	...	...	...	...	...
1915	Correct	s07_syn	-0.210380	0.137472	0.964617	-0.220862
1916	Correct	s07_syn	-0.203112	0.138101	0.970562	-0.212875

Figure 27. Steps 2, 3 and 4 are implemented on S04, S05 and S07.

3.5.2.1.5. Update correctness

As we are trying to understand the correctness, we need to validate the and support our decisions with scientific literature, expert consultation in the field or published guidelines. In the research written by , the knee and trunk angle error are measured for different correct squats and the results obtained show an average error of 19.1 degrees for a knee joint angle error and 4.7 degrees for trunk error [43]. In light of these findings, we have defined our thresholds to classify the groups as follows:

- Angle error between 0 and 10: Correct
- Angle error between 10 and 20: Almost correct
- Angle error greater than 20: Incorrect

By incorporating angle errors into our classification criteria, we aim to provide a more comprehensive and accurate assessment of movement correctness.

We will now visualize the distribution of our data through a graph to assess if the error analysis has resulted in a more balanced dataset. The dataset has undergone necessary pre-processing steps to ensure its suitability for algorithmic analysis. Our goal is to examine if the distribution of different correctness groups has improved, indicating a more balanced representation of the data and enhancing the performance of the algorithms.

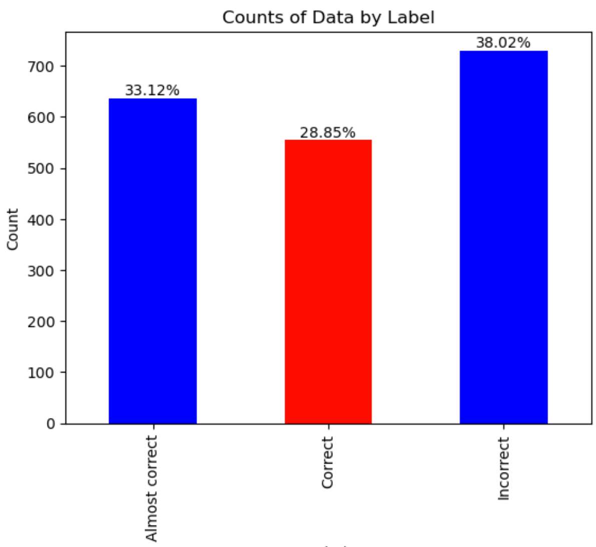


Figure 28. Correctness on the data updated. Now we can see that the new class "Almost correct" is as well represented. The three classes are represented.

3.5.3. Baseline Use Case – CNN model without dimensionality reduction data

Before proceeding feeding the models and implementing feature selection, it is important to define the criteria for determining the importance of dimensions. In this case, our objective is to identify the parts of the body that are important for the squat movement. Therefore, we are going to pre-process the data and exclude certain dimensions that are unlikely to provide us with this specific information.

- **Drop not necessary dimensions**: For example, dimensions such as "Subject" or "Errors" may not directly contribute identifying the joints or body parts involved in the squat. These dimensions are more related to other factors, such as the specific individuals performing the squat or the errors in the movement calculation. By dropping these dimensions, we can focus on the dimensions that directly represent the angles and coordinates of different movements, which are more likely to capture the information we need to understand the importance of different body parts in the squat.

- **Mapping correctness**: We need numerical data, therefore the correctness is going to be substituted for numbers.
- **Normalise dimensions**: In the case our data is not normalised, we are going to normalise it. Not normalising the data before training can cause problems in the networks, making drastically harder to train and decrease its learning speed.
- **Binary encode**: We are going to binary encode the target label.

We can see in the figure 26 the dataframe we are going to feed to the baseline CNN model.

	Correctness	Subject	Hip_x	Hip_y	Hip_z	LHip_x
0	incorrect	s05	-0.158593	0.150440	0.987652	-0.161762
1	incorrect	s05	-0.162997	0.150301	0.988584	-0.167643
2	incorrect	s05	-0.169160	0.148634	0.989643	-0.175345
3	incorrect	s05	-0.179233	0.144897	0.991757	-0.186243
4	incorrect	s05	-0.191349	0.140473	0.991918	-0.199481
...	...	...	...	...	...	...
1915	Correct	s07_syn	-0.210380	0.137472	0.964617	-0.220862
1916	Correct	s07_syn	-0.203112	0.138101	0.970562	-0.212875

Figure 29. The step 5 defines the actions taken on the dataframe before feeding the model CNN1.

In this first use case, a CNN model is going to be developed and trained to calculate the correctness with the whole dataset. CNNs are a type of artificial neural network that are primarily used for image recognition and processing, but they can also be used for other types of data, such as audio, video, and text. For example, you can use a CNN to classify text data, such as in natural language processing tasks like sentiment analysis or language translation.

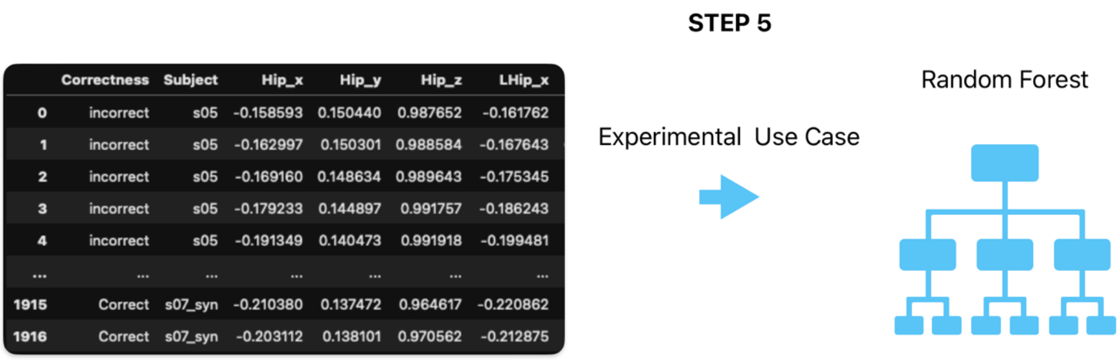
Additionally, CNN models can process and analyse time series data, such as sensor data or financial data. The key is to represent the non-image data in a way that the network can understand and process, such as converting audio to spectrograms or text to word embeddings.

After training the model several times, we have realised that after 200 epochs the accuracy of our model stops increasing, as well as our loss stops decreasing, therefore we have trained the CNN model for a total of 200 epochs. The results will be explained in the Results chapter.

3.5.4. Experimental Use Case – CNN model with dimensionality reduction data

For the experimental use case we are going to use dimensionality reduction techniques like feature selection and feature extraction. Random Forest and ANOVA are the feature selection methods selected for this research. The returned dimensions after implementing these techniques will show us the importance of the different dimensions and their impact on predicting the correctness of the target dimension. This analysis allows us to make informed decisions about feature selection and prioritize the dimensions that contribute the most to the prediction task.

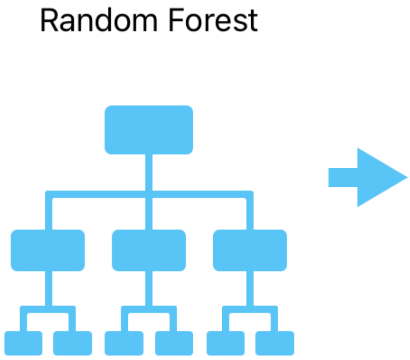
- **Feature selection**: The first technique implemented on the data is Random Forest. We will employ the feature importance scores obtained from tree-based models to select the dimensions that retain at least 80% of the importance. By training a decision tree on our dataset, we can extract these importance scores, which are tailored specifically for classification tasks. These scores provide insights into which dimensions carry the most information and have the strongest influence on predicting the correct class. The same pre-process steps are going to be implemented on the dataframe before feeding it to the random forest model.



	Correctness	Subject	Hip_x	Hip_y	Hip_z	LHip_x
0	incorrect	s05	-0.158593	0.150440	0.987652	-0.161762
1	incorrect	s05	-0.162997	0.150301	0.988584	-0.167643
2	incorrect	s05	-0.169160	0.148634	0.989643	-0.175345
3	incorrect	s05	-0.179233	0.144897	0.991757	-0.186243
4	incorrect	s05	-0.191349	0.140473	0.991918	-0.199481
...	...	...	...	...	...	...
1915	Correct	s07_syn	-0.210380	0.137472	0.964617	-0.220862
1916	Correct	s07_syn	-0.203112	0.138101	0.970562	-0.212875

Figure 30. Random Forest implemented and most relevant features selected.

In the figure 30 we can appreciate how the dataset obtained after applying feature engineering is going to be prepared to feed the random forest model. We are going to use the functions ***random_forest_calculator*** and ***select_features_cumulative_importance*** to obtain the desired results. In figure 31 we can see the results obtained by the random forest model.

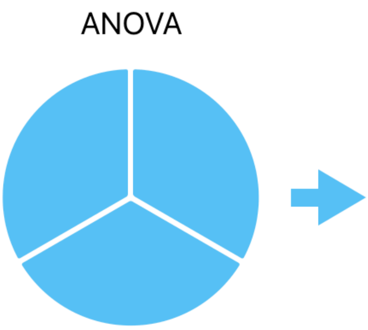



```
Feature Neck_Head_HTop: 0.0796
Feature LShoulder_LElbow_LWrist: 0.0739
Feature RShoulder_y: 0.0612
Feature LElbow_LWrist_TLHand: 0.0584
Feature RElbow_y: 0.0477
Feature LElbow_LWrist_LThumb: 0.0397
Feature RHip_Hip_Spine: 0.0381
Feature Spine_Neck_RShoulder: 0.0332
Feature RElbow_RWrist_TRHand: 0.0329
Feature RAnkle_y: 0.0319
Feature Spine_Neck_LShoulder: 0.0302
Feature RElbow_RWrist_RThumb: 0.0226
Feature RShoulder_RElbow_RWrist: 0.0223
Feature Neck_LShoulder_LElbow: 0.0215
Feature LShoulder_Neck_Head: 0.0185
Feature LShoulder_y: 0.0177
Feature Spine_x: 0.0162
Feature RToe_z: 0.0143
Feature RToe_y: 0.0124
Feature TipRHand_x: 0.0124
Feature RShoulder_Neck_Head: 0.0115
Feature Spine_y: 0.011
Feature TipRToe_x: 0.0084
Feature TipRToe_y: 0.0084
Feature RKnee_x: 0.0083
```

Figure 31. Dimensions obtained with random forest. We can see the feature importance near to the feature.

The results reveal that the top features predominantly consist of angle measurements, specifically those related to the joints in the upper body. Notably, the angles involving the arms, shoulders, upper and lower back, and the connection between the neck, head, and head top exhibit significant importance in determining the correctness of the squat. We are going to retain 75% of the features importance related to our target label, dropping the other 25%.

The same process is going to be repeated for ANOVA, to understand if there are significant differences in the means of certain dimensions across different levels of the "Correctness" variable. To obtain the desired results we have used the function ***anova_calculator***.



	F-statistic	p-value (ANOVA)
Spine_Neck_RShoulder	3454.137374	0.000000e+00
RElbow_y	3347.873410	0.000000e+00
LElbow_LWrist_TLHand	3058.857632	0.000000e+00
RShoulder_y	2951.577741	0.000000e+00
Spine_Neck_LShoulder	2500.383136	0.000000e+00
LShoulder_Neck_Head	2279.825000	0.000000e+00
RElbow_RWrist_TRHand	2151.336784	0.000000e+00
LShoulder_LElbow_LWrist	2137.804976	0.000000e+00
Neck_Head_HTop	1648.919745	0.000000e+00
LShoulder_y	1636.057053	0.000000e+00
LElbow_y	1262.373747	0.000000e+00
LHip_Hip_Spine	1258.089616	0.000000e+00
RHip_Hip_Spine	1221.204064	0.000000e+00
RShoulder_RElbow_RWrist	1122.697053	1.976263e-323
RShoulder_Neck_Head	1085.848091	4.889060e-316
RToe_z	999.326334	4.903503e-298
LAnkle_z	964.662786	1.338640e-290

Figure 32. Anova implemented and most relevant features selected.

In figure 32 we can see the results sorted in descending order of the F-statistic and ascending order of the p-value to prioritize dimensions with larger F-statistics and lower p-values, indicating stronger relationships and statistical significance. The sorted results are printed, showing the dimensions with their corresponding F-statistics and p-values. We are going to retain the top 25 % of the elements obtained.

The ANOVA results reinforce the significance of angles in determining the correctness of movements. The top features obtained from ANOVA, similar to the previous feature importance results, predominantly consist of angles. This consistency indicates that angles play a crucial role in identifying the correctness of movements. By extracting the common features from both lists, we can further explore and analyse their impact on determining movement correctness.

```
['LShoulder_Neck_Head',
 'LElbow_LWrist_LThumb',
 'RElbow_RWrist_RThumb',
 'RShoulder_RElbow_RWrist',
 'Spine_Neck_LShoulder',
 'LShoulder_LElbow_LWrist',
 'LElbow_LWrist_TLHand',
 'RHip_Hip_Spine',
 'RShoulder_Neck_Head',
 'Spine_Neck_RShoulder',
 'Neck_LShoulder_LElbow',
 'Neck_Head_HTop',
 'RElbow_RWrist_TRHand',
 'LHip_Hip_Spine']
```

Figure 33. Top elements from both resulst are selected.

In figure 33 we can see that the results indicate that the angles play a crucial role in understanding the correctness of the movement. The top features are primarily concentrated in the upper body region, including the shoulders, elbows, wrists, and the connections between them. Additionally, the angles related to the spine and hip are also significant. Notably, the coordinates of the body parts align with the angles, emphasizing the importance of the top part of the body and spine, with a notable exception for the right toe. A total of 15 Dimensions is selected from all the angles we had at the beginning. Therefore, we will focus on selecting the top angles from the ANOVA and Random Forest results to proceed with the PCA analysis.

- **Feature extraction:** Once we have obtained the dataframe with the selected features from the feature selection process, the next step is to normalize the data. Normalization is important to ensure that all the features are on a similar scale and have a similar range of values. Normalization can be achieved by applying a scaling technique such as standardization. Standardization transforms the data to have a mean of 0 and a standard deviation of 1.

After normalizing the data, we can proceed applying PCA to generate the synthetic dimensions. In our case, using PCA we generate three principal components, PC1, PC2 and PC3. These components will be added to the reduced feature set, resulting in an expanded feature matrix. Each sample in the dataset will now be represented by the original features as well as the values in the three PCA dimensions. These new features can be seen in the figure 34.

Dimensions from FS

```
['LShoulder_Neck_Head',
 'LElbow_LWrist_LThumb',
 'RElbow_RWrist_RThumb',
 'RShoulder_RElbow_RWrist',
 'Spine_Neck_LShoulder',
 'LShoulder_LElbow_LWrist',
 'LElbow_LWrist_TLHand',
 'RHip_Hip_Spine',
 'RShoulder_Neck_Head',
 'Spine_Neck_RShoulder',
 'Neck_LShoulder_LElbow',
 'Neck_Head_HTop',
 'RElbow_RWrist_TRHand',
 'LHip_Hip_Spine']
```

houlder_Neck_Head	RShoulder_Neck_Head	Correctness	PC1	PC2	PC3
-0.594645	0.120149	1	-1.637416	-0.396913	-1.835980
-0.671973	0.145270	1	-1.658402	-0.432989	-1.871501
-0.711828	0.096849	3	-1.677317	-0.421708	-1.894715
-0.724062	0.053048	3	-1.668873	-0.364496	-1.864247
-0.764471	0.054677	3	-1.697594	-0.319843	-1.793487
...	...	...	...	...	...
-0.298429	-0.174382	1	1.297534	-1.415876	-0.852311
-0.391985	-0.326618	1	0.779540	-1.470239	-1.189048

Figure 34. PCA dimensions and feature selection subset is merged.

By observing the figure 35, we can see the different PC dimensions in a 3D Scatter graph, where the classes can be differentiated with colours. We can make some observations about the separation between the classes:

- There is some overlap between the "Correct" and "Almost correct" classes, as we can see areas where the green and blue points are close to each other. This suggests that these two classes share some similarities in their PCA dimensions.
- The "Correct" class appears to be more distinct and separate from the other two classes. We can see regions where the red points are relatively isolated from the green and blue points.
- The distribution and arrangement of the data points in the 3D space provide insights into the relationship between the PCA dimensions and the correctness labels. It appears that certain combinations of PCA values are more associated with a specific correctness label.

We can realize that the three PC dimensions, created from the feature extraction, help us to understand and recognize correctness. These PC dimensions are going to be added to the angles dataframe with the selected features and will be used to train the CNN model.

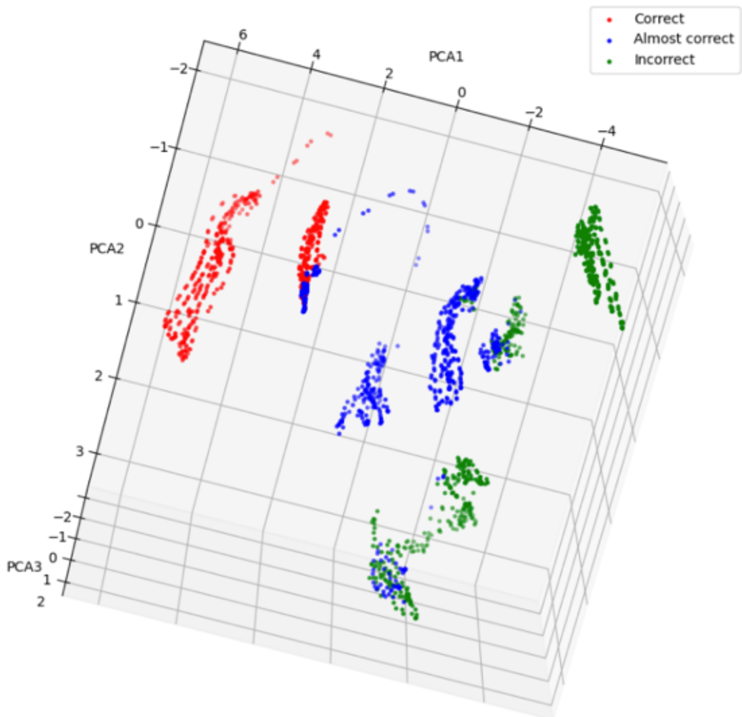


Figure 35. PCA dimensions showing the separation between classes. The different classes can be appreciated clearly.

A second model is going to be trained to process the information with the new dataset and the results will be compared with the results obtained with the results from the first model. The normalized data is going to be used to train our model as we did in the first part of the research. The results from the CNN models are going to be explained and analysed in the chapter results. In figure 36 we can see the input data for the experimental use case (the correctness label is going to be hot encoded).

Shoulder_Neck_Head	RShoulder_Neck_Head	Correctness	PC1	PC2	PC3
-0.594645	0.120149	1	-1.637416	-0.396913	-1.835980
-0.671973	0.145270	1	-1.658402	-0.432989	-1.871501
-0.711828	0.096849	3	-1.677317	-0.421708	-1.894715
-0.724062	0.053048	3	-1.668873	-0.364496	-1.864247
-0.764471	0.054677	3	-1.697594	-0.319843	-1.793487
...	...	...	...	...	...
-0.298429	-0.174382	1	1.297534	-1.415876	-0.852311
-0.391985	-0.326618	1	0.779540	-1.470239	-1.189048

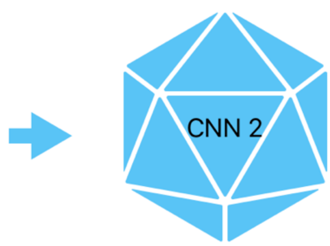


Figure 36. Trained second model with reduced data.

3.5.5. Deployment

This phase doesn't apply to this research, as we were studying the feasibility on implementing dimensionality reduction algorithm over 3D Skeleton data and no models are going to be implemented in different environments.

4. Results

This block will focus on explaining the results obtained in this experiment for each of the two use cases implemented. As it is a classification problem the appropriate techniques to measure are the Accuracy, F1-score and Confusion Matrix.

- **Accuracy**: measures how often the classifier correctly predicts. We can define accuracy as the ratio of the number of correct predictions and the total number of predictions.

- **F1 Score**: explains how many of the actual positive cases we were able to predict correctly with our model. It is a useful metric in cases where False Negative is of higher concern than False Positive. It is important in medical cases where it doesn't matter whether we raise a false alarm, but the actual positive cases should not go undetected.

- **Confusion Matrix**: Confusion Matrix is a performance measurement for the machine learning classification problems where the output can be two or more classes. It is a table with combinations of predicted and actual values.

- Loss: it can be seen as a **distance** between the true values of the problem and the values predicted by the model. Greater the loss is, bigger is the errors the model made on the data.

4.1. Baseline Use Case: No Dimensionality Reduction Applied

As we need to compare our results to measure the performance, the first situation we are going to find is the base case without performing any Dimensionality Reduction algorithms. The CNN model used for the base case is a simple model, composed by a convolutional layer, global max pooling and two dense layers. The data was split following a 20% of total data for validation and 80% for train.

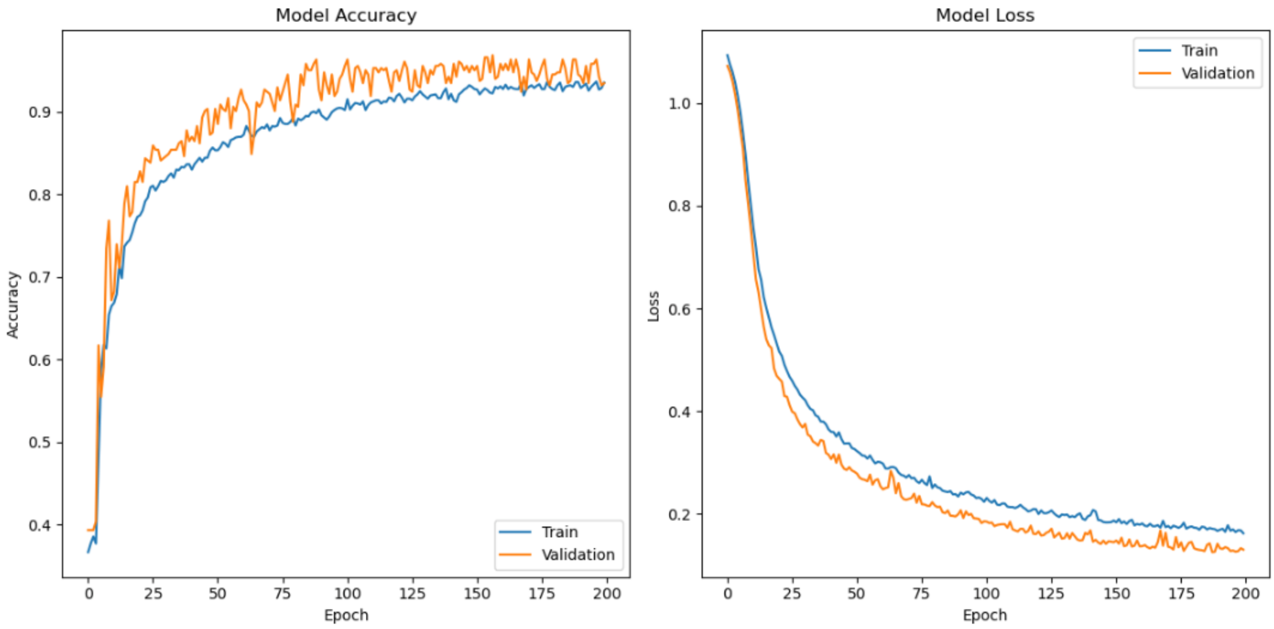


Figure 37. Accuracy and loss for the base use case

In figure 37 we can appreciate the results for the model 1 trained with the dataset with all the dimensions. Our model needed 7.83 seconds to train and validate with the whole dataset. All the coordinates and angles were used, in total 100 columns.

- **Training process Accuracy**: The results are good, and we can see how the model learns to classify the correctness of the different frames. Around the epoch 25 achieves a 0.8 accuracy and reaches 0.88 after 100 epochs. We can realize that the model is improving its accuracy gradually until it reaches the 115 epochs, then it we can't see a big improvement.
- **Validation process Accuracy**: Pretty similar to the training results. We can see how the model learns properly and improve the results obtained in the training phase. With 25 epochs achieve 0.85 accuracy and with 65 epochs reach 0.90, where it stops improving its accuracy, and maintain the range between 0.9 -0.95
- **Training process Loss**: The loss as well gradually decrease and around 150 epochs start to stop decreasing and maintain its value at the 0.20 - 023 range. We could have trained the model for more epochs, but the difference was minimum in the loss.
- **Validation process Loss**: The loss for the validation is quite similar to the training results. But as it happened with the accuracy, we improve a bit the results obtaining a loss in the range 0.16-0.20

After testing the model, the accuracy obtained is 0.932 accuracy and 0.934 f1 score. The graphic shows that the model performs well, but it has difficulties to understand the training data in compare with the validation data. There is a small underfitting because the model is basic for the complexity of the number of dimensions of the data and the number of epochs for training is small. As well we can appreciate a bit of noise because of the quantity of data, but the performance of the model is not part of the scope, but it can be improved in future experiments.

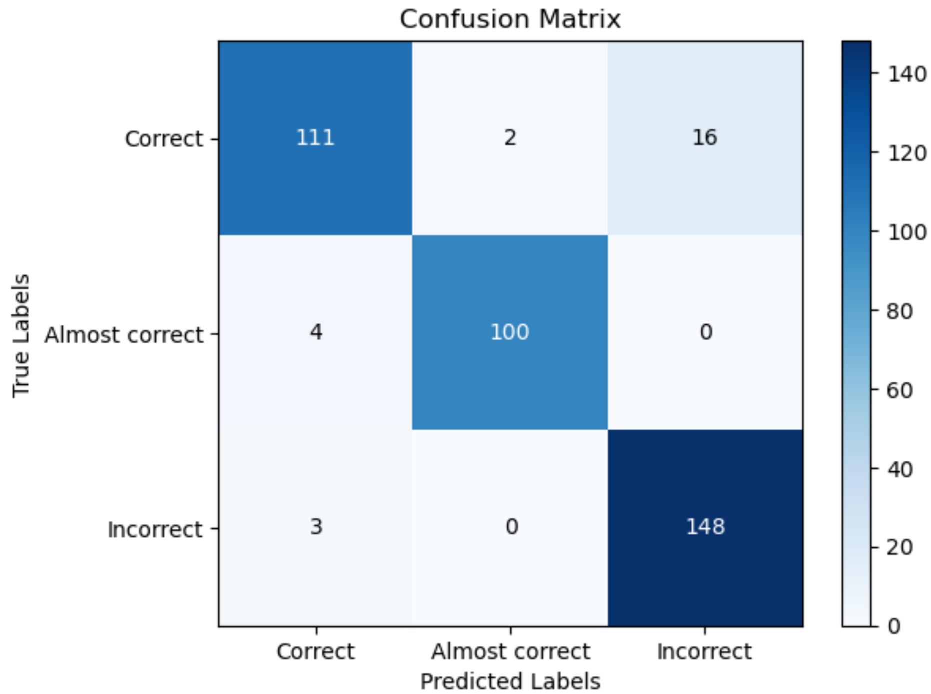


Figure 38. Confusion matrix for base use case. The validation dataset contains 384 elements. The darkest color determines a higher concentration of elements and a lighter color a smaller concentration of elements.

We can see that the results for the incorrect elements are accurate, with 148 elements properly classified and with only 3 errors. For the almost correct element is as well pretty accurate, with 100 elements correctly classified and 4 wrong classified. In contrast the correct elements are the class with less accuracy. A total of 111 element were properly classified, but 18 classified wrong, with 16 classified as incorrect. From our perspective, as we are trying to find the correctness in the movements and apart of sports this tool can be used as well in medical environment, the consideration that a movement incorrect is classified more accurate as incorrect, is more important that a correct movement classified as incorrect. In the case a patient would be doing an exercise incorrect and getting feedback as correct, could lead to a wrong recovery or even worst injure. In the other case a correct movement classified as wrong is not dramatically and the patient could ask for assistant. The same situation could be applied in sports. We expect in a future these problems to be solved.

4.2. Experimental Use Case: Dimensionality Reduction Applied

The second part of the experiment was implemented with the reduced dataset and a new CNN model. The same configuration for the CNN model and data is used, as we did for the baseline use case.

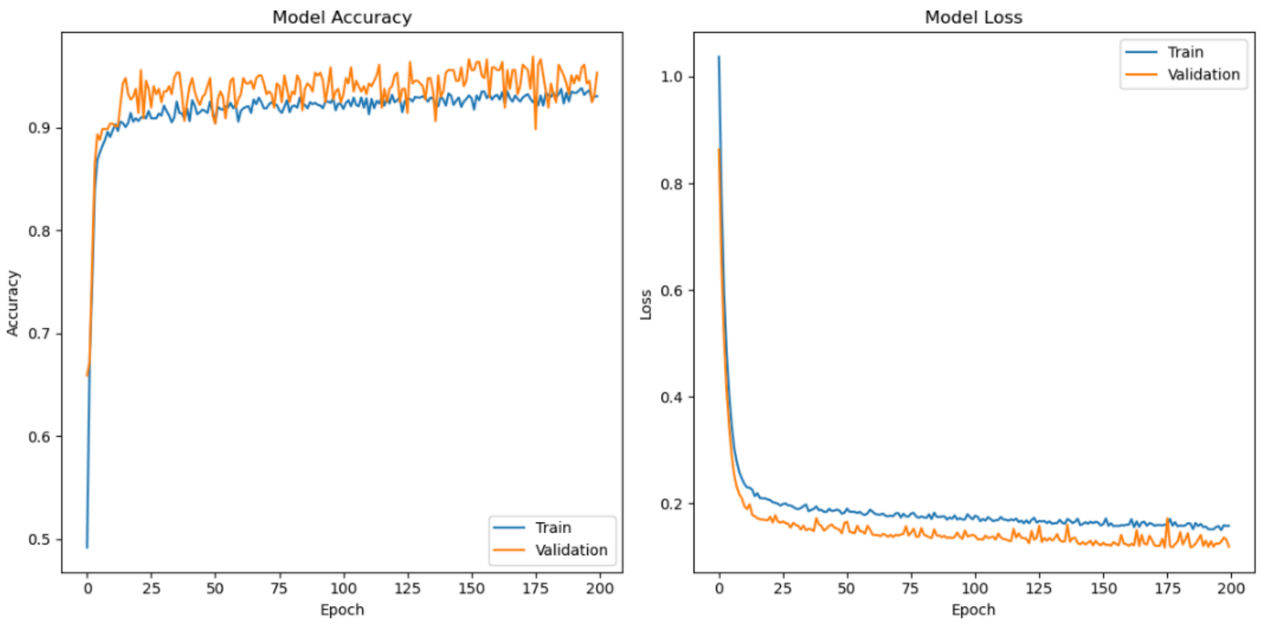


Figure 39. Accuracy and loss for the experimental use case.

The training and validation process of the model was completed in approximately 4.90 seconds for 200 epochs. In figure 39 we can notably appreciate the improvement for the results with the second model.

- **Training process Accuracy**: After the initial phase, the accuracy rapidly increased reaching its 0.9. With 25 epochs it reaches 0.92 and it range over this value for the rest of the epochs, between 0.92 -0.95.
- **Validation process Accuracy**: Pretty similar to the training results until the epoch 20, where it still continues learning. We can see how the model improve the results obtained in the training phase achieving 0.95 with 25 epochs. After these points it maintains the range between 0.94 -0.96
- **Training process Loss**: The loss as well decrease rapidly until the epoch 15, reaching a value of 0.25. After that point decrease slowly until the epoch 50, where it reaches 0.2. After this point stops decreasing and maintain its value at the 0.19 – 021 range. We could

have trained the model for more epochs, but the model was not able to drop anymore the loss.

- **Validation process Loss**: The loss for the validation is quite similar to the training results. But as it happened with the accuracy, we improve a bit the results obtaining a loss of 0.2 with 15 epochs and after those points maintaining the loss in the range 0.10-0.15

When comparing this model with the one trained using all dimensions, we observed a remarkable difference.

- The model trained with the reduced dimensions achieved its highest accuracy of 0.9 in just 25 epochs for the training phase, while the model using all dimensions required 150 epochs to reach a similar accuracy level.
- The validation results, similar to the results with the training are better in compare with the model trained with all dimensions. While it ranges between 0.93-0.6 after 25 epochs, the first model barely reaches 0.95 after 100 epochs.
- The loss for the training as well decreased more rapidly with the second model in compare with the first model. We can see how the second model learns much faster and reach a lower loss.
- The loss for the validation follows the same pattern as the loss for the training, reaching faster a lower loss in compare with the first model.
- Additionally, the training time for the model with selected dimensions was notably faster, taking only 4.90 seconds compared to the 7.83 seconds of the model with all dimensions, which means that the second model trained 38% faster.

This implies that by choosing the appropriate dimensions, we can significantly enhance the training speed, accuracy and loss of the model.

The second model shows slight improvements in accuracy and F1 score compared to the first model. The accuracy and F1 score remain the same at 0.953, while we had 0.932 accuracy and 0.934 f1-score with the model containing all the dimensions, indicating a 2% improvement. Here as well we can appreciate similar results related to the performance of the model in compare with the baseline use case. The model performs a bit better because the number of the dimensions is smaller, and the model can understand better the data. There is still noise in the results and as well, these problems can be solved with an improved model, another split of the data and a larger number of epochs for the training.

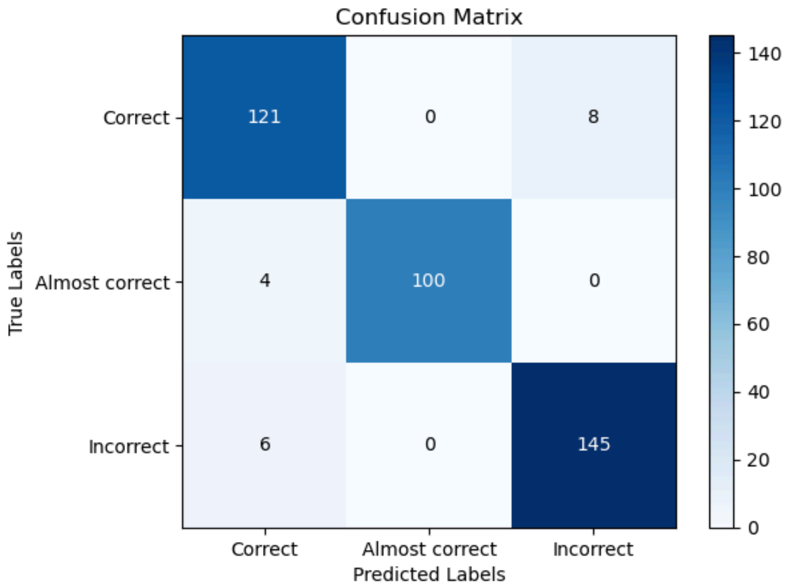


Figure 40. Confusion matrix for the experimental use case.

The confusion matrix reflects some improvements, but as well some worsen. Meanwhile the correct label improves the classification by 10 elements, the incorrect class decreased by 3. The almost correct class maintain the same number of elements. Overall, it improves the results in compare with the results from the first model, even though we have a slight increase in wrong classified incorrect elements.

Overall, the results demonstrate better performance than the first model, suggesting the potential for implementing these algorithms in future research endeavours.

Feature engineering techniques helped us to reduce the original 105 dimensions to 18 dimensions. After a meticulous study of the data, 100 dimensions were selected for this research, were joints and angles were the dimensions considered. With this data, dimensionality reduction algorithms like Random Forest were implemented, providing insights into the influential dimensions for predicting correctness, and ANOVA, helping us to identify dimensions with strong relationships to the target variable. These algorithms achieved to reduce the remained dimensions to only 15 dimensions. Moreover, PCA transformed the obtained dimension subset, into uncorrelated variables that captured the maximum data variance, effectively representing the data with only 3 dimensions. Finally, the original CO_2 footprint for baseline use case was 2.81e-05, meanwhile for the experimental use case was 1.85e-05.

After using these dimensionality reduction techniques, we have obtained a dataframe with 18 dimensions, reducing the quantity of the data by 82% and detecting the important dimensions and landmarks that define the correctness of the selected movement. We have improved the accuracy of the model by 2%, reduced the training time by 38% and as well reduced the CO2 footprint by 35%, consuming less computational resources. We can confirm the potential that dimensionality reduction algorithms have implemented with 3D Skeleton data.

5. Discussion and Future work

This research highlights the significant impact of dimensionality reduction algorithms, such as Random Forest and PCA, on 3D Skeleton data analysis. We were able to prove that the dimensionality reduction algorithms can be implemented on 3D Skeleton data, reducing successfully the number of dimensions, detecting the important dimensions to understand the correctness of specific movements, improving the accuracy of CNN models when classifying correctness and reducing the use of resources.

The importance of following a proper methodology, such as CRISP-DM, in machine learning projects was also emphasized. Using the methodology, we could understand how important is to pre-process properly the data and a big part of the time invested in this research was used in the initial phases “Business Understanding”, “Data Understanding”, “Data Preparation” and “Modelling”.

As expected, we have obtained pretty good results, but as well we have discovered the huge challenge when working with this 3D Skeleton data. The complexity of working with data in 3D space, such as preparing the data to be able to work with it or calculating the errors between the same frames for videos with different lengths, added extra complexity to this research. Developing additional functions to automatically align poses and detect pose boundaries increased the project's complexity but yielded excellent results that can benefit future researchers in the field.

The creation of extra synthetic data was another challenge, as it required creating coordinated movements with relationships between each point and are not random points in a 3D space. After carefully consideration, interpolation techniques gave us the best results, effectively enhancing the model performance and achieving a more balanced representation of the target classes.

The research also highlighted the importance of understanding biomechanics, particularly in the context of computer vision and sports. The implications of body biomechanics in our study revealed that expert knowledge is crucial. Surprisingly, our findings showed that, while the lower body is important in squatting, the upper body, specifically angles between upper landmarks, played a more consistent and significant role [43]. This contradicted our initial assumptions, emphasizing the importance of incorporating biomechanical expertise in similar studies. The theory confirms these results, confirming that the lower body's significance in squatting can vary, while the upper body components remain relatively consistent.

We have successfully reached the goals defined at the beginning of this project:

- A deep learning-based model like CNN for pose comparison was developed and we have evaluated its performance and accuracy, in compare with a second CNN model, obtaining better results.
  - Better model training time, improved by 38%.
  - Better model accuracy, improved by 5%.
  - Smaller dimensional space, reduced by 80%
  - Improved resources use, reduced the CO2 footprint by 35%.
- We have identified the key dimensions in the dataset that enable accurate and efficient pose comparison, highlighting the angles between the landmarks of the body.
- Certain specific conditions and correlations of a pose, within the selected angles were found. In this case, detected the importance of the upper body parts and spine with hips when doing squats allowed us to understand better the correctness of the pose.
- We have explored and understood the data to be able to apply the correct dimensionality reduction techniques. A strategy to prepare the data and implement the dimensionality reduction techniques were developed.

Throughout the course of this research, we encountered several challenges and complexities, balking the initial planification. The scarcity of available 3D Skeleton data and the complexity of calculating the errors between the different frames created significant obstacles. Existing datasets often lacked the required quality, documentation, or specific information, especially for movements like the squat exercise. Consequently, we had to generate our own data to ensure accuracy and relevance.

Moreover, even though it is easy to find libraries or information for pose estimation problems, finding code or investigations direction pose comparison proved challenging. Alignment of data and calculating errors between frames were essential steps in comparing poses or movements, whether comparing subjects to an instructor or an ideal standard of correctness. Despite the additional effort, the CRISPDM methodology allowed us to successfully achieve our goals, yielding promising and positive results. Because of these reasons, we had to modify the planification done at the beginning of the research several times.

Originally, we wanted to use several models to compare frames and sequences of frames, but because of the lack of time we had to focus on a single model for pose comparison, specifically using CNN. Initially, we had intended to explore the use of Transformers or LSTM to better understand movement sequences and spatial relationships between points. As well because of lack of time we were not able to implement other types of dimensionality reduction techniques like LDA, Kendall or t-SNE.

Those changes were necessary to ensure the success of this research proving the feasibility of implementing dimensionality reduction algorithms on 3D Skeleton data.

Through this research, we have successfully reduced resource consumption and mitigated the negative environmental impact associated with AI technologies, particularly in terms of CO2 emissions. The implementation of these technologies not only benefits the environment but also has a positive social impact by improving systems that support health quality of humans. By enhancing human pose technologies, we can develop efficient systems that can be used in several fields, like sports, avoiding injuries, which leads to a decrease of health quality in humans along a life span. The dataset used in our study consisted of diverse subjects, including individuals of different genders, heights, and body types.

However, for future research, it is recommended to obtain more diverse datasets that include subjects from various races and backgrounds to minimize biases. It is important to note that all subjects in the dataset were represented by numerical identifiers, ensuring the privacy and anonymity of their personal information.

This research explores a relatively unexplored domain within the field of computer vision, presenting exciting opportunities and possibilities for future exploration and advancement:

- **Data preparation**: a significant part of our time was used in the data preparation phase, a critical step to achieve good results. We focused on shaping the data to fit our models by aligning different videos and splitting them into sequences. To ensure consistency, we have developed functions that detect the start and end for each movement from a video. After that, we use a second function to extract same number of frames for all the sequences, by dropping several frames in the process. However, a drawback of our current approach is the loss of valuable information due to dropping frames to achieve an equal number of frames in each sequence. In future research, it would be beneficial to explore the development of a new function that aligns videos while retaining as many frames as possible.

- **Dimensionality reduction algorithms**: for this research we used algorithms like Random Forest and ANOVA to try to maintain the most important and relevant dimensions that define the correctness. Other algorithms for feature selection like Kendall, could be implemented to benchmark which ones obtain better results. Speaking about feature extraction PCA was the algorithms selected for this research as it works best with continuous data and the movements are continuous when performing a movement. From other side it would be interesting as well to use other algorithms like LDA or t-SNE, but LDA doesn't work well if the data is not balanced and t-SNE has a quadratic time and space complexity in the number of data points, which means that it is particularly slow, computationally heavy and resource draining. Therefore, for this research PCA was the algorithm selected to be used.

- **Models**: In this research, CNN models were selected as primary models. CNNs are models usually used in the field of computer vision, to extract features from images and enable understanding of visual data. However, CNN models can also effectively handle numerical data, such as the coordinates of 3D Skeleton data. When working with 3D spatial data, it is essential to consider the temporal and spatial relationships between the points in the 3D space. Other advanced algorithms, such as LSTM or Transformer models, have a better understanding of the sequential frames and spatial connections of points. It would be interesting to conduct a benchmark comparison between these state-of-the-art and the CNN model to explore their effectiveness in interpreting the correctness of movements.

This research serves as a steppingstone for future investigations in several directions. Exploring alternative models holds promise in comprehending the underlying sequential patterns of movement sequences, going beyond the limitations of frame-level analysis. Additionally, the alignment and sequence extraction algorithms employed here can be harnessed for other projects requiring video sequence normalization, providing the basis for further analysis and comparisons across diverse movement sequences.

I hope future students and researchers can use this code and that the research helps them providing a solid framework to build upon. By improving the existing models, refining the code implementation, and delving into additional functions, there is ample room to advance the field.

6. Glossary

3D – Three dimensions
ANOVA – Analysis of variance
CNN – Convolutional Neural Network
CO2 - Carbon dioxide
CRISPDM - The CRoss Industry Standard Process for Data Mining
LDA - Linear Discriminant Analysis
LSTM - Long short-term memory
MAPE - Mean Angle Percentage Error
MJPE - Mean Joint Percentage Error
PCA – Principal Componen Analysis
t-SNE – t-Distributed Stochastic Neighbor Embedding

7. Bibliography

[1] Wang, F. and Li, Y. (2013) Beyond physical connections: Tree Models in human pose estimation, arXiv.org. Available at: https://arxiv.org/abs/1305.2269. https://doi.org/10.48550/arXiv.1305.2269

[2] Odemakinde, E. (2023) Human pose estimation with deep learning - ultimate overview in 2023, viso.ai. Available at: https://viso.ai/deep-learning/pose-estimation-ultimate-overview/

[3] Huang, Y. (2019) FollowMeUp sports: New benchmark for 2D human keypoint recognition. Available at: https://www.researchgate.net/publication/351839687_FollowMeUp_Sports_New_Benchmark_for_2D_Human_Keypoint_Recognition. https://doi.org/10.48550/arXiv.1911.08344

[4] LaFortune, J., & Macuga, K. L. (2018). Learning movements from a virtual instructor: Effects of spatial orientation, immersion, and expertise. Journal of Experimental Psychology: Applied, 24(4), 521–533. https://doi.org/10.1037/xap0000189

[5] Mündermann, L., Corazza, S., & Andriacchi, T. P. (2006, March 15). The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications - Journal of NeuroEngineering and Rehabilitation. BioMed Central. https://doi.org/10.1186/1743-0003-3-6

[6] Odemakinde, E. (2023, February 23). Human Pose Estimation with Deep Learning - Ultimate Overview in 2023 - viso.ai. viso.ai. https://viso.ai/deep-learning/pose-estimation-ultimate-overview/

[7] Xiong, X., Min, W., Zheng, WS. et al. S3D-CNN: skeleton-based 3D consecutive-low-pooling neural network for fall detection. Appl Intell 50, 3521–3534 (2020). https://doi.org/10.1007/s10489-020-01751-y

[8] (2021) CRISP-DM help overview. Available at: https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview.

[9] Locke, S. (2017) CRISP-DM and why you should know about it, Locke Data. Available at: https://itsalocke.com/blog/crisp-dm-and-why-you-should-know-about-it/.

[10] Caggiano, V. and Durandau, G. (2022) MyoSuite -- a contact-rich simulation suite for musculoskeletal
motor ... Available at: https://www.researchgate.net/publication/360936263_MyoSuite_--_A_contact-rich_simulation_suite_for_musculoskeletal_motor_control

[11] Bullock, G.S. et al. (2021) "Baseball pitching biomechanics in relation to pain, injury, and surgery: A
systematic review," Journal of Science and Medicine in Sport, 24(1), pp. 13–20.

[12] Mehta, D. et al. (2020) XNect: Real-time multi-person 3D motion capture with a single RGB camera,
arXiv.org. Available at: https://arxiv.org/abs/1907.00837

[13] Toshev, A. and Szegedy, C. (2014) DeepPose: Human pose estimation via Deep Neural Networks,
arXiv.org. Available at: https://arxiv.org/abs/1312.4659

[14] Panariello, D. et al. (2022) Biomechanical analysis of the upper body during overhead industrial tasks
using electromyography and motion capture integrated with Digital Human Models - International
Journal on Interactive Design and Manufacturing (ijidem), SpringerLink. Springer Paris.

[15] Rogez, G. and Schmid, C. (2016) MoCap-guided data augmentation for 3D pose estimation in the
wild, arXiv.org. Available at: https://arxiv.org/abs/1607.02046

[16] Meister, D.W., Ladd, A.L. and Butler, E.E. (2011) Rotational biomechanics of the Elite Golf Swing:
Benchmarks for amateurs.

[17] Barnes, K.R. and Kilding, A.E. (2015) Running economy: Measurement, norms, and determining
factors, Sports medicine - open. U.S. National Library of Medicine.

[18] Maharaj-Pariagsingh, L. and Mohammed, P.S. (2021) DanceTutor: An its for coaching novice ballet
dancers using pose recognition of whole-body movements, SpringerLink. Springer International
Publishing. Available at: https://link.springer.com/chapter/10.1007/978-3-030-78270-2_47

[19] Douvantzis, P. et al. (2013) Dimensionality reduction for efficient single frame hand pose estimation,
SpringerLink. Springer Berlin Heidelberg. Available at:
https://link.springer.com/chapter/10.1007/978-3-642-39402-7_15

[20] Kwak, N., Choi, S.-I. and Choi, C.-H. (2008) Feature extraction for regression problems and an
example application for pose estimation of a face, SpringerLink. Springer Berlin Heidelberg. Available
at: https://link.springer.com/chapter/10.1007/978-3-540-69812-8_43

[21] Yu a et al. (2017) Exemplar-based 3D human pose estimation
with sparse spectral embedding, Neurocomputing. Elsevier.
https://doi.org/10.1016/j.neucom.2016.09.137

[22] G., I.R. (2023) CRISP-DM, Metodología de Datos, iPMOGuide. Available at:
https://ipmoguide.com/crisp-dm-metodologia-de-datos/.

[23] Domingos, P. (2012). A few useful things to know about machine learning. Communications Of The ACM, 55(10), 78. http://dx.doi.org/10.1145/2347736.2347755.

[24] Lopez Yse, D. Straightforward guide to dimensionality reduction, Pinecone. Available at: https://www.pinecone.io/learn/dimensionality-reduction/.

[25] Brownlee, J. (2020) How to choose a feature selection method for machine learning, MachineLearningMastery.com. Available at: https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/.

[26] L. Breiman, (2001), 'Random Forests', Mach. Learn., vol. 45, no. 1, pp. 5–32.

[27] B. H. Menze et al., (2009) 'A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data', BMC Bioinformatics, vol. 10, p. 213.

[28] Prasetiyowati, M.I. (2020) (PDF) feature selection to increase the random forest method ..., Feature selection to increase the random forest method performance on high dimensional data. Available at:

https://www.researchgate.net/publication/347712551_Feature_selection_to_increase_the_random_forest_method_performance_on_high_dimensional_data.

[29] Analysis of variance (2023) Wikipedia. Available at: https://en.wikipedia.org/wiki/Analysis_of_variance.

[30] M. Palmer, Principal Components Analysis, Ordination.okstate.edu., from http://ordination.okstate.edu/PCA.htm

[31] Fieraru, M., Zanfir, M. and Cristian Pirlea, S. (2021) Home: Fit3d dataset, Home | Fit3D Dataset. Available at: https://fit3d.imar.ro/.

[32] Lateef, R.A. and Abbas, A.R. (2022) Tuning the hyperparameters of the 1D CNN model to improve the performance of human activity recognition, Engineering and Technology Journal. Available at: https://etj.uotechnology.edu.iq/article_173806.html.

[33] Kuzminykh, D. and Polykovskiy, D. (2018) 3D molecular representations based on the wave transform for Convolutional Neural Networks, Molecular pharmaceutics. Available at: https://pubmed.ncbi.nlm.nih.gov/29473756/.

[34] Odemakinde, E. (2023a) Human pose estimation with deep learning - ultimate overview in 2023, viso.ai. Available at: https://viso.ai/deep-learning/pose-estimation-ultimate-overview/#:~:text=There%20are%20two%20types%20of,detecting%20poses%20for%20multiple%20objects.

[35] Ingwersen, C.K. et al. (2023) SportsPose -- A Dynamic 3D Sports pose dataset, arXiv.org. Available at: https://arxiv.org/abs/2304.01865.

[36] Soomro, K., Zamir, A.R. and Shah, M. (2012) UCF101: A dataset of 101 human actions classes from videos in the wild, arXiv.org. Available at: https://arxiv.org/abs/1212.0402.

[37] Mykhaylo, A. et al. (2014) Papers with code - MPII human pose dataset, Dataset | Papers With Code. Available at: https://paperswithcode.com/dataset/mpii-human-pose.

[38] Nagymáté, G. and Kiss, R.M. (2018) Application of optitrack motion capture systems in human movement analysis: A systematic literature review, Recent Innovations in Mechatronics. Available at: https://ojs.lib.unideb.hu/rIim/article/view/3845.

[39] Schroeder, S. et al. (2022) Accuracy measurement of different marker based motion analysis systems for biomechanical applications: A round robin study, PloS one. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9273086/. 10.1371/journal.pone.0271349

[40] Van Hooren, B. et al. (2023) The accuracy of markerless motion capture ... - wiley online library. Available at: https://onlinelibrary.wiley.com/doi/10.1111/sms.14319?af=R. https://doi.org/10.1111/sms.14319

[41] Needham, L., Evans, M., Cosker, D.P. et al. The accuracy of several pose estimation methods for 3D joint centre localisation. Sci Rep 11, 20673 (2021). https://doi.org/10.1038/s41598-021-00212-x

[42] Fieraru, M., Zanfir, M. and Cristian Pirlea, S. (2021) CVF Open Access. Available at: https://openaccess.thecvf.com/content/CVPR2021/papers/Fieraru_AIFit_Automatic_3D_Human-Interpretable_Feedback_Models_for_Fitness_Training_CVPR_2021_paper.pdf.

[43] Jeulink, K. (2020) Bachelor assignment 03-07-2020 1 the detection of (in)correct motion during resistance training, The detection of (in)correct motion during resistance training. Available at: https://essay.utwente.nl/81958/1/Jeulink_BA_EEMCS.pdf.

[44] Bauld, S. (2018) Biomechanical analysis of squat and deadlift, performance. Available at: https://stuartbauld.wixsite.com/performance/single-post/2018/02/27/biomechanical-analysis-of-squat-and-deadlift-1.

8. Annexes

Repository - https://github.com/danielGOB/TFM/tree/main

The different functions mentioned in this dissertation are going to be introduced here with the docstring explaining briefly the inputs and outputs of each function:

def **process_json_files**(data_dir):
    """
    This function iterates over the folders in the data directory and extracts pose data from the JSON files.
    It calculates angles between consecutive joints and organizes the extracted data in separate lists.

    Parameters:
    data_dir (str): The path to the directory containing the JSON files.

    Returns:
    three lists: 'data', 'raw_data', and 'jangles'.
        - 'data' contains the complete pose data, including labels, folder names, flattened coordinates, and angles.
        - 'raw_data' contains pose data without the calculated angles.
        - 'jangles' contains pose data with only the calculated angles.

    """

----

def **calculate_angle**(a, b, c):
    """
    Calculate the angle (in degrees) between three points represented as vectors.

    Parameters:
    a (array-like): Coordinates of the first point.
    b (array-like): Coordinates of the middle point.
    c (array-like): Coordinates of the last point.

    Returns:
    float: The angle (in degrees) between the vectors formed by the points.
    """

```
def interpolate_frames(df, num_intermediate_frames):
    """
    This function takes a DataFrame 'df' containing sequential frames and performs interpolation
    to generate a specified number of intermediate frames between each pair of consecutive frames.

    Parameters:
    df (pandas.DataFrame): The DataFrame containing sequential frames.
    num_intermediate_frames (int): The number of intermediate frames to be generated between each
pair of consecutive frames.

    Returns:
    The DataFrame with interpolated frames.
    """
--------------------------------
def sequence_extractor(df):
    """
    Extracts sequences from a DataFrame.

    Args:
        df (pandas.DataFrame): The DataFrame containing the data.

    Returns:
            - local_maxima_before_indices: Indices of the first local maxima before each sequence.
            - smallest_minima_indices: Indices of the smallest local minima for each sequence.
            - local_maxima_indices: Indices of the first local maxima after each sequence.
    """
--------------------------------
def alignment_sequence(df, sorted_indexes, local_maxima_before_indices,
smallest_minima_indices, local_maxima_indices):
    """
    Aligns and extracts sequences from a DataFrame based on provided indices.

    Args:
        df (pd.DataFrame): The input DataFrame containing angle data.
```

    sorted_indexes (list): The sorted indices of the sequences.
    local_maxima_before_indices (list): The indices of the first local maxima before each sequence.
    smallest_minima_indices (list): The indices of the smallest local minima for each sequence.
    local_maxima_indices (list): The indices of the first local maxima after each sequence.

  Returns:
    pd.DataFrame: The aligned and extracted sequences as a DataFrame.

  """

```
def get_sorted_indices(lst):
  """
  Returns the sorted indices of a given list.

  Parameters:
    lst (list): The list for which to obtain the sorted indices.

  Returns:
    sorted_indices (numpy.ndarray): The sorted indices of the list.
  """
```

```
def get_equally_spaced_elements(lst, num_elements):
  """
  Returns the indices of equally spaced elements from a given list.

  Parameters:
    lst (list): The list from which to obtain the indices.
    num_elements (int): The desired number of equally spaced elements.

  Returns:
    indices (numpy.ndarray): The indices of equally spaced elements.
  """
```

```
def calculate_errors(df, trainer_subject='s08'):
  """
  Calculates the errors between trainee and trainer coordinates and angles for each frame in the given
DataFrame.
```

Args:
df (pd.DataFrame): The input DataFrame containing the data.
trainer_subject (str): The subject identifier of the trainer. Defaults to 's08'.

Returns:
pd.DataFrame: A DataFrame containing the calculated errors for each frame, including coordinate errors,
angle errors, trainee angle, subject, sequence, and part.

"""
----
def **update_correctness**(df):
"""
Update the correctness label in a DataFrame based on angle error values.

Args:
df (pandas.DataFrame): DataFrame containing angle error values.

Returns:
pandas.DataFrame: DataFrame with updated correctness labels in the 'Correctness' column.

"""
----
def **random_forest_calculator(**df_final):
"""
Calculates the feature importances using a Random Forest classifier.

Parameters:
df_final (pandas.DataFrame): The DataFrame containing the data.

Returns:
List[Tuple[str, float]]: A list of tuples containing the feature name and its importance.
"""
----
def **select_features_by_cumulative_importance**(feature_scores, cumulative_threshold):
"""
Selects features based on their cumulative importance scores.

Args:

feature_scores (dict): A dictionary of feature scores, where the keys are feature names and the values are importance scores.
cumulative_threshold (float): The cumulative importance threshold, expressed as a percentage between 0 and 1.

Returns:
list: A list of selected feature names.

"""

----
def **anova_calculator**(df_final):
"""
Calculates the Analysis of Variance (ANOVA) for each column in the provided DataFrame.

Parameters:
df_final (pandas.DataFrame): The DataFrame containing the data to perform ANOVA on.

Returns:
pandas.DataFrame: A DataFrame with the F-statistic and p-value (ANOVA) for each column.
"""

CNN MODEL ARCHITECTURE

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)              Output Shape              Param #
=================================================================
 reshape_1 (Reshape)       (None, 17, 1)             0

 conv1d_1 (Conv1D)         (None, 15, 32)            128

 global_max_pooling1d_1 (Gl  (None, 32)              0
 obalMaxPooling1D)

 flatten_1 (Flatten)       (None, 32)                0

 dense_2 (Dense)           (None, 128)               4224

 dense_3 (Dense)           (None, 3)                 387

=================================================================
Total params: 4739 (18.51 KB)
Trainable params: 4739 (18.51 KB)
Non-trainable params: 0 (0.00 Byte)
```

None