UNIVERSITY of STIRLING

Division of Computing Science and Mathematics
Faculty of Natural Sciences
University of Stirling

Anomaly Detection in Action Videos

Detecting anomalies in deadlift exercise

Author:

Eric Tom Mathews

Supervisor:

Dr Jefersson dos Santos

Dissertation submitted in partial fulfilment for the degree of
Master of Science in Artificial Intelligence

September 2023

Abstract

In today's fast-paced world, maintaining physical fitness and wellbeing is paramount. Regular exercise routines have become a cornerstone of a healthy lifestyle, and individuals from all walks of life are embracing various forms of physical activity to stay in shape and improve their overall health. However, while the benefits of exercise are widely acknowledged, the importance of proper form and technique is often underestimated. Incorrect exercise forms can lead to a range of issues, from diminished effectiveness and suboptimal results to serious injuries. Recognising the critical need for effective exercise form analysis, this dissertation embarks on a journey to develop an innovative system capable of assessing exercise forms and identifying anomalies.

The Problem:

Exercise enthusiasts often encounter difficulties in maintaining correct posture while performing an exercise. This can stem from a lack of feedback and guidance, especially when exercising alone or outside a fitness facility. Form errors may include improper posture, deviations from proper lifting techniques, and inconsistencies in movement patterns. Such errors not only hinder the efficacy of workouts but can also result in injuries, some of which may be irreversible. The absence of readily available tools that can offer precise, personalised feedback during exercise routines poses a significant challenge. Recognising this challenge as a critical problem to solve, the dissertation sets out to develop a system that can accurately assess exercise forms, identify anomalies, and provide users with valuable feedback to optimise their workouts and minimise the risk of injuries.

Objectives:

The primary objectives of this dissertation were rooted in the need to address the problem of exercise form analysis:

1. *Dataset Creation:* The first objective was to curate a comprehensive dataset containing instances of correct deadlift forms. This dataset served as the foundation for training and evaluating the anomaly detection system, ensuring a robust understanding of what constitutes proper form.
2. *Pose Estimation:* Advanced pose estimation techniques were employed to accurately track key body joints and positions during deadlift exercises. Pose estimation, a subfield of computer vision, was crucial for extracting essential data from video footage.
3. *Machine Learning Integration:* To differentiate between correct and incorrect exercise forms, the dissertation integrated machine learning algorithms into the process. These algorithms, including classification of deadlift stages and anomaly detection models, were instrumental in identifying patterns associated with deviations from proper forms.
4. *Performance Evaluation:* The system's performance was evaluated using a range of metrics, including accuracy, precision, recall, and F1-score. These evaluations aimed to

assess the system's effectiveness in detecting anomalies and providing users with real time feedback.

Methodology:

The methodology employed in this dissertation was a multifaceted approach that encompassed data collection, annotation, pose estimation, machine learning integration, and performance evaluation. The creation of the dataset involved collecting video recordings of deadlift exercises performed with correct form. Pose estimation techniques, leveraging computer vision algorithms, were then applied to the video footage to precisely track key body joints, providing a detailed representation of each exercise repetition, a process that required annotation to identify deadlift stages. Machine learning models were integrated into the system to process the pose estimation data and identify patterns indicative of incorrect forms. These models, including classification algorithms and anomaly detection techniques, played a pivotal role in distinguishing between correct and deviant exercise forms. Finally, the system's performance was evaluated on a test set using metrics including accuracy, precision, recall and F1-score.

Achievements:

The dissertation's accomplishments are noteworthy. It successfully addressed its objectives, beginning with the creation of a dataset exclusively comprising correct deadlift forms—a crucial foundation for training the anomaly detection system. Advanced pose estimation techniques allowed for precise tracking of key body joints, enabling a thorough analysis of exercise forms. Machine learning algorithms, including classification and anomaly detection models, were skilfully integrated, enabling the system to identify deviations from proper forms. Performance evaluation demonstrated the system's capability to detect anomalies with a high degree of accuracy, precision, recall, and F1-score.

In conclusion, this dissertation represents a significant milestone in the intersection of technology and physical fitness. While it may not claim to be the ultimate fine-tuned model, it stands as a promising step forward. The dissertation has laid the foundation for the integration of computer vision and machine learning in exercise form analysis, offering a path towards safer and more effective workout routines. As the digital and physical worlds continue to converge, the possibilities for utilising technology to promote healthier lifestyles are limitless. This project is not the final destination but rather a pivotal point in the journey towards a future where exercise form analysis becomes an accessible and indispensable tool for fitness enthusiasts.

Attestation

I understand the nature of plagiarism, and I am aware of the University's policy on this.

I certify that this dissertation reports original work by me during my University project except for the following:

- The Mediapipe pose estimation implementation mentioned in Sections 5.1.2, 5.1.3 and 6.2 was taken from [5].
- The skelemotion representation code implementation mentioned in Sections 5.1.5 and 6.3 was taken from [30].
- The autoencoder model implementation and prediction with a prior mentioned in Section 5.2 and 6.4 was taken from [31].

Signature **Date:** 14th September 2023

Acknowledgements

I extend my heartfelt thanks to Dr. Jefersson dos Santos, my dedicated supervisor, for his unwavering support, guidance, and expertise throughout the progress of this dissertation. His insightful feedback and encouragement have been instrumental in shaping the research and its outcomes.

I would like to express my sincere gratitude to the IT team at the University of Stirling for providing access to GPU resources. Their assistance in facilitating the computational requirements of this project has been indispensable, enabling the implementation of machine learning models and enhancing the quality of the project.

I would also like to thank Dr. Carlos Caetano for his invaluable assistance in integrating his work on SkeleMotion representation into the creation of the skeleton image representation of the exercise movement. His collaboration greatly enriched the project and extended its capabilities.

Dr. Wen-shin Lee deserves special recognition for her support during the ethics approval process and project demonstrations. Her guidance ensured that the project was conducted ethically and met all necessary standards.

Furthermore, I am grateful to Dr. Yuanlin Gu for his support during the poster presentation and project demonstrations. His insights and feedback were invaluable in refining the project and its presentation.

I am deeply appreciative of the University of Stirling Sports Center for generously providing the environment in which the deadlift dataset was created. This environment was instrumental in collecting the necessary data for this dissertation, contributing significantly to the project's success.

Lastly, I would like to acknowledge the unwavering support of my family and friends. Your belief in me has been a constant source of motivation and strength throughout this academic journey.

This dissertation represents a collective effort, and I am deeply indebted to everyone who played a part in its realisation.

Table of Contents

List of Figures

List of Tables

1 Introduction

Engaging in regular exercise is pivotal for maintaining optimal health and overall wellbeing. Daily physical activity yields a multitude of advantages, including improved physical fitness, enhanced muscular and bone strength, positive mental health, improved posture, and increased flexibility – all of which contribute to extended longevity. Emphasising proper exercise form is equally vital, serving as a key preventative measure against injuries. A study involving physically active individuals revealed that approximately 56% experienced exercise-related injuries [1]. This case study focuses on identifying anomalies and incorrect forms within exercise videos, particularly during the execution of deadlifts and thus minimising the chance of injury. The choice of deadlifts stems from their compound nature, involving multiple joints. The success of this case study could potentially pave the way for applying the methodology to a wide range of other exercises.

1.1 Context

"Safety should never be a priority. It should be a precondition." [2] *(Paul H O'Neill, chairman and CEO of the Pittsburgh industrial giant* Alcoa *)*

Detecting anomalies in exercise videos and accurately estimating movement is not merely a technical endeavour; it holds profound motivational significance for individuals pursuing an enhanced fitness journey. Firstly, the capacity to identify anomalies ensures that exercises are performed in the most effective and safest manner possible. By pinpointing deviations from proper form, individuals are empowered to make real time adjustments, optimising the benefits of each repetition. This awareness fosters a strong mind-to-muscle connection, cultivating a deeper understanding of the body's mechanics and aligning efforts with fitness goals. With this technology as a guide, individuals can proactively address potential pitfalls and prevent injuries that might otherwise hinder progress.

Secondly, the pursuit of anomaly detection and movement estimation ignites motivation by providing concrete evidence of growth and dedication. Observing improvements in form and technique through data-driven feedback generates a sense of achievement and progress. This, in turn, nurtures confidence and reinforces the notion that diligent work yields results. As movements become more fluid, strength increases, and performance refines, individuals are encouraged to push boundaries and set higher aspirations. The technology's objective tracking of evolution not only amplifies enthusiasm for exercise but also propels individuals to maintain consistency, surmount plateaus, and embrace the transformative potential of their fitness journeys.

What are Deadlifts?

The deadlift is a fundamental compound exercise (involving multiple joints) that holds a central place in strength training. It involves lifting a barbell or weighted object from the ground to a standing position. To execute a standard deadlift, an individual starts by standing in front of the barbell with their feet shoulder-width apart. Maintaining a neutral spine and braced core, they bend at the hips and knees, gripping the barbell with either a double overhand or mixed grip. With a strong grip and engaged posterior chain, the lifter lifts the barbell off the ground by extending their hips and knees until they are fully upright. The movement is reversed to lower the weight back to the ground with controlled form.

Bottom Position (Start / End) | Neutral Position | Top Position

Figure 1: Deadlift Stages

Deadlifts are a powerhouse exercise with a range of benefits. They intensely target muscles such as the lower back, glutes, hamstrings, and quadriceps, promoting strength and muscle development. Additionally, deadlifts engage the core, enhancing stability and supporting good posture. Beyond physical gains, the deadlift fosters mental resilience due to its demanding nature. Proper form is pivotal to avoid injury, so beginners are advised to seek guidance to ensure safety and effectiveness. Whether as a standalone exercise or as part of a comprehensive strength regimen, the deadlift epitomizes functional strength and offers a robust foundation for overall fitness.

Figure 1 illustrates the three distinct stages of a deadlift: the Bottom Position, Neutral Position, and Top Position. The individual executing the deadlift commences from the bottom position, progresses through the neutral position to reach the top position, and then returns to the bottom position, traversing through the neutral position once more. Within the context of an exercise routine, a singular instance of this sequence is termed a "repetition" (rep). Multiple repetitions of this movement may be performed, collectively forming a "set" – a sequential series of repetitions.

What are anomalies in Deadlifts?

Anomalies in deadlifts refer to deviations or irregularities from the correct form and technique while performing the exercise. These anomalies can compromise the effectiveness of the lift and increase the risk of injury. Detecting and addressing anomalies is crucial to ensure that the deadlift is performed safely and efficiently. Some common anomalies in deadlifts include:

1. *Rounded Back:* One of the most significant anomalies is rounding of the lower back during the lifting phase. This can place excessive stress on the spine and lead to injuries. Maintaining a neutral spine is essential for distributing the load properly and protecting the lower back.

2. *Jerky or Uncontrolled Movement:* Rapid and jerky movements during the lift can strain the muscles and increase the risk of injury. Deadlifts should be executed with controlled and deliberate movements to ensure proper muscle engagement and alignment.
3. *Barbell Drift:* Allowing the barbell to drift away from the body during the lift can alter the biomechanics and place additional strain on the back and shoulders. The barbell should remain close to the body throughout the movement.

Identifying and addressing these anomalies through proper coaching, feedback, and video analysis is vital for performing safe and effective deadlifts. This dissertation centers around the analysis of deadlifts and the detection of any anomalies using pose estimation in conjunction with other machine learning techniques.

1.2 Purpose and Motivation

The purpose of this dissertation is twofold, driven by a combination of personal passion and a deep appreciation for the potential of Artificial Intelligence:

1. *Passion for Fitness and Injury Prevention:* Over the past eight years, I've been on a fitness journey that has had its share of challenges and injuries. This personal experience has motivated me to create a solution that can minimize the risks associated with fitness-related injuries. Often, access to a personal fitness trainer for form correction is neither easy nor affordable, and some injuries can be irreversible. My goal is to develop an application that empowers users to perform exercises correctly, preventing them from repeating mistakes, and reinforcing proper form.

2. *Harnessing the Power of Artificial Intelligence:* My fascination with the field of Artificial Intelligence, particularly neural networks, has driven me to explore its potential in solving this problem. I'm inspired by the remarkable capabilities of neural networks and their ability to learn and adapt. This dissertation serves as a platform to apply and expand my knowledge in computer vision and generative modelling within the realm of fitness and exercise analysis.

It's important to note that this dissertation strikes a balance between these motivations and the constraints of time and resources. While it may not produce a state-of-the-art application for correcting exercise form, it represents a significant step forward in understanding the intersection of AI and fitness. Future opportunities and avenues for improvement will be explored in subsequent discussions.

1.3 Scope and Objectives

The scope and objectives of this dissertation lay the groundwork for a comprehensive exploration into the realm of exercise analysis and anomaly detection, with a particular focus on the deadlift exercise. As physical fitness goals increase in prominence in contemporary lifestyles, the need for accurate, personalised, and technologically driven solutions for optimising workout routines has become more evident. This dissertation aims to bridge this gap by delving into the integration of pose estimation techniques and machine learning algorithms to develop a robust system capable of identifying anomalies within deadlift performances. By precisely tracking key body joints and movements, this project endeavours to offer a valuable tool for individuals seeking to enhance their exercise technique while minimizing the risk of injuries. With an eye on both technical advancement and user applicability, the outlined scope and objectives form a cohesive framework designed to contribute to the evolving landscape of fitness technology and empower users in their pursuit of safer and more effective workout practices.

1. *Anomaly Detection:* The dissertation will primarily focus on developing a robust system for identifying anomalies in deadlift exercises using pose estimation and machine learning. It will explore various types of anomalies such as improper posture, deviations from proper lifting technique, and inconsistencies in movement patterns.
2. *Creation of Deadlift Dataset:* A fundamental goal of this project is the creation of a comprehensive deadlift dataset. This dataset will include video recordings of an individual performing the deadlift exercise with good form. The dataset is carefully curated by following the assumptions mentioned below.
3. *Pose Estimation:* The scope will encompass the application of advanced pose estimation techniques to accurately track the key body joints and positions during the deadlift exercise. This involves using computer vision algorithms to analyse video footage and extract relevant data.
4. *Machine Learning Integration:* The dissertation will involve the integration of machine learning algorithms to process the pose estimation data and identify patterns associated with correct and incorrect deadlift forms. This integration may include classification algorithms, anomaly detection models, and potentially neural networks for enhanced accuracy.
5. *Data Collection and Annotation:* Data is gathered from an individual executing a deadlift, which is subsequently fed into a pose estimator to identify crucial keypoints. The training dataset exclusively comprises instances of correct deadlift form, strategically facilitating the anomaly detector's capability to identify deviations from this established norm.
6. *Performance Evaluation*: The objectives will involve evaluating the developed system's performance on various metrics, such as accuracy, precision, recall, and F1-score. The system's ability to accurately detect anomalies and provide real-time feedback will be a key focus.
7. *User-Friendly Implementation:* This dissertation aims to ensure accessibility to individuals engaging in deadlifts, making it user-friendly and replicable for various exercises. The functionality should extend beyond the creator and cater to anyone seeking to improve their exercise techniques, promoting inclusivity and widespread usability.

By achieving this scope and these objectives, the dissertation intends to contribute to the field of exercise analysis and provide a valuable tool for individuals aiming to improve their deadlift technique while minimising the risk of injuries.

1.4 Assumptions

Before delving into the intricacies of this dissertation, it is important to clarify the underlying assumptions:

1. *Single Performer Dataset:* Focusing on a dataset featuring a solitary performer executing the deadlift exercise is a deliberate choice. It ensures a clear and unambiguous context for analysing exercise form. By eliminating the presence of multiple individuals, the system can concentrate solely on the actions and movements of the performer, reducing potential complexities that might arise from distinguishing between different subjects.

2. *Stable Camera Placement:* The assumption of a stable camera position directly in front of the exercise performer is a critical foundation. It guarantees a consistent viewpoint, which is crucial for accurate pose estimation and exercise form analysis. Variations in camera angles or positions can introduce distortions and inaccuracies in skeleton keypoint detection, potentially leading to erroneous assessments of form.
3. *Single-Person Frame:* The constraint that each video must contain only one person within the frame is rooted in the need for precise isolation of the exercise performer. This restriction simplifies the process of identifying and tracking the relevant individual throughout the video, ensuring that the system's analysis is solely focused on the correct subject.
4. *Comprehensive Framing:* Enforcing the requirement that video frames encompass the entire body of the individual throughout the lift is essential for holistic analysis. It enables the system to consider the complete range of motion and form throughout the exercise, thereby providing a more accurate assessment of the deadlift technique.
5. *Repetition-Based Predictions:* Making predictions based on entire repetitions rather than individual frames aligns with the practical context of exercise assessment. It mirrors how fitness enthusiasts typically evaluate their form by observing complete repetitions. This approach captures the fluidity and continuity of movements, offering a more realistic and meaningful analysis of exercise form.

In summary, these assumptions serve as a deliberate and necessary framework for the project, enabling a focused exploration of exercise form analysis. They provide clarity, consistency, and context, which are indispensable elements for developing a robust and effective system for form assessment. By adhering to these assumptions, the project establishes a solid foundation upon which to build its methodology and approach.

1.5 Achievements

This dissertation represents a pivotal step in my learning journey, particularly in the domain of anomaly detection and the practical development of real-world applications. It has provided me with hands-on experience and an in-depth understanding of how to leverage artificial intelligence techniques for addressing tangible problems. The process of conceptualising, designing, and implementing this project has been a valuable learning curve, pushing the boundaries of my knowledge and challenging me to find creative solutions. The experience gained from building this anomaly detection system will undoubtedly serve as a strong foundation for future endeavours in the intersection of AI and fitness technology.

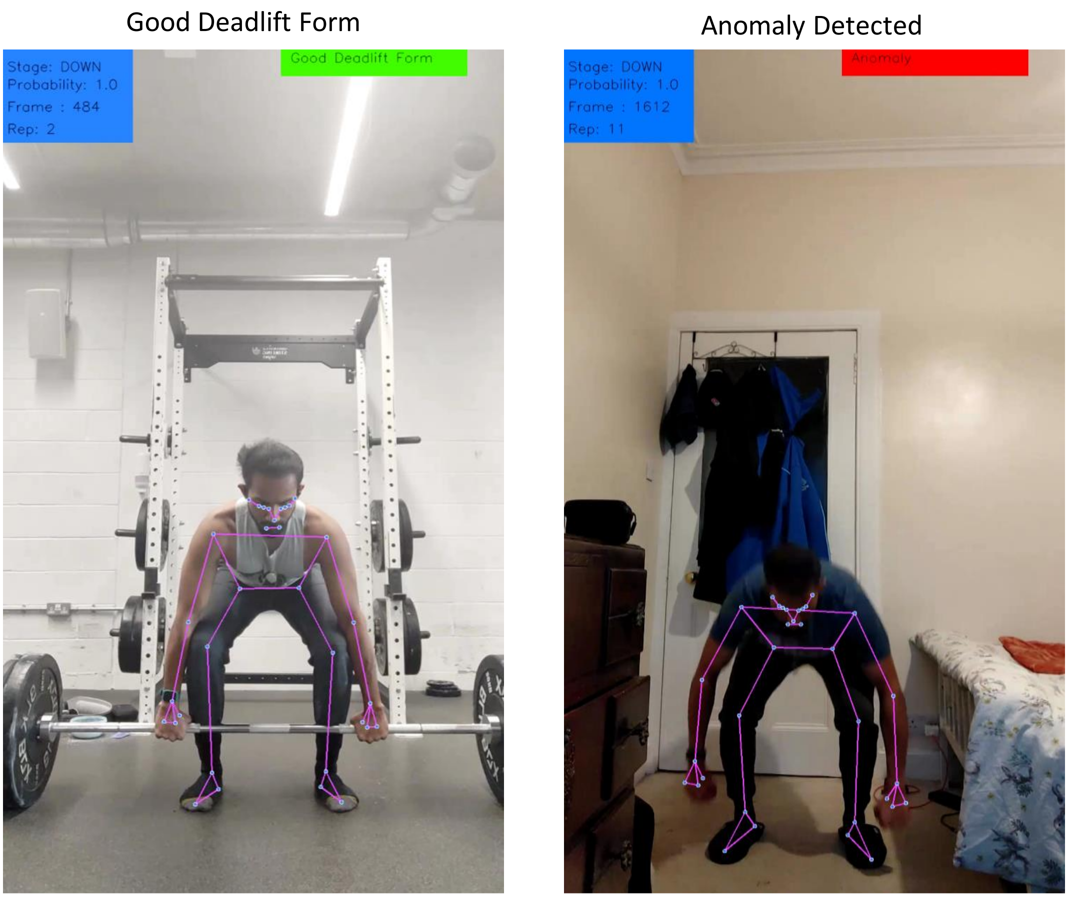


Figure 2: Results

The final autoencoder model stands as a testament to the potential of this project. Achieving an ***88% overall accuracy*** on a randomly selected test set is a promising indication of its capabilities. Moreover, the low training reconstruction ***Mean Squared Error (MSE) of 0.0079*** underscores the model's ability to capture intricate patterns within the data. As demonstrated in the figure above, the system identifies anomalies, current exercise stages, and repetitions. Despite these impressive results, I acknowledge that there is room for improvement. By expanding the dataset with more samples and fine-tuning the model's architecture, we can enhance its performance even further. These results not only demonstrate the effectiveness of the anomaly detection system but also highlight the potential for its application in identifying anomalies, tracking exercise stages, and monitoring repetitions, which holds great promise for enhancing the safety and effectiveness of fitness training.

1.6 Overview of Dissertation

This dissertation unfolds across seven distinct chapters, each contributing a vital piece to the overarching narrative of the project:

Chapter 1 serves as the project's introduction, laying a contextual foundation for the entire dissertation. It explains the purpose and motivation that underpin the project, delineates its scope and objectives, and provides a snapshot of the achievements attained through its course.

Chapter 2 delves into the state of the art in the realm of anomaly detection, offering a comprehensive exploration of the latest technologies deployed to tackle similar challenges. It scrutinizes analogous projects, drawing insightful comparisons and contrasting elements to provide a well-rounded understanding.

Chapter 3 is the bridge between theory and application. It presents the essential background information imperative for grasping the subsequent intricacies. Topics such as pose estimation, SkeleMotion representation, autoencoder models, exploratory data analysis (EDA) employing t-distributed stochastic neighbor embedding (t-SNE), and prediction mechanisms involving a prior distribution are explained.

Chapter 4 bifurcates the project into its fundamental pipelines: one for training and the other for prediction. This chapter dissects the project's core structure, offering a clear roadmap for the reader to navigate through the intricacies of each pipeline.

Chapter 5 embarks on the journey through the training pipeline, unraveling the steps involved in crafting the dataset, conducting exploratory data analysis (EDA), training the autoencoder model, and concluding with the evaluation of the training pipeline's results.

Chapter 6 is dedicated to the prediction pipeline, explaining the intricate process of handling input videos, identifying the initiation and conclusion of each exercise repetition to generate SkeleMotion representations, implementing prediction mechanisms guided by a prior distribution, and ultimately presenting the output results. Additionally, this chapter offers insights into the time statistics governing the prediction process.

Finally, in ***Chapter 7***, the dissertation gives a comprehensive conclusion. It synthesizes and evaluates the project's entirety, providing critical insights into its limitations while outlining promising avenues for future research and development.

2 State-of-The-Art

Exercise form analysis using AI represents a transformative leap in the fitness and health industry. These systems have the potential to revolutionise how we approach exercise. They offer real time, personalised feedback on exercise technique, helping individuals optimise their form and reduce the risk of injury. AI-driven exercise analysers can cater to users of all fitness levels, from beginners looking to establish proper form to seasoned athletes fine-tuning their performance. Moreover, these systems make fitness more accessible by providing guidance and motivation through virtual trainers and gamified workouts. As technology continues to advance, we can anticipate even more sophisticated and immersive exercise form analysis tools that empower individuals to achieve their fitness goals safely and effectively.

The state of the art in exercise analysers has advanced significantly in recent years, primarily driven by advancements in computer vision, artificial intelligence, and wearable technology. Some key trends and technologies in the field:

1. *Computer Vision and Pose Estimation:* Exercise analysers now leverage computer vision and pose estimation techniques to track body movements during exercises accurately. These systems can provide real-time feedback on form and posture, helping users perform exercises correctly and avoid injury. They often use deep learning models to identify key joints and body parts.

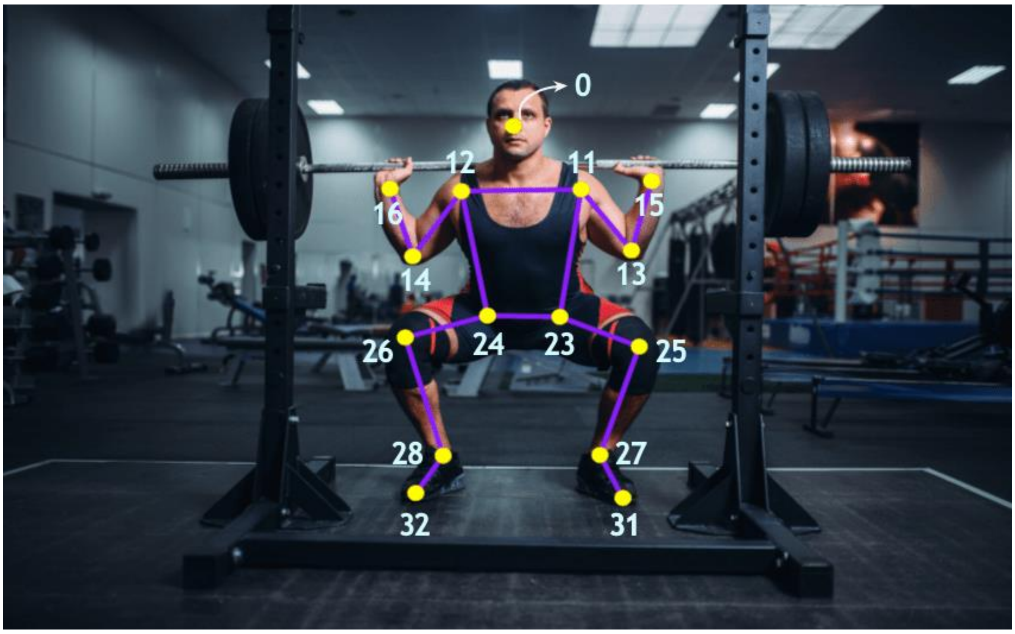


Figure 3: Computer Vision and Pose Estimation in analysing Exercise [20]

2. *Wearable Devices:* Wearable fitness trackers and smartwatches have become ubiquitous. These devices can monitor various aspects of exercise, including heart rate, steps taken, calories burned, and even sleep patterns. Some advanced wearables can provide guidance on exercise routines and offer personalised recommendations.

Figure 4: Everyday wearable technologies [21]

3. AI-Powered Personal Trainers: AI-powered virtual personal trainers are gaining in popularity. These apps use AI algorithms to create personalised workout plans, track progress, and provide feedback. They can adapt workouts based on the user's fitness level, goals, and feedback.

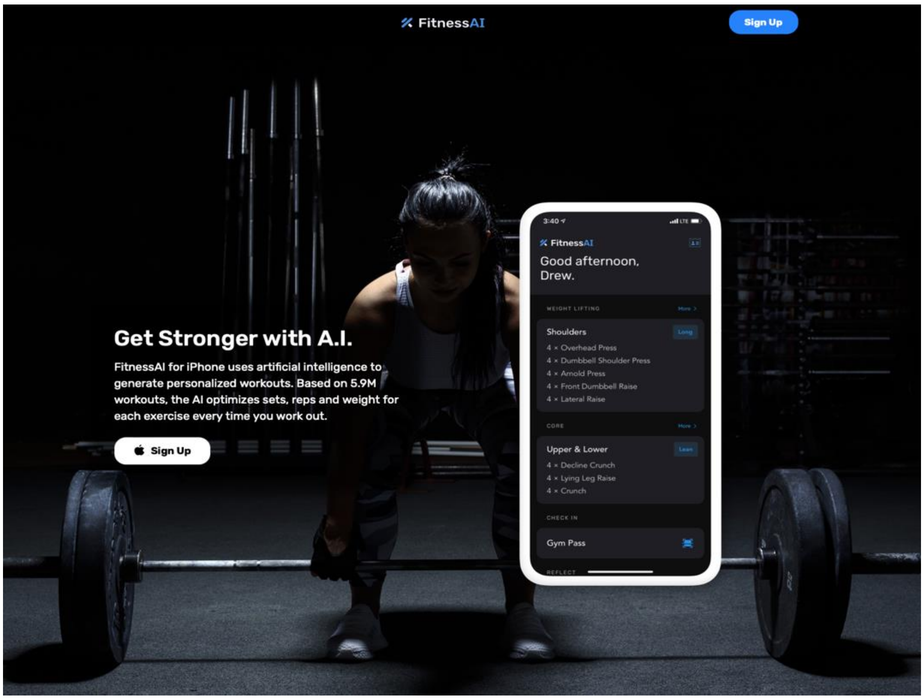


Figure 5: Fitness AI, an AI powered fitness App [22]

4. *Gamification:* Gamification elements are being integrated into exercise analysers to make workouts more engaging. Users can compete with friends, earn rewards, and track achievements, which can motivate them to stay consistent with their exercise routines.

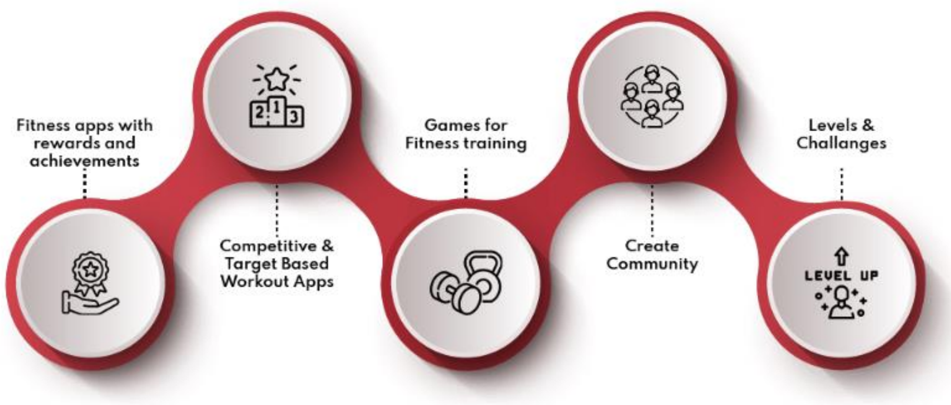


Figure 6: Fitness App Engaging with Gamification [23]

5. *Biometric Data Integration:* Some exercise analysers incorporate biometric data, such as blood pressure and oxygen saturation levels, to provide a more holistic view of health and fitness. This can help individuals tailor their exercise routines to improve specific health metrics.

Figure 7: Biometric Data and Fitness [24]

6. *Virtual Reality (VR) and Augmented Reality (AR):* VR and AR technologies are being used to create immersive exercise experiences. Users can participate in virtual fitness classes or games that encourage physical activity.

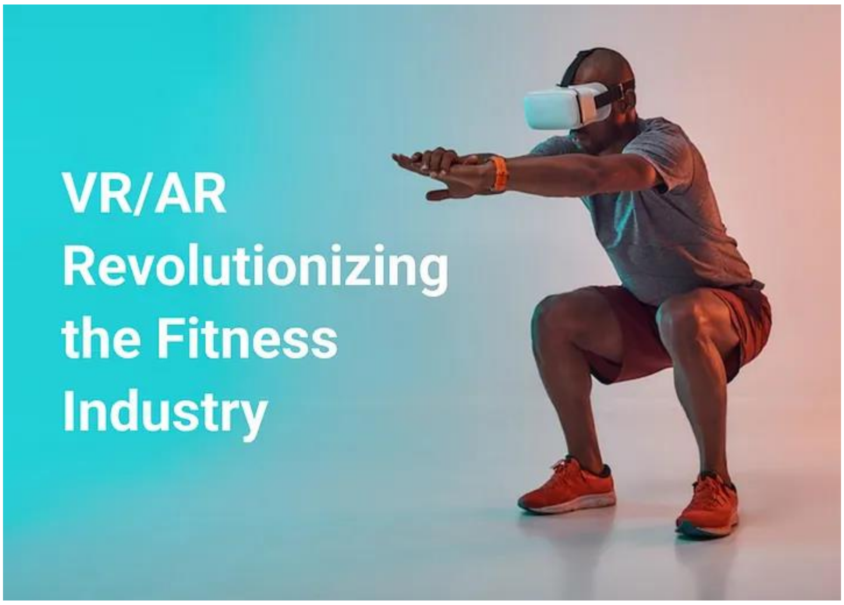


Figure 8: VR/AR in Fitness [25]

7. *Machine Learning for Anomaly Detection*: Machine learning models are employed to detect anomalies or incorrect exercise forms. These models can analyse movements and provide real time feedback to ensure users are performing exercises safely and effectively.

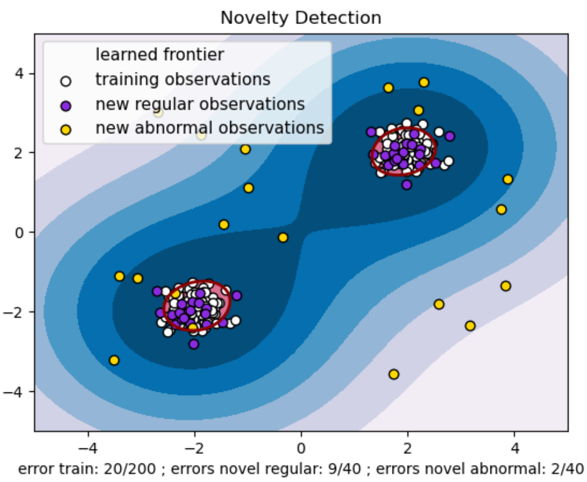


Figure 9: Novelty Detection [26]

2.1 Similar Works

While this dissertation introduces a unique approach through anomaly detection techniques, it's essential to acknowledge the existing state-of-the-art technologies in the field of exercise form analysis and feedback. The platforms mentioned below share common motivations and methods with this dissertation, albeit with distinct implementations. These technologies collectively represent the growing importance of leveraging AI and computer vision for enhancing exercise routines and optimising performance. By recognising the broader landscape of related solutions, we can appreciate the diversity of approaches aimed at improving fitness and promoting healthy lifestyles.

2.1.1 Onform: Video Analysis App

Onform [27] is a video analysis and coaching platform designed to enhance the coaching experience for both in-person and remote athletes. It enables coaches to provide valuable video feedback and maintain effective communication with their athletes, regardless of their physical location. Onform offers features like voice-over feedback on exercises, allowing coaches to provide precise guidance. It also facilitates the comparison of two videos, enabling coaches and athletes to analyse form changes or track progress over time. Onform primarily focuses on serving the sport of golf, making it a valuable tool for golf coaches and athletes seeking to refine their skills and analyse their performance through video analysis and remote coaching. This paid platform shares common objectives with the dissertation, particularly in the realm of form assessment with skeleton tracking and repetition counting, showcasing the practical applications and demand for advanced technology in fitness and sports coaching.

Figure 10: Onform Applications [27]

2.1.2 Exer Health

Exer Health [28] stands as a motion-AI digital health application designed to facilitate both patients and healthcare providers in assessing and monitoring various aspects of musculoskeletal (MSK) conditions. Its core functionality includes evaluating range of motion, mobility, and strength, making it particularly valuable for tracking progress before and after surgical procedures. Exer Health employs computer vision, similar to the techniques explored in this dissertation, to precisely measure joint angles and count repetitions, enabling real time form correction during exercises. What sets Exer Health apart is its ability to automatically detect physical biomarkers on the user's body, eliminating the need for additional sensors or hardware. These markers are then leveraged to gather accurate measurements, offering healthcare providers valuable data for remote patient monitoring, a feature that aligns with the dissertation's focus on remote exercise analysis. While Exer Health is primarily centred on physiotherapy and healthcare patients, its use of pose estimation and computer vision techniques underscores the relevance of these technologies in enhancing healthcare and exercise assessment practices.

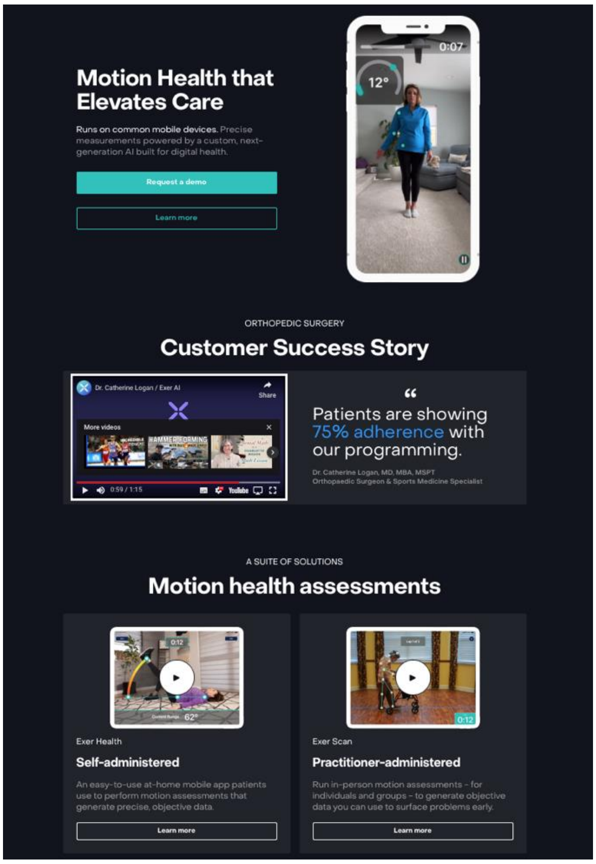


Figure 11: Exer Health Home Page [28]

2.1.3 FormCheck – AI Video Analysis

Form Check [29] offers a comprehensive solution for exercise enthusiasts, providing the means to record, measure, and refine their exercise movements and bar paths. Users can harness its measurement systems to assess various aspects of their performance, including symmetry, angles, and body part positions, thus enabling a thorough analysis of their workouts. It allows users to customise their exercise routines by creating personal targets and notifications while receiving real time video feedback during their workouts. Additionally, Form Check offers a unique feature by allowing users to set a desired pace (Reps per Minute) for a set, and it measures each repetition to determine if the user is maintaining the set pace. This application caters to a wide range of exercises, from squats and pull-ups to push-ups and deadlifts, making it a versatile tool for fitness enthusiasts. Its functionality closely aligns with the dissertation's focus on exercise form analysis and feedback, emphasising the significance of technology in enhancing fitness routines and optimising performance.

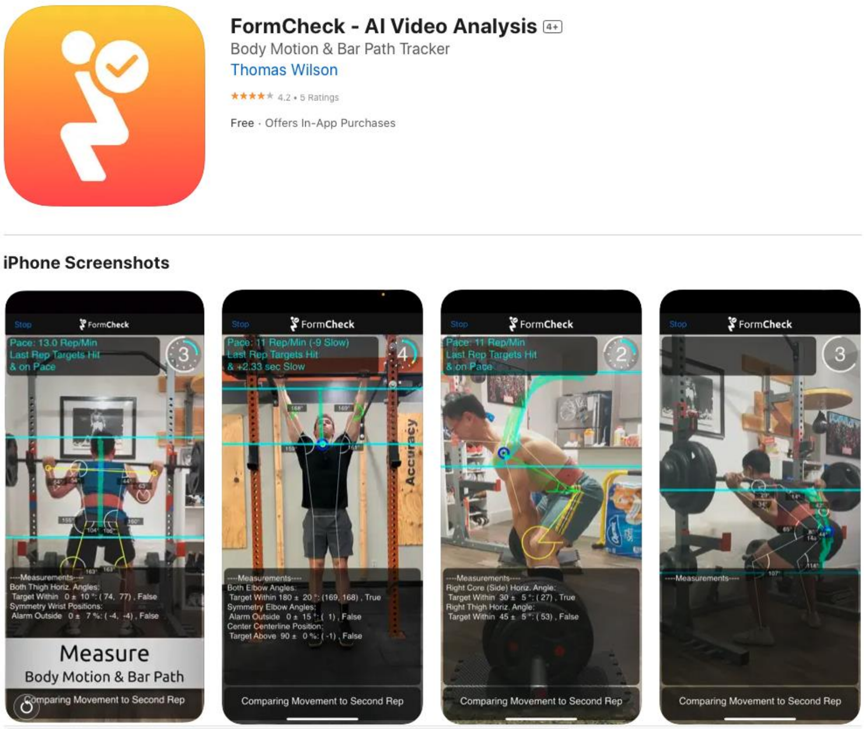


Figure 12: Form Check Application [29]

3 Background

3.1 Human pose Estimation

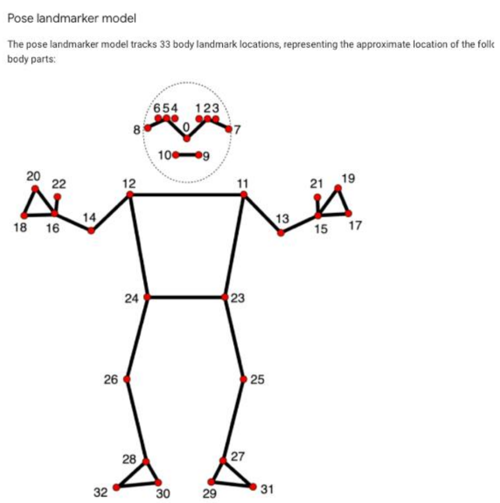
Human pose estimation within the realm of fitness is currently experiencing heightened interest, largely attributed to the recent strides made in the field of neural networks. Human pose estimation, a technology rooted in computer vision, holds the capability to detect and analyze human postures [3]. The article [3] expounds upon the implementation of pose estimation within an AI fitness coaching application, delineating the process through the subsequent steps:

1. Segmenting the input video based on exercise commencement and conclusion.
2. Identifying 2D and 3D keypoints on the user's physique.
3. Dissecting the exercise into distinct phases.
4. Scanning for common errors.
5. Drawing comparisons between frames of the input video and reference frames.

Within this dissertation, we will combine the selected strategies highlighted above while integrating terminologies delineated in [4] and [5]. Nicholas Renotte, an influential figure on YouTube renowned for his content pertaining to data science, machine learning, and deep learning, expounds on his development of an AI bot geared towards learning deadlifts. This innovative approach employs Google's Mediapipe pose estimator [8] to guide the learning process.

The Mediapipe model is adept at detecting 33 landmarks in each frame, each associated with distinct x, y, and z coordinate values. Importantly, the detected pose can be conceptualised as a disconnected graph consisting of three distinct components.

Landmark Model



Keypoints

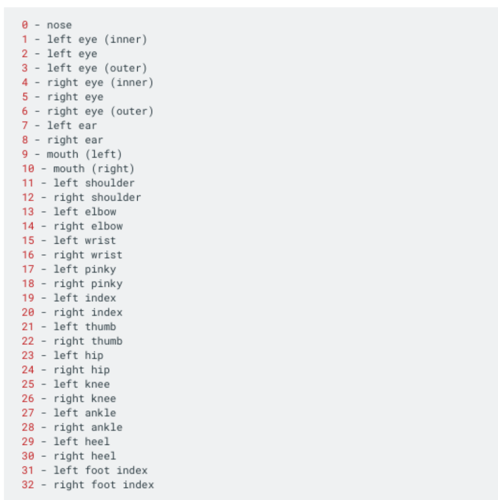


Figure 13: Mediapipe Pose Landmarks [8]

3.2 SkeleMotion Representation

Expanding the horizon, C. Caetano, J. Sena, and others [7] present a fresh perspective on skeleton joint sequences. Their proposition introduces a novel representation rooted in motion data, particularly for 3D action recognition. This framework introduces three distinct representation modes: Magnitude, Orientation, and the Tree Structure Reference Joint Image (TSRJI). This method relies on encoding temporal dynamics by explicitly utilising motion information, specifically focusing on the magnitude and orientation of skeleton joints. In this dissertation, we leverage this representation, predominantly focusing on Orientation. This enables the creation of a skeleton image portrayal depicting joint movements intrinsic to the deadlift exercise, subsequently utilised as the foundation for our anomaly detection dataset.

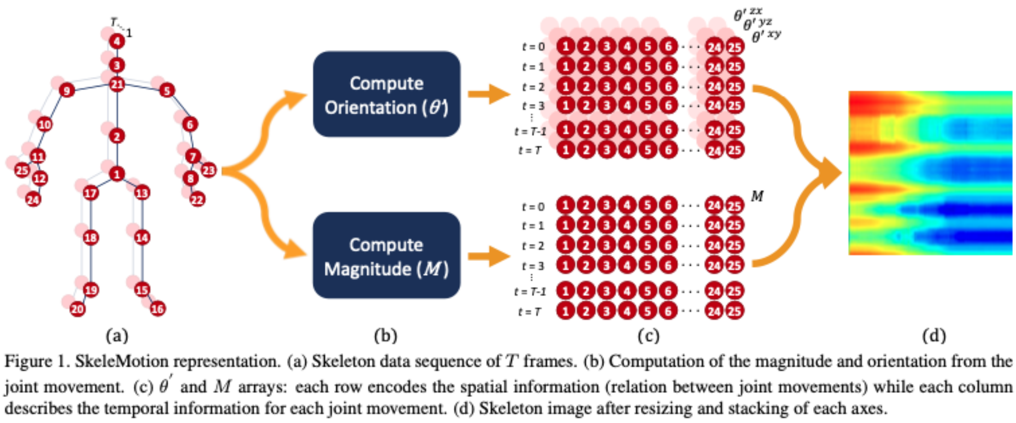


Figure 14: SkeleMotion Representation workflow [7]

3.3 Exploratory Data Analysis with t-SNE

T-distributed Stochastic Neighbor Embedding (t-SNE) [9] serves as a valuable resource for visualising data existing in high-dimensional spaces. It operates by transforming the similarities among data points into joint probabilities, with the aim of minimising the Kullback-Leibler (KL) divergence between the joint probabilities of the low-dimensional embedding and the original high-dimensional data.

Figure 15 illustrates the t-SNE embeddings of the MNIST dataset [11], a transformation of image data into a two-dimensional visualisation. Within the figure, ten distinct clusters are evident, each corresponding to a different digit. t-SNE proves valuable for visualising complex, high-dimensional data and allows for the clear differentiation of these digit clusters.

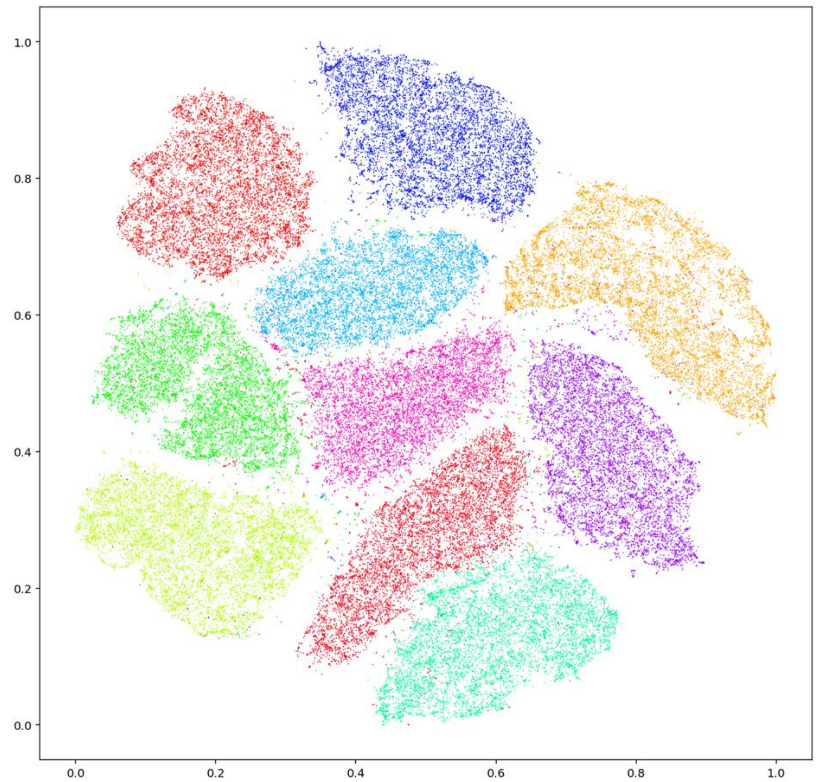


Figure 15: t-SNE embeddings of MNIST dataset [10]

3.4 Autoencoder Model

An autoencoder [13] is a specialised artificial neural network designed to master efficient codings of unlabelled data. It accomplishes this by learning two fundamental functions: an encoding function responsible for transforming input data, and a decoding function that reconstructs the original input data from its encoded representation as shown in Figure 16. Essentially, the autoencoder strives to uncover an efficient data representation, referred to as encoding, usually for the purpose of dimensionality reduction. This process enables the network to capture meaningful patterns within the data without explicit labelling, making it a valuable tool in various machine learning tasks, including anomaly detection.

The fundamental idea is that an autoencoder learns to encode the input data into a lower-dimensional latent space and then decode it to reconstruct the original data. During training, the autoencoder becomes highly proficient at reconstructing normal data but struggles with anomalous patterns that it hasn't encountered before. When presented with new data, if the reconstruction error (typically measured as Mean Squared Error or some other loss function) is significantly higher than usual, it signals an anomaly. This is because the autoencoder, having never seen such anomalies during training, struggles to faithfully reconstruct them, resulting in a higher error. By setting a threshold on the reconstruction error, autoencoders can effectively flag unusual or anomalous data points, making them a powerful tool for detecting outliers and deviations from normal patterns in various domains, including image analysis, time series data, and more.

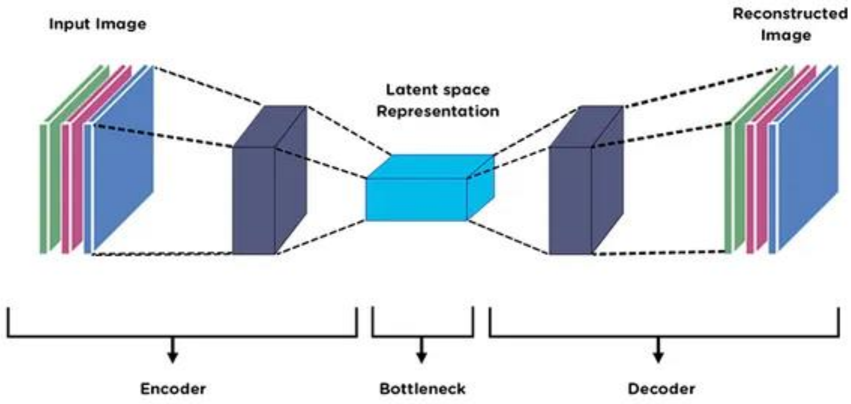


Figure 16: Basic Autoencoder Architecture [14]

3.5 Anomaly Detection with a prior distribution

Furthermore, L. Beggel, M. Pfeiffer, and their colleagues [6] introduced a strong method for spotting unusual things in images using Adversarial Autoencoder techniques. Instead of just relying on the usual way of measuring differences in the reconstructed images (Mean Squared Error), they added extra information by using a prior distribution on the hidden representation. This makes the system better at finding oddities in the pictures and makes the whole detection process more reliable.

To predict with a prior, this dissertation makes use of the kernel density estimation. Kernel density estimation (KDE) [17] is a statistical technique that uses kernel smoothing principles for the purpose of estimating probability density functions. This method is particularly valuable when dealing with non-parametric data analysis, as it allows for the estimation of the probability density function of a random variable. It achieves this by assigning weights, known as kernels, to data points and then deriving a smoothed density estimation from these weighted values.

A Gaussian curve, also known as a Gaussian distribution or normal distribution [18], is a probability distribution characterised by a symmetrical, bell-shaped curve. It's defined by two parameters: the mean (average), which marks the center of the curve, and the standard deviation, which controls its width. This distribution is widely used in statistics and data analysis because it accurately describes many natural phenomena and random processes. The 68-95-99.7 Rule associated with it highlights that about 68% of data falls within one standard deviation of the mean, roughly 95% within two standard deviations, and nearly 99.7% within three, making it a fundamental tool for modelling and understanding data variability in various fields.

Kernel Density Estimation (KDE) and the Gaussian distribution are closely linked through the choice of kernel functions. KDE uses tiny Gaussian curves centred on each data point, and when combined, these curves create a KDE estimate that often closely matches the true underlying distribution, making the Gaussian distribution a fundamental component of KDE for modelling and understanding data distributions.

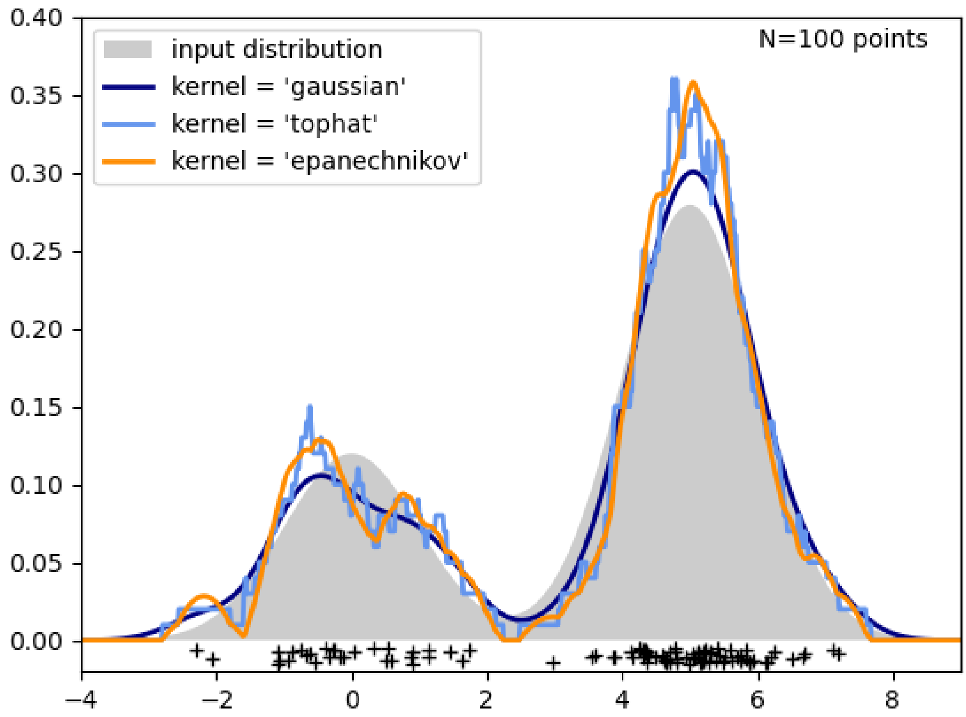


Figure 17: KDE with different kernels [19]

In Figure 17, the results of Kernel Density Estimation (KDE) are displayed, showcasing three different kernel choices applied to a sample of 100 data points originating from a bimodal distribution. It's evident that the choice of kernel shape significantly influences the smoothness of the resulting distribution. This dissertation adopts the use of KDE with a Gaussian kernel specifically in the latent space generated by the encoder of the autoencoder using the dataset for making predictions.

4 Workflow

Before delving into the intricate technical aspects, it is essential to outline the two principal workflows that guided the creation of this Dissertation, as illustrated in Figure 18. This overview will provide a foundational understanding of the structured approach taken throughout the development process.

The training pipeline serves as the foundational framework for constructing the anomaly detection model. It initiates with the dataset creation phase and seamlessly transitions into the training of the autoencoder model. Subsequently, the model undergoes fine-tuning based on observed results and errors, ultimately culminating in its deployment for the evaluation phase. This systematic approach ensures the model's development, refinement, and robust performance throughout the process.

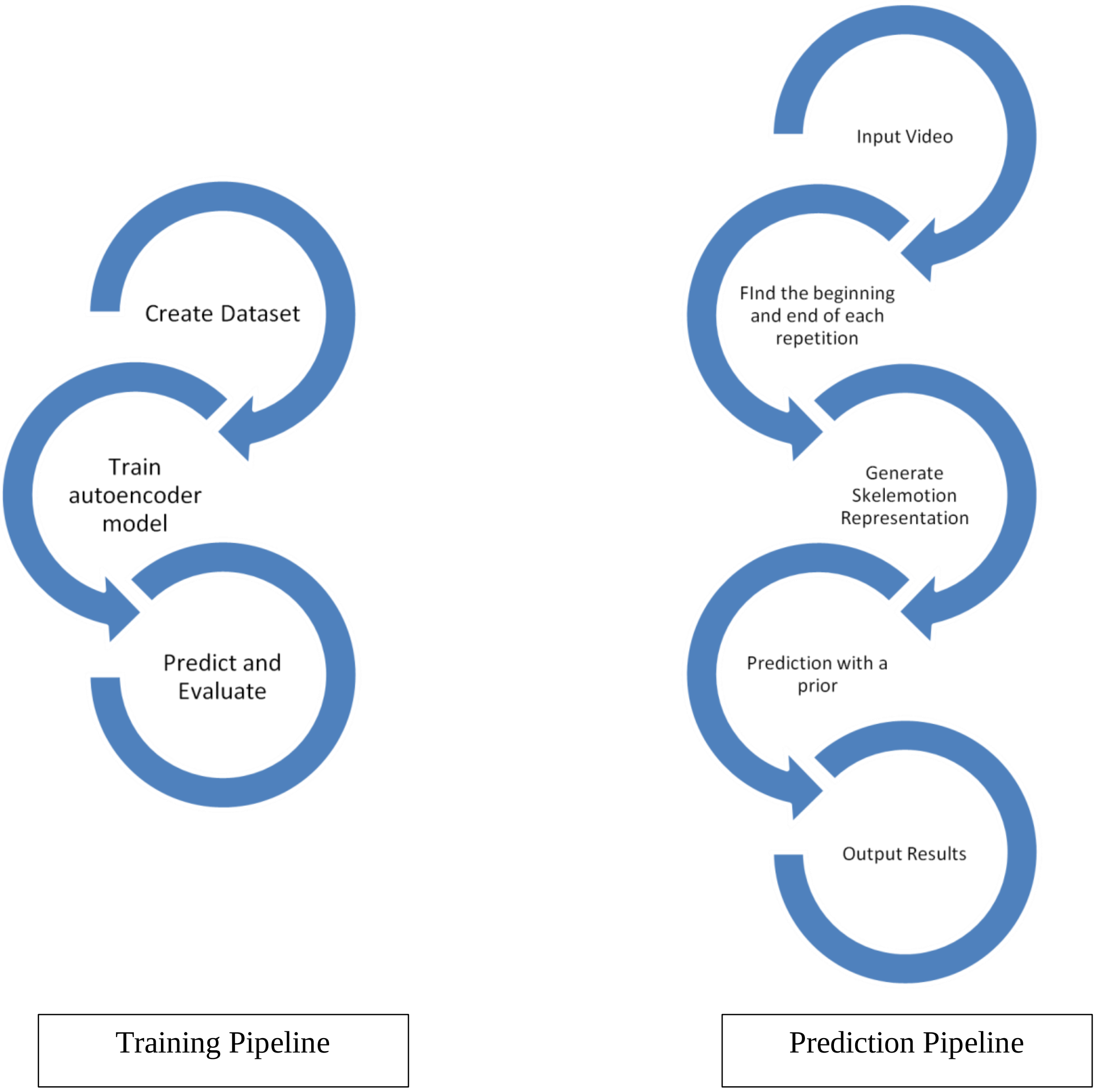


Figure 18: Training and Prediction Pipelines

In contrast, the prediction pipeline serves a distinct purpose, primarily focused on the prediction process. It commences with the analysis of input video footage, proceeds to identify the initiation and culmination of each deadlift repetition, and subsequently generates the Skelemotion representation. This representation is then fed into the pre-trained model to determine whether the repetitions qualify as anomalies or adhere to the expected pattern, leveraging a prior distribution for added reliability. The output of this prediction process is presented visually to the user, providing immediate feedback.

It's important to note that this is a high-level overview of both pipelines, and a more comprehensive exploration of its intricacies will be provided later in this report.

5 Training Pipeline

5.1 Create Dataset

The dataset to train the anomaly detector is derived from a fusion of the terminologies discussed in the previous section. The following diagram explains the workflow in creating the dataset:

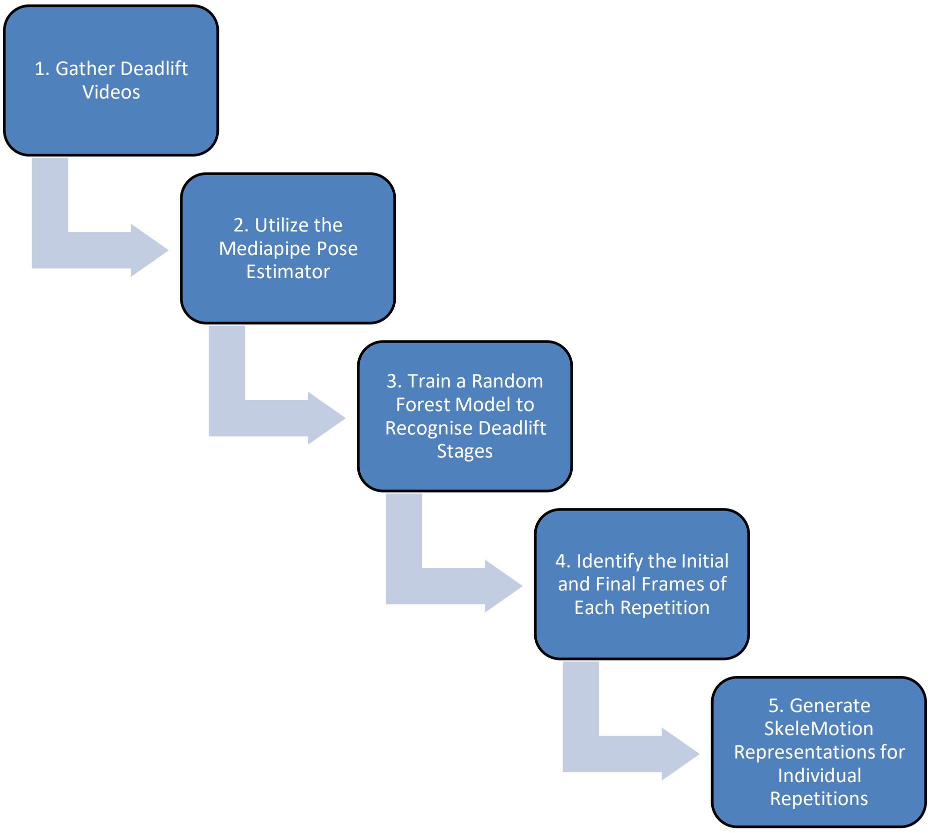


Figure 19: Dataset Workflow

5.1.1 Gather Deadlift Videos:

A total of 249 instances of correct deadlift repetitions were captured across 20 distinct videos featuring my own execution of the exercise. It's essential to highlight that the individual performing the exercise does not impact the ultimate predictions, as only the skeletal landmarks are retained for subsequent stages.

Every video was consistently recorded from a frontal perspective, employing a stable mobile camera setup. This approach was chosen deliberately to ensure comprehensive body coverage during execution. It's worth noting that individuals interested in utilizing this project can readily adopt this uncomplicated setup. The primary objective here was to maintain simplicity, thereby enabling anyone to replicate the process with ease.

Figure 20 presents a selection of sample frames extracted from the video dataset, showcasing various stages of the deadlift exercise. It's worth emphasising that, for precise skeleton keypoint detection in the subsequent Mediapipe stage, it is crucial that each frame captures the complete body, from head to toe.

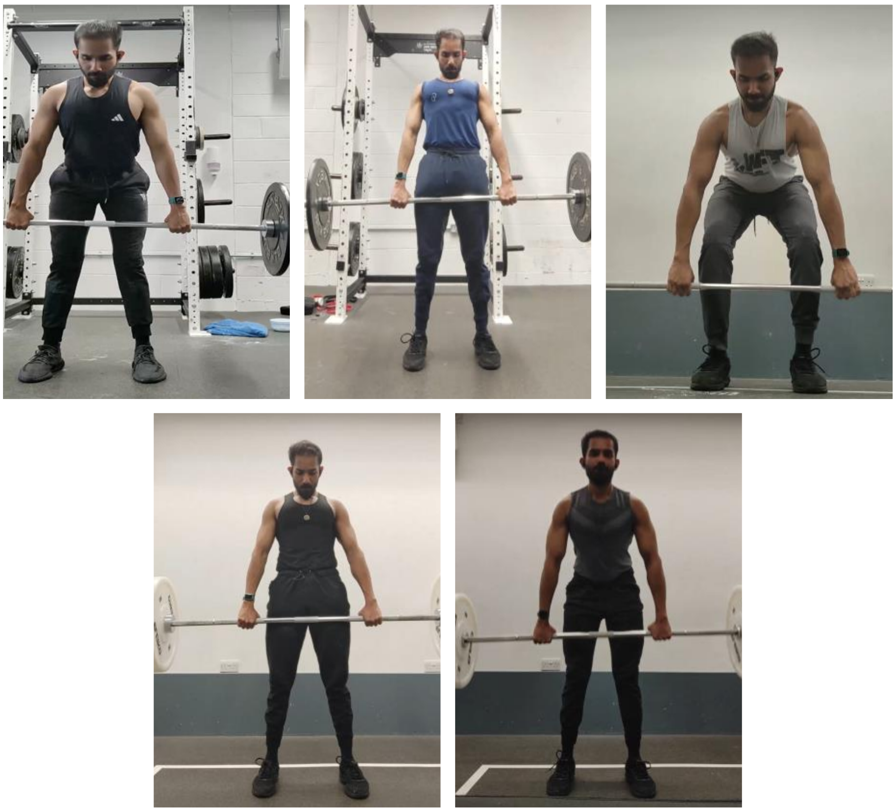


Figure 20: Snapshot of frames from the dataset

5.1.2 Utilize the Mediapipe Pose Estimator

Using the OpenCV and Mediapipe Python packages, each frame is processed, enabling the Mediapipe model to detect the 33 landmarks in each frame. To generate data for the Random Forest model to identify the three stages of the deadlift, the keypoints associated with each stage are exported into a CSV file with manual annotation. The script "Deadlift Stages/annotation.py" accomplishes this task by pressing the 'u' key for the up / top position when the lift is at its peak and the 'd' key for the down / bottom position corresponding to the frame displayed during the script execution.

The annotation script is useful when there is a need to incorporate a new exercise into the system. Figure 21 provides a glimpse of the annotation feed window for the video that requires annotation. This script accommodates two essential inputs: the path to the video file and the path to the CSV file. Furthermore, it facilitates the annotation of multiple videos into a single, consolidated CSV file. This consolidated dataset can then be used to train a model capable of learning and recognising specific poses, contributing to the versatility and adaptability of the system.

Figure 21: Annotation Feed Window

Figure 22 provides a glimpse of the CSV file, wherein the categorisation of deadlift stages is denoted in the first column. Following this classification, the subsequent columns capture the x, y, and z coordinate values of all 33 keypoints, accompanied by their respective visibility values. Notably, the class column contains only two distinct values, "up" and "down." In the ensuing phase, the Random Forest model undergoes training to comprehend these distinct stages. If the probability associated with either class is less than 0.7, the model categorizes it into the third class, denoted as the "Neutral position". This approach enhances the model's ability to discern variations within the deadlift sequence effectively.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	class	x1	y1	z1	v1	x2	y2	z2	v2	x3	y3	z3	v3	x4	y4	z4	v4	x5	y5	z5	v5
2	up	0.54613554	0.33470678	-0.3146372	0.99999988	0.55996686	0.32712853	-0.3011883	0.9999994	0.56674683	0.32791013	-0.3011541	0.99999946	0.57371598	0.32888249	-0.3011941	0.99999946	0.53894776	0.32621777	-0.3008897	0.99999923
3	up	0.54701662	0.34582698	-0.2186321	0.99999952	0.56083751	0.33616465	-0.203484	0.99999762	0.56653047	0.33629945	-0.2035013	0.99999768	0.57319534	0.33661714	-0.2035268	0.99999779	0.53862184	0.3357254	-0.2025789	0.99999738
4	up	0.54988086	0.36932984	-0.1846064	0.99999887	0.56167996	0.35777766	-0.1754354	0.99999428	0.56732559	0.35744345	-0.175468	0.99999434	0.57345045	0.35679671	-0.1755419	0.99999464	0.5395295	0.35824069	-0.1741631	0.99999362
5	up	0.55524778	0.37121853	-0.2315978	0.99999875	0.56569517	0.35972193	-0.2198686	0.9999941	0.5713712	0.35930255	-0.2198792	0.99999428	0.5784058	0.35871452	-0.2199322	0.99999446	0.54417127	0.35989401	-0.2196045	0.9999935
6	up	0.55523282	0.37135133	-0.265155	0.99999857	0.56669945	0.36020201	-0.2551708	0.99999416	0.5724749	0.35983557	-0.2551347	0.9999941	0.57960433	0.35922071	-0.2551618	0.99999446	0.54484886	0.36004117	-0.2539211	0.99999374
7	down	0.54853249	0.52314216	-0.6479468	0.99999976	0.56155723	0.51243252	-0.6514235	0.99999946	0.56980103	0.51099628	-0.6512927	0.99999958	0.57739562	0.50944054	-0.6514246	0.99999952	0.5376991	0.51196951	-0.6553379	0.9999994
8	down	0.54883575	0.54116231	-0.651974	0.99999976	0.5615654	0.52995557	-0.6541674	0.99999923	0.56949455	0.52840048	-0.6540865	0.99999952	0.57693315	0.52665603	-0.6542607	0.99999934	0.53710949	0.53000063	-0.6568188	0.99999881
9	down	0.55520242	0.56042904	-0.6657436	0.99999976	0.56738615	0.54848266	-0.6683494	0.99999911	0.57456076	0.5465107	-0.6683002	0.99999946	0.58146501	0.54422355	-0.6684871	0.99999923	0.5441587	0.54914105	-0.6699739	0.99999839
10	down	0.55089444	0.58225828	-0.7467585	0.99999863	0.56253487	0.57150447	-0.7513416	0.99999678	0.56992942	0.569731	-0.7513252	0.99999678	0.5767132	0.56758863	-0.7515566	0.99999708	0.53792781	0.57178181	-0.7513542	0.99999398
11	down	0.54155016	0.57976246	-1.0565103	0.99999374	0.55460918	0.56880528	-1.059016	0.99999201	0.56310403	0.5669626	-1.058961	0.99998951	0.57003289	0.56483161	-1.0591931	0.99999189	0.52669466	0.56858981	-1.0527805	0.99998516
12	down	0.54138446	0.57466912	-1.0972762	0.99999571	0.55370528	0.56253535	-1.0926553	0.99999493	0.5620479	0.56090248	-1.0926422	0.99999285	0.56922501	0.55883557	-1.0929393	0.9999947	0.52742797	0.56276935	-1.0888463	0.99999028
13	down	0.55456364	0.56402868	-1.250821	0.99999207	0.56675446	0.55147344	-1.2385848	0.99999171	0.57416272	0.55039471	-1.2386254	0.99998868	0.58164322	0.54892069	-1.2389667	0.99999154	0.54267806	0.55081093	-1.2340469	0.99998271
14	up	0.54959708	0.33112419	-0.2508887	0.99999946	0.56166863	0.32282564	-0.231284	0.99999851	0.56776923	0.32362205	-0.2312908	0.99999857	0.57439154	0.32474625	-0.2313087	0.99999875	0.54096961	0.32166755	-0.2319577	0.99999809
15	up	0.54617846	0.33555767	-0.2388965	0.99999946	0.55975354	0.33020192	-0.2175798	0.99999744	0.5660826	0.33079356	-0.2175555	0.99999756	0.5719319	0.33126095	-0.2175368	0.99999779	0.53886575	0.32908934	-0.217206	0.9999972
16	up	0.54796302	0.35559013	-0.145642	0.99999881	0.56123346	0.34592143	-0.1247682	0.99999505	0.56752318	0.34594625	-0.1248124	0.99999535	0.57386339	0.34586823	-0.1249021	0.99999583	0.53796589	0.34607935	-0.1247281	0.99999452
17	up	0.54753506	0.35939381	-0.1868991	0.99999678	0.56115478	0.34919685	-0.1674749	0.99998868	0.56791633	0.34921417	-0.1674742	0.99998897	0.5739997	0.34908089	-0.1675375	0.99999046	0.53702307	0.34909764	-0.1649854	0.99998784
18	up	0.54911274	0.35829601	-0.2115043	0.99999726	0.56277156	0.34804958	-0.1932494	0.99998975	0.56939906	0.34802926	-0.1932526	0.99998999	0.57581431	0.34790802	-0.1932961	0.99999112	0.53803933	0.34811449	-0.1918576	0.99998915
19	up	0.54570043	0.36194882	-0.2259835	0.99999881	0.55947912	0.35027531	-0.2204174	0.99999481	0.56669128	0.35004762	-0.220461	0.99999487	0.57257098	0.34971991	-0.2205154	0.99999529	0.53547299	0.34950316	-0.2166601	0.99999452
20	down	0.54303604	0.5215987	-0.6367443	0.9999997	0.55553073	0.50938433	-0.6362452	0.99999911	0.56397456	0.50742322	-0.6362	0.99999934	0.57075107	0.50543123	-0.6364238	0.99999928	0.53007692	0.51050735	-0.6357638	0.99999869
21	down	0.54082537	0.55331516	-0.9979694	0.99999619	0.55383861	0.54096884	-0.9950409	0.9999938	0.56245589	0.53897089	-0.9950033	0.99999285	0.56912589	0.53694177	-0.9952554	0.9999944	0.52821988	0.54211682	-0.992015	0.99998909
22	down	0.54058391	0.56920034	-1.2267642	0.99999434	0.55251533	0.55676234	-1.2171906	0.99999207	0.56094724	0.55510426	-1.2171735	0.99998993	0.56757444	0.55308211	-1.217468	0.99999261	0.52657485	0.55728328	-1.2119497	0.99998558
23	down	0.54637831	0.57372111	-1.3026508	0.99999624	0.55929977	0.56166303	-1.2904114	0.99999475	0.56684893	0.56033444	-1.2903923	0.99999303	0.57412136	0.55878276	-1.2907202	0.9999947	0.53462887	0.56184733	-1.2855406	0.99999022
24	down	0.55071092	0.57839942	-1.3818007	0.99999642	0.56318498	0.56594032	-1.3648113	0.99999458	0.57021379	0.56452972	-1.3648112	0.99999243	0.57748067	0.5628866	-1.3651702	0.9999941	0.53912169	0.56624281	-1.3595327	0.99998981
25	down	0.54891157	0.57321799	-1.3113916	0.99999529	0.56236082	0.56173819	-1.295638	0.99999219	0.57005793	0.56060928	-1.2956244	0.99998927	0.57771569	0.55910689	-1.2959501	0.99999136	0.53737915	0.56152129	-1.2904826	0.9999854

Figure 22:Annotated csv file

The CSV file contains a total of 502 rows and 133 columns. Within these columns, 210 entries correspond to the "up" class, while the remaining 292 entries pertain to the "down" class, as visually depicted in Figure 23. It's worth noting that there is minimal class imbalance, obviating the necessity for balancing procedures prior to training the Random Forest model. This balanced distribution enhances the model's ability to generalise effectively across the two distinct classes.

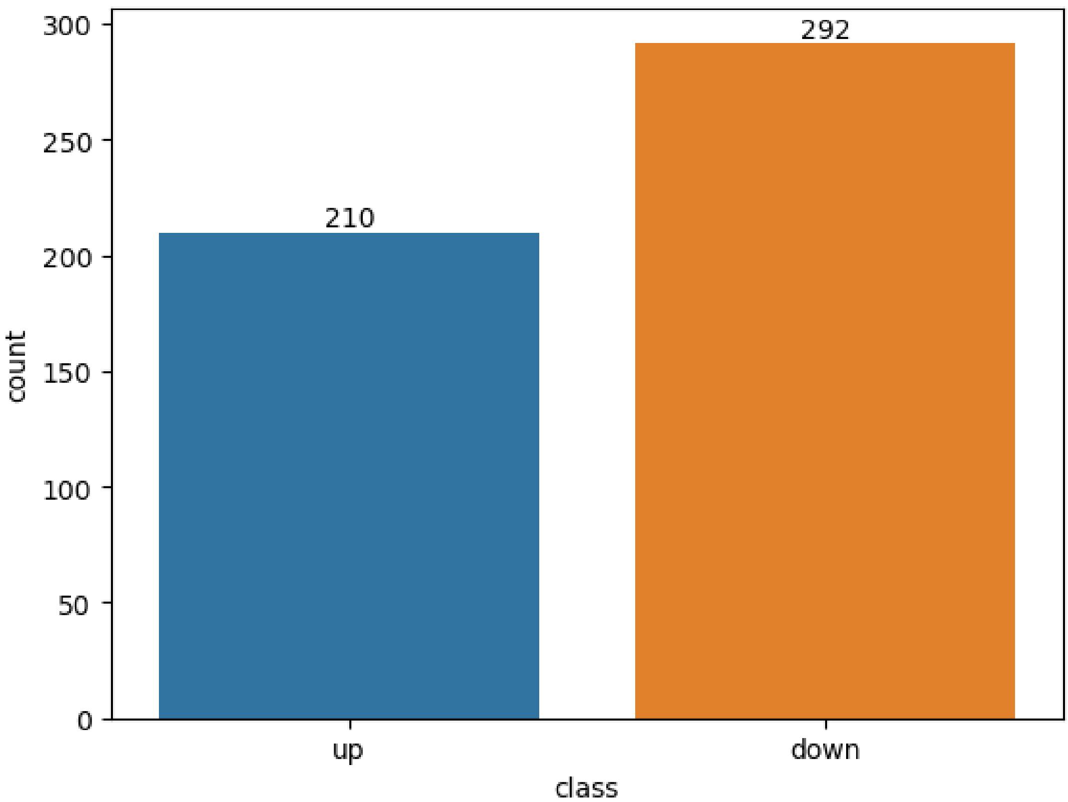


Figure 23: Class distribution of the annotated csv

5.1.3 Train a Random Forest Model to recognise Deadlift Stages

The annotated CSV data is utilised to train a Random Forest model, leveraging the capabilities of the Sklearn package. This model is trained using an 80% portion of the data, while the remaining 20% is reserved for testing and evaluation. The Random Forest model exhibits exceptional performance, boasting accuracy, precision, and recall values all at 100%. These metrics signify the model's ability to accurately identify and classify the distinct stages of the deadlift exercise, establishing it as a high-performing and reliable tool for this purpose. Table 1 presents an overview of the hyperparameters employed in the Random Forest model.

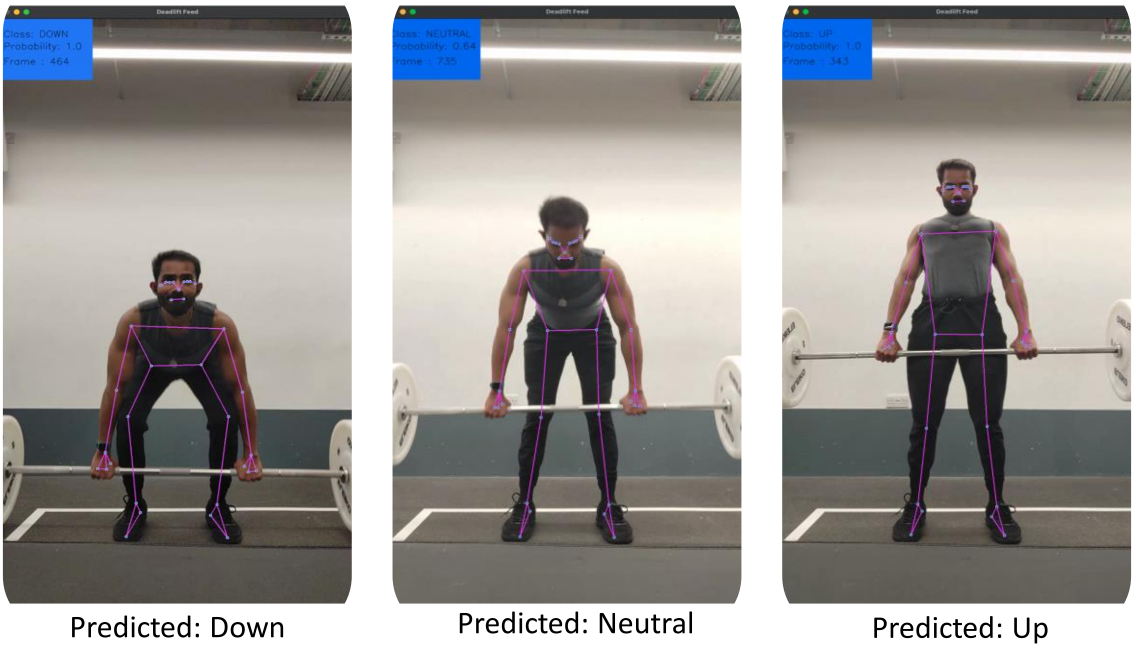


Figure 24: Random Forest predictions with Mediapipe

Figure 24 provides an insightful representation of the Random Forest model's predictions at different stages of the lift. The left blue rectangle visually displays essential information, including the class label, probability, and the corresponding frame number. It is crucial to highlight that the training of each deadlift stage was accomplished using Mediapipe coordinate values, imparting the Random Forest model with the capability to effectively recognise these distinct stages. Notably, the accuracy of Mediapipe in detecting the skeleton keypoints throughout the entire lift is evident, reaffirming its role as a reliable tool in this context.

Hyperparameters:	Values:
n_estimators:	100
min_samples_split	2
min_samples_leaf	1
max_features	'sqrt'
criterion	'gini'

Table 1: Random Forest Hyperparameters

5.1.4 Identify the Initial and Final Frames of Each Repetition

With the Random Forest model now proficient in identifying each stage of the deadlift, the next crucial step is to pinpoint the starting and ending frames of each repetition. This is an essential prerequisite for generating the SkeleMotion representation in the subsequent phase.

To achieve this, a logical sequence is developed within the Random Forest model. It actively seeks frames that commence from the "down" stage, progress through the "neutral" phase to the "up" stage, and finally return to the "down" stage, constituting a full repetition. Determining the "down" stage frame relies on assessing a reference joint value, specifically keypoint 16 (right wrist), and identifying when it reaches its maximum value, signifying the lowest point in the exercise. Simultaneously, the frame numbers associated with each repetition are calculated, enabling the recording of the starting and ending frames for each repetition, facilitating future reference and analysis.

5.1.5 Generate SkeleMotion Representations for Individual Repetitions

The final step in this data pipeline involves the generation of SkeleMotion representations for each repetition. This crucial task is accomplished by the "GenerateSkelemotion/GenerateSkeleton.py" file. By providing the folder location of the deadlift videos, the script systematically generates representation images for all repetitions within that folder. It's noteworthy that the Orientation (Image Type 2) representation [7] is employed, specifically emphasizing the orientation of joint movements throughout the lift.

Figure 25 provides a glimpse of randomly selected sample SkeleMotion Orientation representation images for each repetition. These image representations serve as the foundational dataset for training the anomaly detector, forming a crucial component in the subsequent phases of the project.

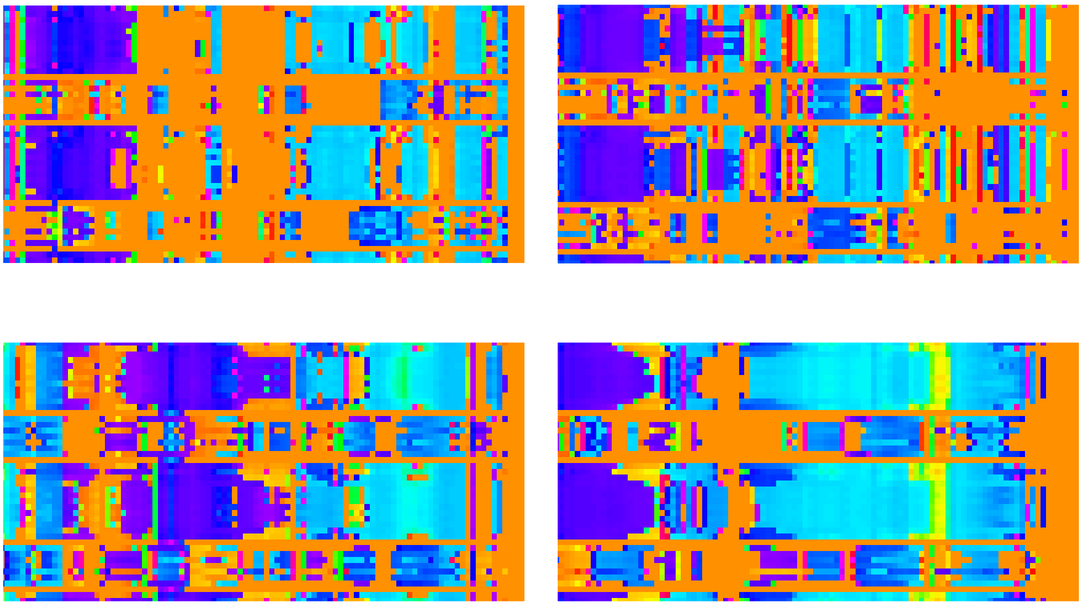


Figure 25:Sample Skelemotion Image representation of the deadlift movement

5.1.6 EDA with t-SNE

The next step involves analysing the representation to discern any distinguishable patterns between anomalies and normal deadlifts. For this comparison, a dataset comprising incorrectly performed deadlifts was collected, albeit in a limited quantity, for the purpose of visual differentiation.

To carry out this analysis, a pretrained ResNet-18 model [12] is used to extract features from the representations. These extracted features are then flattened into a 4096-dimensional vector for each image. Subsequently, this 4096-dimensional feature set undergoes dimensionality reduction using t-SNE (n_components = 2, perplexity = 10), transforming it into a more manageable 2-dimensional format. This reduction simplifies the visualisation process, making it easier to discern patterns and differences between the representations of normal and anomalous deadlifts.

In Figure 26, the blue cluster represents the normal deadlift data, while the orange data points represent anomalies. This visual representation clearly illustrates a significant distinction in the data space between normal and anomalous representations. This compelling evidence reinforces the decision to proceed with the current dataset of deadlifts, as it underscores features in differentiating between the two.

[85]: <AxesSubplot:xlabel='0', ylabel='1'>

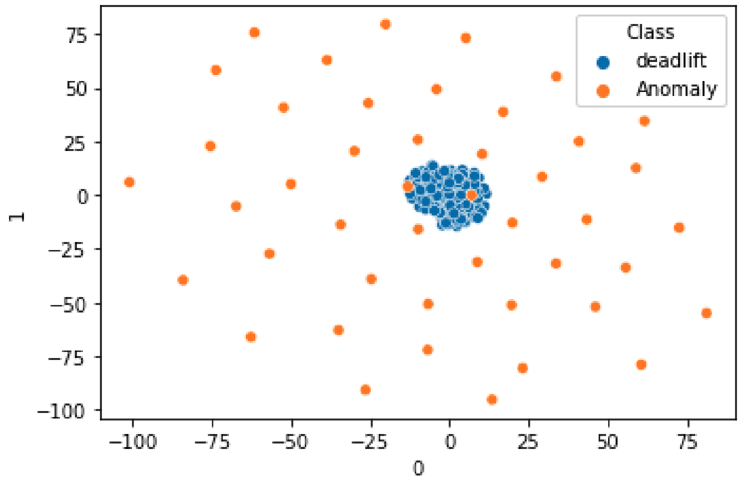


Figure 26: t-SNE plot of normal and anomalous deadlifts

It's important to note that a perplexity value of 10 was selected for this visualisation. Perplexity determines the number of nearest neighbours considered during the t-SNE dimensionality reduction process, aiding in the optimal visualisation of data patterns.

5.2 Autoencoder Model

The dataset is loaded with the specified split, as outlined in Table 2, where 80% of the data is designated for training, and the remaining 20% is reserved for validation purposes. To facilitate the construction of the anomaly detector, each image is resized to a consistent dimension of 256 x 256 pixels.

Split	Images
Training	199
Validation	50

Table 2: Training Validation Data Distribution

Several experiments were conducted on the dataset, exploring various convolutional and filter sizes. The goal was to identify an architecture for the autoencoder that could consistently produce optimal results. Figure 27 provides an in-depth depiction of the encoder-decoder architecture utilized for anomaly detection in the context of the deadlift exercise. This architectural design is crucial in capturing and distinguishing patterns within the data, ultimately enabling effective anomaly detection.

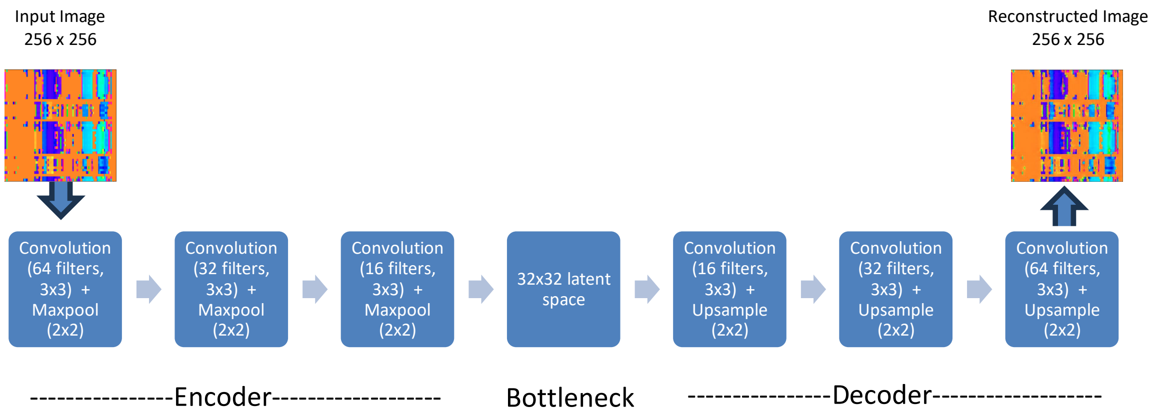


Figure 27: Autoencoder Architecture

5.2.1 Encoder

The encoder module comprises three sequential convolution operations, each employing 64, 32, and 16 filters of size 3x3. Following each convolution operation, a max-pooling layer is applied, reducing the image dimensions by half (2x2). This process effectively encodes the input image into a compact 32 x 32 latent space. Subsequently, this bottleneck latent space serves as the input for the decoder module, allowing for the reconstruction of the original image from this condensed representation.

5.2.2 Decoder

The decoder module mirrors the architecture of the encoder, featuring convolutional filters of 16, 32, and 64. In contrast to the encoder, the decoder employs upsampling techniques, effectively doubling the image dimensions after each convolution operation. The primary objective of the decoder is to reconstruct the original image from the condensed 32x32 latent space, ensuring that the essence of the input data is preserved throughout this reconstruction process. This bidirectional architecture plays a pivotal role in the autoencoder's ability to efficiently encode and decode data representations.

5.2.3 Loss Function and Optimisation

The loss function adopted for this autoencoder is the Mean Squared Error (MSE) [16], and to optimise the learning process, the Adam optimizer [15] is employed. The primary goal of the optimizer is to minimize the pixel-wise MSE discrepancy between the original image and its reconstructed counterpart. As the model advances during the training phase, the encoder evolves in its capacity to proficiently compress the input image into the 32 x 32 latent space, while the decoder concurrently refines its ability to recreate the original image from this compact latent representation. This progressive learning mechanism results in the autoencoder effectively capturing essential data features and reproducing them with fidelity during the reconstruction process.

5.2.4 Training and Validation

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 256, 64)	1792
max_pooling2d (MaxPooling2D)	(None, 128, 128, 64)	0
conv2d_1 (Conv2D)	(None, 128, 128, 32)	18464
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 32)	0
conv2d_2 (Conv2D)	(None, 64, 64, 16)	4624
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 16)	0
conv2d_3 (Conv2D)	(None, 32, 32, 16)	2320
up_sampling2d (UpSampling2D)	(None, 64, 64, 16)	0
conv2d_4 (Conv2D)	(None, 64, 64, 32)	4640
up_sampling2d_1 (UpSampling2D)	(None, 128, 128, 32)	0
conv2d_5 (Conv2D)	(None, 128, 128, 64)	18496
up_sampling2d_2 (UpSampling2D)	(None, 256, 256, 64)	0
conv2d_6 (Conv2D)	(None, 256, 256, 3)	1731

Total params: 52067 (203.39 KB)
Trainable params: 52067 (203.39 KB)
Non-trainable params: 0 (0.00 Byte)

Figure 28: Autoencoder Architecture

The 7-layered autoencoder model has a total of 52,067 trainable parameters, as depicted in the Figure 28. It underwent an extensive training regimen, spanning 500 epochs, utilising batches of 4 for training and incorporating 50 validation images. Over the course of these 500 epochs, the model exhibited a reduction in the Mean Squared Error (MSE) loss metric, dropping from an initial value of 0.09 to an optimised 0.0079 for the training set. Impressively, the validation loss converged to an equally low value of 0.0103 as shown in the Figure 29. This substantial reduction in loss signifies the model's ability to effectively capture and recreate the essential characteristics of the input data during both training and validation, attesting to its robust performance.

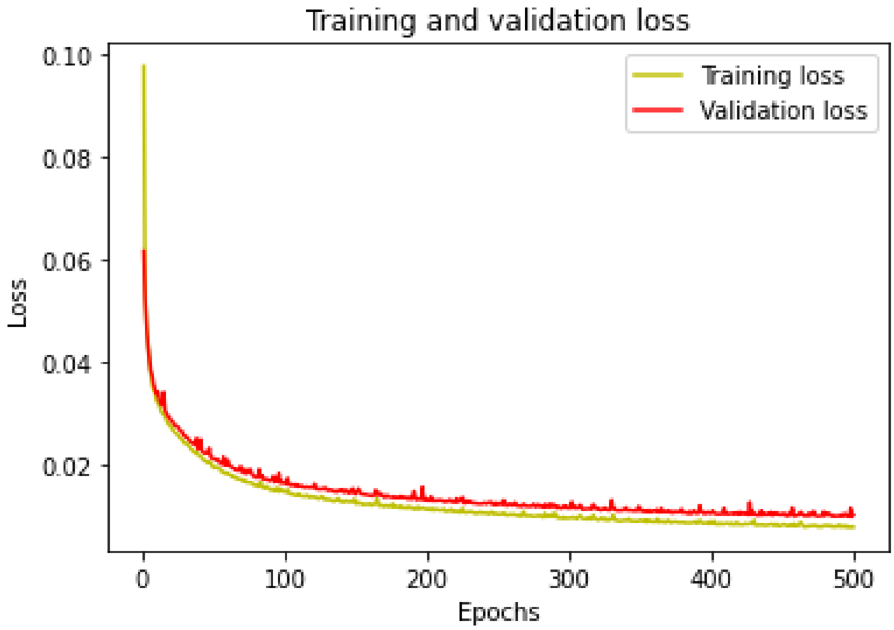


Figure 29: Training and Validation Loss Curve

5.2.5 Reconstruction

To assess the model's reconstruction capability, a random image from the validation set was chosen and presented to the autoencoder for reconstruction, as demonstrated in Figure 30. The original image appears on the left, while the reconstructed image is displayed on the right. Notably, the reconstructed image closely mirrors the original, showcasing a high level of fidelity and accuracy in the model's ability to capture and replicate the underlying patterns within the training data. This successful reconstruction provides strong assurance that the model effectively learned and retained the intricate data patterns during the training process, thus validating its proficiency.

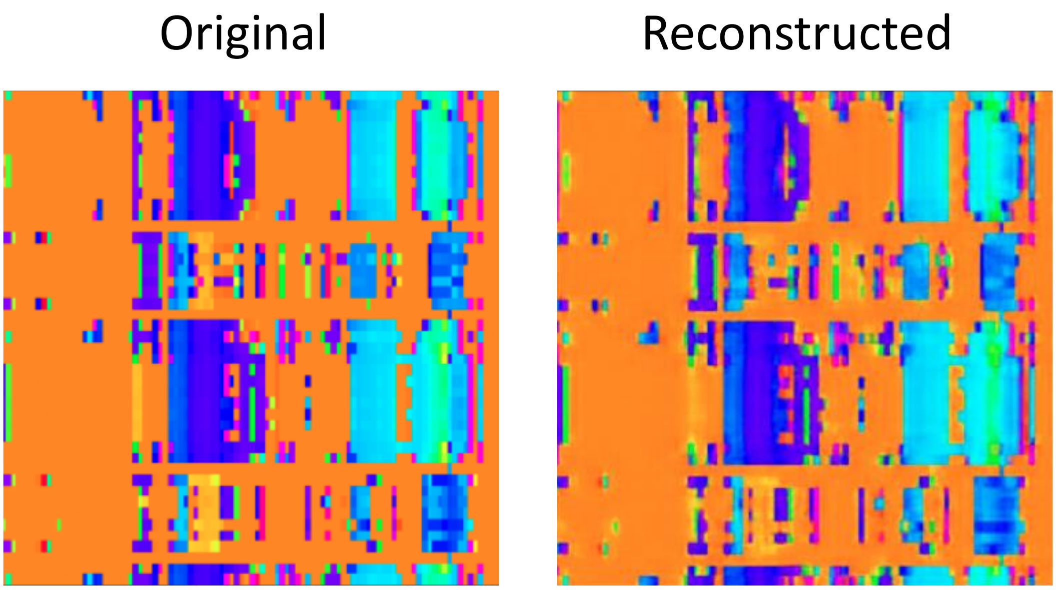


Figure 30: Reconstruction from Autoencoder

5.3 Predict and Evaluate

5.3.1 Prediction with a prior

As highlighted earlier, relying solely on the mean squared error (MSE) difference between pixels in the original and reconstructed representations may not be the most effective approach for detecting anomalies, especially when dealing with subtle deviations. An alternative method, suggested by L. Beggel, M. Pfeiffer, and their colleagues [6], involves incorporating a prior distribution into predictions by considering the latent space during prediction and comparing it to a prior distribution, such as a Gaussian distribution fitted on the training set. This approach enhances the reliability of predictions.

Metric	Threshold
Density:	-240000
Reconstruction Error:	0.032

Table 3: Anomaly Thresholds

When an image is submitted for prediction, tasked with determining whether it represents an anomaly or a proper deadlift form, the autoencoder's encoder module converts the image into the 32 x 32 latent space. The prediction process then involves examining a Kernel density estimation of the latent spaces from the training samples, yielding a density value for the image under evaluation. To make a prediction, this density value is tested against a predefined threshold for anomalies. The function classifies the image as an anomaly if its predicted density falls below the threshold, and if the reconstruction error (MSE) exceeds another threshold, as depicted in the Table 3. Conversely, if these conditions are not met, the function classifies the test image as a representation of a proper deadlift form. This combined approach ensures more robust anomaly detection in comparison to relying solely on MSE.

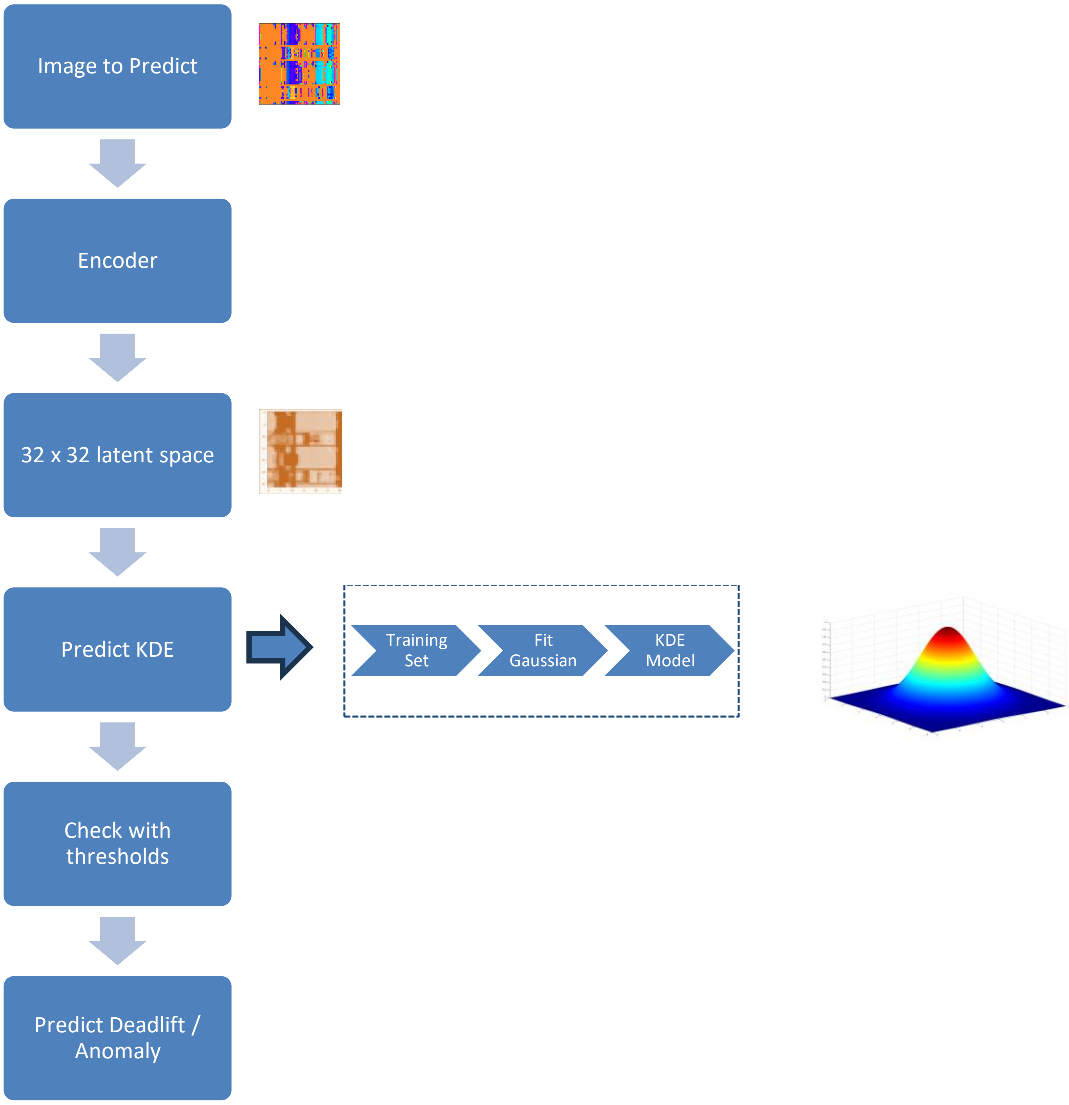


Figure 31: Prediction Workflow

5.3.2 Testing and Evaluation

To facilitate testing and evaluation, additional video footage was collected and subsequently subjected to the pipeline, resulting in the creation of Skelemotion representations for these movements. In total, this supplementary dataset comprises 29 instances of proper deadlift forms and 21 instances of anomalous actions. This augmented dataset provides a diverse range of scenarios, enhancing the robustness of the testing and evaluation process by encompassing various real-world situations and variations in movement as shown in the Figure 32.

Proper Deadlift Form

Anomaly created in a different setting

Proper deadlift form in a different setting

Figure 32: Test Set Samples

To maintain a level of comfort and privacy during the creation of anomaly instances, all anomalous actions were recorded in a separate, controlled setting, thereby eliminating the potential embarrassment associated with generating anomalies in a public space. This separation in recording settings was a deliberate measure to prevent bias during testing. It's noteworthy that all proper deadlift forms recorded in this distinct setting were accurately detected as good deadlift forms by the model.

The confusion matrix depicted in Figure 33 provides a comprehensive overview of the model's performance on this specific test set. It's crucial to acknowledge that anomalies can manifest in various ways, and due to the constraints of conducting exhaustive testing, these results may exhibit variation with different test sets. Therefore, these findings serve as a general performance indicator, recognising the potential for variations in results based on the nature and diversity of anomalies encountered.

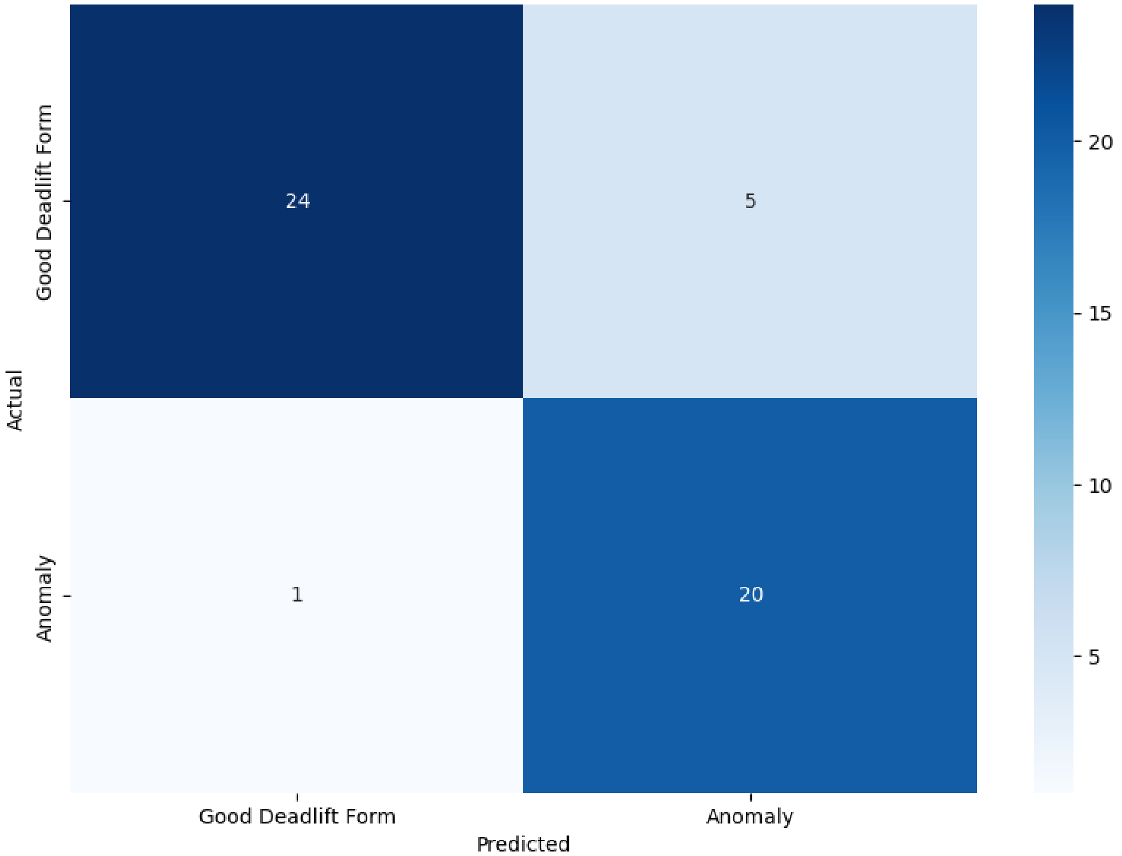


Figure 33: Confusion Matrix of the test set

Referring to the confusion matrix, the model demonstrated a commendable performance by correctly identifying 24 instances of good deadlift forms, albeit with a slight misclassification of 5 cases. Furthermore, it effectively detected 20 anomalies, with only 1 instance incorrectly classified as a proper form. For a more comprehensive understanding of the model's performance, the classification report depicted in Figure 34 provides additional insights and detailed metrics that elucidate its efficiency in distinguishing between good deadlift forms and anomalies.

	precision	recall	f1-score	support
Good Deadlift Form	0.96	0.83	0.89	29
Anomaly	0.80	0.95	0.87	21
accuracy			0.88	50
macro avg	0.88	0.89	0.88	50
weighted avg	0.89	0.88	0.88	50

Figure 34: Classification Report of the predictions

The classification report offers a comprehensive breakdown of the model's performance on the test set. Notably, the model achieved an impressive overall accuracy rate of 88%, and it exhibited high precision, recall, and f1-score values in the 80s and 90s range. These metrics collectively provide a detailed snapshot of the model's performance on the test data.

However, it's important to emphasise that these results, while encouraging, do not provide an absolute guarantee of the model's capabilities in distinguishing all anomalies in real-world scenarios. This dissertation primarily serves as a case study aimed at learning and showcasing the technologies employed, rather than striving to build a state-of-the-art anomaly detection method. Consequently, the model's performance in practical, diverse settings may vary, and its generalisability to real-world applications warrants further exploration and refinement.

6 Prediction Pipeline

Having explored the theoretical foundations and operational intricacies that form the backbone of this dissertation, it is now an apt moment to elucidate the prediction pipeline. To initiate this pipeline, one can simply execute the "Predict/predict.py" script, providing the path to the video file as an input parameter.

6.1 Input Video

In the prediction pipeline, the input video undergoes a frame-by-frame processing procedure. Each frame is passed through the Mediapipe framework, which detects the skeleton keypoints within the frame. These detected keypoints are subsequently fed into the trained random forest model, tasked with identifying the specific deadlift stage represented by each frame.

As previously discussed, each repetition of the deadlift commences from the bottom position, progresses through the neutral phase to reach the top position, and then returns to the down stage, thereby completing a full repetition. During this process, the skeleton keypoints, along with the random forest's stage prediction, are recorded in a structured dataframe. Each row within this dataframe corresponds to a frame from the video, and it includes valuable information such as the predicted stage and its associated probability score.

6.2 Beginning and end of each repetition

Employing the same techniques utilised during the dataset creation process, I leverage keypoint 16, corresponding to the right wrist, as a reliable reference joint. This reference point assists in determining the lowest bottom position within the entire movement involved in a repetition. By identifying this point, I establish the starting frame of a repetition. Similarly, the process continues by tracing the movement from neutral to up, back to neutral, and eventually to down, enabling me to pinpoint the lowest down position. This comprehensive analysis facilitates the identification of both the starting and ending frames for each repetition, effectively segmenting the video into distinct, analysable units.

6.3 Generate Skelemotion representation

Subsequently, the frames spanning from the identified starting frame to the ending frame are fed into the Skelemotion representation creation module. This module generates a unique representation image for each repetition, building upon the insights gained from the previous section. These newly created images are then organised within a designated folder, conveniently located within the provided file path. These images serve as crucial inputs, propelling the pipeline further towards its ultimate objective.

6.4 Prediction with a prior

Within the prediction pipeline, the pre-trained autoencoder model is loaded, specifically leveraging its encoder component. In this phase, the encoder model is loaded and employed to predict encoded image vectors on the training dataset, similar to the prediction process expounded in the training pipeline. These generated encoded image vectors are then fitted into a Gaussian Kernel Density Estimation (KDE) module.

Simultaneously, the images produced in the preceding step are submitted to the autoencoder, resulting in the calculation of a reconstruction error for each image. Meanwhile, the KDE module computes a score for these new images, grounded in the training dataset. This dual approach involves comparing both the reconstruction error and the KDE-derived scores to their corresponding thresholds, as outlined in the training pipeline. This comparative evaluation decides whether each repetition adheres to the expected pattern or exhibits anomalies, thereby detecting anomalies within the newly generated images.

6.5 Output Result Video

Once the prediction is performed, the module automatically displays the output as shown in Figure 35, while also saving the result video into the same folder location. The blue rectangular box at the top left side of the frame provides the following information:

- *Stage*: Current stage of the deadlift
- *Probability:* Probability of being in the UP or DOWN position of the lift
- *Frame:* The frame number being displayed on the screen
- *Rep:* The current repetition of the lift

The top right rectangle furnishes one of the following details:

- A **Green** Box signifies that no anomaly was detected in that particular repetition, affirming it as a 'Good Deadlift Form.'
- A **Red** Box indicates that an anomaly was detected in that specific repetition, categorising it as an 'Anomalous' repetition.

This output video offers a visual understanding for the user to analyse the deadlift form, pinpointing any repetitions that deviated from the standard norms of the exercise.

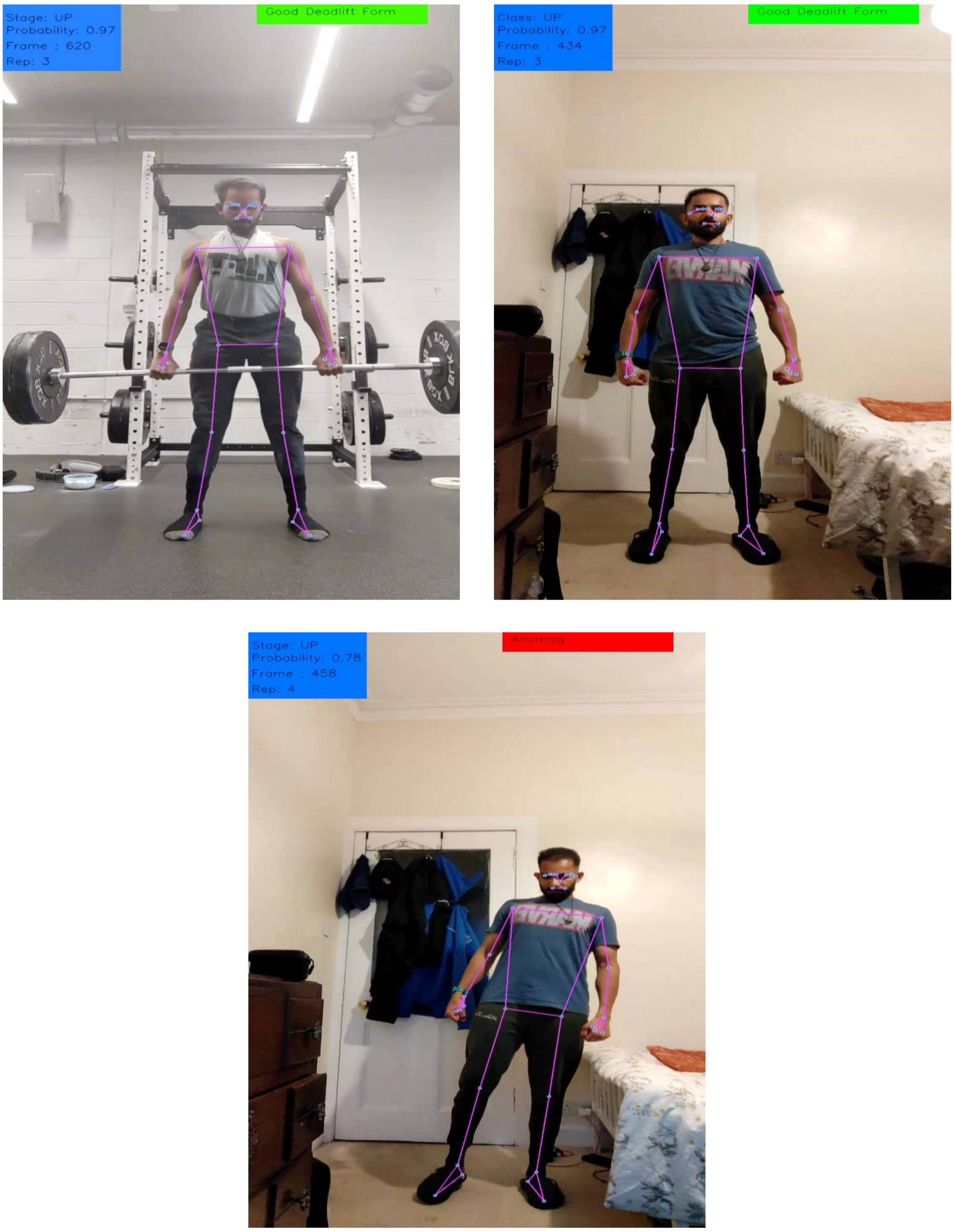


Figure 35: Prediction Result Samples

6.6 Time statistics for prediction

Understanding the time efficiency of the prediction pipeline in processing videos, especially after analysing each repetition and presenting results, is important. For this evaluation, the experiment was conducted on a personal computer with the following specifications:

- *Computer:* Apple's MacBook Pro
- *Processor:* Apple M1
- *RAM:* 8GB

- *Operating System:* MacOS Ventura 13.4.1

A random test video was chosen to assess the pipeline's performance, yielding the subsequent findings:

- *Number of frames:* 1841
- *Number of exercise repetitions:* 10
- *Total time taken:* 1 minute.
- *Average Time to process a frame*: 0.0326 seconds.

These results shed light on the efficiency of the prediction pipeline in handling video processing tasks, providing valuable insights into its real-world performance. It's essential to underscore that these time metrics serve as general guidelines and may exhibit variations across different systems, particularly those equipped with robust GPU computing capabilities. System-specific factors, such as hardware specifications and computing resources, can significantly influence the processing speed of the prediction pipeline.

7 Conclusion

In the era of technological advancements, the fusion of artificial intelligence and computer vision has opened up exciting possibilities across various domains. This dissertation embarked on a journey to explore the application of these cutting-edge technologies in the realm of exercise form analysis, specifically focusing on the deadlift exercise. This concluding chapter encapsulates the key findings, accomplishments, limitations, and avenues for future work, drawing together the threads of this dissertation's narrative.

7.1 Accomplishments and Contributions

The primary motivation behind this project was twofold: a passion for fitness and a profound interest in the capabilities of artificial intelligence. Over the course of this dissertation, significant accomplishments have been achieved:

1. *Skelemotion Representation:* A novel concept named "Skelemotion" representation was introduced by C. Caetano etc all [7], which captures the temporal dynamics of skeleton keypoints during a physical exercise. This representation proved to be instrumental in analysing exercise form.
2. *Autoencoder-based Anomaly Detection:* An autoencoder-based anomaly detection model was designed to assess the quality of each repetition within the deadlift exercise. This model learned to distinguish between correct and anomalous repetitions, providing a valuable tool for users to enhance their exercise form.
3. *Integration of Computer Vision:* Pose estimation techniques were employed to extract skeleton keypoints from video frames. These keypoints served as the foundational data for Skelemotion representations and anomaly detection.
4. *Predictive Pipeline:* A prediction pipeline was devised, allowing users to receive feedback on their exercise form. The pipeline analyses video input, identifies the stages of the exercise, and classifies each repetition as Good Deadlift Form or Anomalous Form.
5. *Statistical Analysis:* Techniques such as t-SNE and Kernel Density Estimation (KDE) were employed for exploratory data analysis and anomaly score calculation. These statistical methods provided insights into the distribution of data points and the identification of anomalies.
6. *Performance Evaluation:* The autoencoder-based model exhibited an overall accuracy of 88% on a randomly selected test set. This performance, while promising, leaves room for further improvement with the expansion of the dataset and model refinements.

7.2 Limitations and Lessons Learned

While this dissertation has made substantial strides in the domain of exercise form analysis, several limitations and challenges were encountered:

1. *Data Variability*: The system's performance heavily relies on the quality and diversity of the training data. Due to the time limitation in completing the project, the model is built on a limited amount of data to test the feasibility of the approach. Expanding the dataset

with more variations in body types, fitness levels, and environmental conditions could enhance its robustness.

2. *Small Anomalies Undetected:* The rounding of the back is a subtle change in the deadlift form where the lifter's back is slightly rounded during the lift. While it was expected to be categorised as an anomaly, the model primarily identifies deviations from the ideal form. Detecting such nuanced variations, like subtle back rounding, might require more complex models or specialised anomaly detection techniques. Figure 36 shows the example of undetected rounded back.

Figure 36: Undetected Rounded Back

3. *No Real time Feedback:* This approach currently lacks the capability of providing real time feedback. The process of identifying the start and end of each repetition to create the Skelemotion representation introduces a delay. To offer immediate feedback, optimisations in data processing and model inference speed would be necessary.

4. *Dependency on Skeleton Keypoints*: The accuracy of the approach heavily relies on the precision of the skeleton keypoints provided by the pose estimation model. Inaccurate or missing keypoints can result in degraded system performance. Enhancing the pose

estimation accuracy or implementing redundancy checks for keypoint validation could mitigate this limitation.

5. *Resource Intensive:* Running the prediction pipeline on low-end devices or in resource-constrained environments might not be feasible due to the high computational demands of computer vision and machine learning models. Exploring model optimisation techniques and lightweight architectures could make the system more accessible to a broader user base.

7.3 Future Scope

The journey does not conclude here; rather, it opens doors to exciting future possibilities:

1. *Dataset Expansion:* Expanding the dataset's breadth and depth is paramount. Including a wider spectrum of exercises, encompassing various fitness levels, and accounting for diverse environmental conditions will enable the system to generalise effectively across a myriad of real-world scenarios.
2. *Model Refinements:* The heart of the system, the autoencoder-based anomaly detection model, is a dynamic entity that demands continual refinement. Staying abreast of advancements in generative modelling and incorporating novel techniques can push the boundaries of the system's accuracy and robustness.
3. *Real-time Feedback:* The pursuit of instantaneous feedback is an unceasing goal. To achieve this, optimising the prediction pipeline to minimise delays and exploring hardware acceleration options are essential steps in the system's evolution.
4. *Mobile Application:* A dedicated mobile application is a natural progression. Its development can democratise the benefits of exercise form analysis, placing them at the fingertips of a global audience. With mobile devices being ubiquitous, such an application has the potential to revolutionise fitness monitoring.
5. *Community Involvement:* Involving the fitness community in dataset augmentation and system enhancement is a collaborative approach that could yield tremendous results. Crowdsourced data and collective intelligence can lead to more accurate analysis and a stronger sense of ownership among users.
6. *Generalization to Other Exercises:* Extending the system's capabilities to assess a broader spectrum of exercises is a transformative step. From squats to yoga poses, the system could become an all-encompassing fitness companion, catering to the diverse needs and interests of fitness enthusiasts.
7. *Clinical Applications:* Exploring the application of this technology in clinical settings, particularly in the fields of physical therapy and rehabilitation, has immense potential. It can contribute to more efficient and data-driven healthcare solutions.
8. *Gamification and Engagement:* Leveraging gamification principles can incentivise users to maintain proper form and stay motivated in their fitness journey. Incorporating features like leaderboards, challenges, and rewards, can foster a sense of community and competition among users.

In conclusion, this dissertation represents the confluence of passion, technology, and innovation. It symbolises the fusion of personal fitness aspirations and the transformative potential of artificial intelligence. While it may not lay claim to the title of a state-of-the-art solution, it serves as a foundation, a case study, and a trailblazing step towards a healthier, more informed future.

The journey has illuminated the path to a novel realm of exercise form analysis, characterised by Skelemotion representations, specialized datasets, and autoencoder-based anomaly detection. It is a testament to the vision of leveraging technology to minimise the risk of injuries, democratise fitness knowledge, and empower individuals to attain their fitness goals.

As the digital and physical worlds continue to converge, the possibilities are infinite. This dissertation represents not just an end point but a launching pad into a future where AI-driven exercise form analysis becomes an integral part of the fitness landscape. The path forward is illuminated, and the journey continues with unwavering determination towards a future where fitness is safer, smarter, and more accessible to all.

References

[1] A. Grice, S. R. Kingsbury, and P. G. Conaghan, "Nonelite exercise-related injuries: Participant reported frequency, management and perceptions of their consequences," Scandinavian Journal of Medicine & Science in Sports, vol. 24, no. 2, pp. e86–e92, Apr. 2014. Link: https://doi.org/10.1111/sms.12115

[2] "Paul O'Neill CEO of Alcoa - It's all about safety," www.youtube.com. Link: https://www.youtube.com/watch?v=tC2ucDs_XJY

[3] MobiDev, "3D Human Pose Estimation in AI Fitness Coach Apps," The Startup, Jan. 18, 2021. Link: https://medium.com/swlh/3d-human-pose-estimation-in-ai-fitness-apps-670d93d453a9

[4] "AI Learns to Do Deadlifts," www.youtube.com. Link: https://www.youtube.com/watch?v=H7cGq0xIHbc

[5] nicknochnack, "MediaPipePoseEstimation/Media Pipe Pose Tutorial.ipynb at main · nicknochnack/MediaPipePoseEstimation," GitHub, 2021. Link: https://github.com/nicknochnack/MediaPipePoseEstimation/blob/main/Media%20Pipe%20Pose%20Tutorial.ipynb

[6] L. Beggel, M. Pfeiffer, and B. Bischl, "Robust Anomaly Detection in Images using Adversarial Autoencoders," arXiv.org, Jan. 18, 2019. Link: https://arxiv.org/abs/1901.06355

[7] C. Caetano, J. Sena, F. Brémond, J. A. dos Santos, and W. R. Schwartz, "SkeleMotion: A New Representation of Skeleton Joint Sequences Based on Motion Information for 3D Action Recognition," arXiv.org, Jul. 30, 2019. Link: https://arxiv.org/abs/1907.13025

[8] "Pose landmarks detection task guide | MediaPipe," Google Developers. Link: https://developers.google.com/mediapipe/solutions/vision/pose_landmarker

[9] "sklearn.manifold.TSNE — scikit-learn 0.21.3 documentation," Scikit-learn.org, 2014. Link: https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html

[10] Wikipedia Contributors, "t-distributed stochastic neighbor embedding," Wikipedia, Jul. 04, 2019. https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding

[11] Wikipedia Contributors, "MNIST database," Wikipedia, Feb. 22, 2019. Link: https://en.wikipedia.org/wiki/MNIST_database

[12] "resnet18 — Torchvision main documentation," pytorch.org. Link: https://pytorch.org/vision/main/models/generated/torchvision.models.resnet18.html

[13] Wikipedia Contributors, "Autoencoder," Wikipedia, May 18, 2019. Link: https://en.wikipedia.org/wiki/Autoencoder

[14] D. Birla, "Autoencoders," Medium, Mar. 12, 2019. Link: https://medium.com/@birla.deepak26/autoencoders-76bb49ae6a8f

[15] D. Kingma and J. Lei Ba, "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION," Jan. 2017. Link: https://arxiv.org/pdf/1412.6980.pdf

[16] Wikipedia Contributors, "Mean squared error," Wikipedia, Mar. 30, 2019. Link: https://en.wikipedia.org/wiki/Mean_squared_error

[17] Wikipedia Contributors, "Kernel density estimation," Wikipedia, Sep. 11, 2019. Link: https://en.wikipedia.org/wiki/Kernel_density_estimation

[18] Wikipedia Contributors, "Normal distribution," Wikipedia, Nov. 12, 2019.Link: https://en.wikipedia.org/wiki/Normal_distribution

[19] "2.8. Density Estimation," scikit-learn. Link: https://scikit-learn.org/stable/modules/density.html

[20] "Build an AI Fitness Trainer using MediaPipe to Analyze Squats," learnopencv.com, Dec. 20, 2022. Link: https://learnopencv.com/ai-fitness-trainer-using-mediapipe/

[21] Y. VanDerPloeg, "Wearable Tech - the new generation of performance tools," BioLAB, Jun. 12, 2020. Link: https://www.biolab.vic.edu.au/wearable-tech-the-new-generation-of-performance-tools/

[22] "FitnessAI — Get Stronger, Faster with Artificial Intelligence," www.fitnessai.com. Link: https://www.fitnessai.com

[23] L. Voutik, "Make Your Fitness App More Engaging with Gamification," Quytech Blog, Jan. 11, 2023. Link: https://www.quytech.com/blog/fitness-app-with-gamification/

[24] Frontiersin.org, 2018. Link: https://www.frontiersin.org/files/Articles/329783/fphys-09-00743-HTML-r1/image_m/fphys-09-00743-g001.jpg

[25] Onix-Systems, "How VR/AR revolutionized the fitness industry?," Medium, Aug. 04, 2023. Link: https://onix-systems.medium.com/how-vr-ar-revolutionized-the-fitness-industry-c8551e77dd4a

[26] "2.7. Novelty and Outlier Detection — scikit-learn 0.23.2 documentation," scikit-learn.org. Link: https://scikit-learn.org/stable/modules/outlier_detection.html

[27] “Home | OnForm - Video Analysis App and Coaching Platform,” OnForm. Link: https://www.onform.com

[28] “Exer Health - Objective motion health data.,” Link: www.exer.ai. https://www.exer.ai

[29] “FormCheck - AI Video Analysis,” App Store, Dec. 24, 2021. Link: https://apps.apple.com/ca/app/formcheck-ai-video-coaching/id1563633996

[30] carloscaetano, “Skeleton Images Representation (SkeleMotion and TSRJI),” GitHub, Sep. 14, 2023. https://github.com/carloscaetano/skeleton-images

[31] “python_for_microscopists/260_image_anomaly_detection_using_autoencoders/260_image_anomaly_detection_using_autoencoders.py at master · bnsreenu/python_for_microscopists,” GitHub. https://github.com/bnsreenu/python_for_microscopists/blob/master/260_image_anomaly_detection_using_autoencoders/260_image_anomaly_detection_using_autoencoders.py

Appendix 1 – Folder Structure

```
Anomaly-Detection-in-Deadlift/
   requirements.txt
   README.md
----GenerateSkelemotion/
      GenerateSkeleton.py #Generate Skelemotion Images
--------skeleton_images/ #code from [7]
         GenerateSkeletonImages.py
         KinectData.py
         ImgType.py
----TSNE/
      TSNE.ipynb # EDA
----Evaluate/
      evaluate.ipynb #Final Evaluation
--------Anomaly/ #21 PNG images
--------Correct/ #29 PNG images
----Predict/
      encoder500epoch.keras #trained encoder model
      predict.py # prediction pipeline implementation
      autoencoder500epoch.keras #trained autoencoder model
----Train/
      Autoencoder.ipynb #Training of autoencoder
----Deadlift Stages/
      annotation.py #annotation of deadlifts
      annotation_data.csv
      deadlift_bot_rf.pkl #random forest model to identify deadlift stages
      train.py #training code for random forest model
----Data/
--------Orientation/
------------anomaly/
----------------anomaly/ #45 PNG images
------------Data/
```

----------------valid/

--------------------deadlift/ #50 PNG images

----------------deadlift/ #249 PNG images

----------------train/

--------------------deadlift/ #199 PNG images

Appendix 2 – User guide

Instructions for Running Predict.py:

1. *Navigate to the predict/predict.py directory.*

   - Use your file explorer or terminal to go to the directory where predict.py is located.

2. *Open a terminal within this directory.*

   - You can open a terminal or command prompt in this directory by right-clicking and selecting "Open Terminal" (or "Open Command Prompt" on Windows).

3. *Activate the virtual environment where the necessary packages are installed.*

   - If you're using a virtual environment (e.g., conda or venv), activate it using the appropriate command.

4. *Execute the prediction script.*

   - Run the following command in the terminal:

```
python predict.py --video <path_to_deadlift_video>
```

   Replace `<path_to_deadlift_video>` with the actual file path to your Deadlift video.

5. *Wait for the result window to open.*

   - The prediction process may take some time depending on the video's length and complexity. Please be patient.

6. *Access your video results.*

   - Once the prediction is complete, the video results will be automatically saved in the same directory as your input video file.

These steps should help you run the `predict.py` script effectively.

Appendix 3 – Installation guide

Installation Guide for Required Packages in a New Virtual Environment:

Follow these steps to create a new virtual environment and install the packages listed in the `requirements.txt` file:

Step 1: Set up a New Virtual Environment (Optional but Recommended)

If you don't have `virtualenv` installed, you can install it globally using `pip`:

```
pip install virtualenv
```

Now, create a new virtual environment (you can name it as you like):

```
virtualenv myenv
```

Activate the virtual environment:

- On Windows:

  ```
  myenv\Scripts\activate
  ```

- On macOS and Linux:

  ```
  source myenv/bin/activate
  ```

Step 2: Clone the github repository and install the Required Packages

Clone from : https://github.com/erictom97/Anomaly-Detection-in-Deadlift

Navigate to the directory containing your `requirements.txt` file and use `pip` to install the packages:

```
pip install -r requirements.txt
```

This command will read the `requirements.txt` file and install all the packages listed in it.

Step 3: Verify Installation

To ensure that the packages have been successfully installed, you can check the installed packages within your virtual environment:

```
pip list
```

You should see a list of packages, including those specified in `requirements.txt`.

That's it! You've successfully set up a new virtual environment and installed the required packages from the `requirements.txt` file.