Project plan for degree project

DV1478: BACHELOR PROJECT IN COMPUTER SCIENCE

April 3, 2025

Title	3D Pose Estimation and Time-Series Classification for Distinguishing Normal and Fatigued States.
Classification	3D Pose Estimation, Fatigue Classification , Time-Series data.

Student 1	Name	Yaswanth Rahul Yarlagadda
	E-Mail	yayr24@student.bth.se
	Person nr	031222T112
	Program	Bachelor's in computer science.
Student 2	Name	Shaheer Dudekula
	E-Mail	shdd24@student.bth.se
	Person nr	040808T099
	Program	Bachelor's in computer science.
Supervisor	Name & title	Shahryar Eivazzadeh , University Lecturer.
	E-Mail	shahryar.eivazzadeh@bth.se
	Department	Computer Science

1 Introduction

Understanding human motion plays a crucial role in practical applications such as healthcare [10], sports performance analysis, rehabilitation, ergonomics, and occupational safety. Detailed analysis of human movements enables early detection of fatigue, prevention of injuries, enhancement of physical performance, and improvements in ergonomic design for better productivity and comfort.

Recent advancements in computer vision technology have significantly transformed human motion analysis by eliminating the need for intrusive sensors. Nowadays, accurate and detailed human movements can be captured effortlessly through ordinary video recordings and analyzed using sophisticated computational methodologies. In this thesis, we present a practical, accessible, and cost-effective approach using multiple smartphones strategically positioned at various angles in an indoor environment. This setup provides extensive coverage and robust data acquisition, facilitating precise capture of diverse motion patterns.

Our methodology begins with the precise extraction of 2D joint positions from smartphone-recorded videos using the MMPose framework, a state-of-the-art open-source toolkit known for high-accuracy human pose estimation—unlike prior works that require motion-capture hardware. MMPose [7] effectively detects human body joints across varying lighting conditions, camera angles, and complex backgrounds, making it highly suitable for our intended indoor applications. Following this, the extracted 2D joint positions are converted into accurate 3D representations through a pose-lifting process, utilizing the SemGCN PoseLifter. This advanced Graph Convolutional Network (GCN) specifically addresses challenges in depth estimation and occlusions, resulting in reliable 3D joint positions even from monocular viewpoints. The outcome of this pipeline is an enriched time-series dataset comprising comprehensive 3D joint positions, velocities, and accelerations, ready for advanced analysis[5], [6].

Subsequently, we perform a detailed classification of the movement patterns into two main categories: normal and fatigued. For robust classification performance, we employ Long Short-Term Memory (LSTM) neural networks, specifically chosen for their superior capability to analyze temporal sequences and capture long-range dependencies inherent in human motion data. Beyond basic classification, our approach further aims to quantitatively predict fatigue levels [11], offering valuable insights into the progression and severity of fatigue during physical activities.

Ultimately, this research aims to develop a scalable, highly accurate, and user-friendly pipeline for human motion recognition and analysis. The framework's flexibility allows it to be easily extended or adapted for various types of physical activities beyond walking. By removing intrusive sensors and relying solely on accessible technologies like smartphones and advanced computational models, our methodology significantly enhances the potential for widespread adoption in health monitoring, sports analytics, rehabilitation programs, ergonomic assessments, and occupational safety management. Thus, it empowers users, trainers, clinicians, and ergonomists to better understand, evaluate, and optimize physical activities and conditions in diverse practical settings.

2 Ethical, Societal, and Sustainability Aspects

2.1 Ethical Aspects

Our project prioritizes ethical responsibility by ensuring informed consent and voluntary participation. Participants may withdraw at any time without consequences. Privacy is rigorously protected, and data collected is exclusively utilized for research purposes and securely managed [1].

2.2 Societal Aspects

Transparency is fundamental, guaranteeing that participants fully comprehend our methods and data usage. This open communication fosters trust and encourages active community participation and engagement in research activities.

2.3 Sustainable Aspects

We emphasize environmental responsibility by employing efficient computational tools like MMPose, which minimize resource usage and energy consumption, thereby promoting sustainability and resource efficiency.

3 Aim:

Aim: To perform accurate 2D pose estimation and lift it to 3D pose estimation from video data recorded using a three-camera setup, generating comprehensive 3D time-series datasets. These datasets will be classified into normal and fatigued categories, quantitatively predicting fatigue levels. Additionally, the study aims to visualize and analyze similarities and differences between normal and fatigued movements using a robust computational framework integrating MMPose [7], and Long Short-Term Memory (LSTM) neural networks [8].

3.1 Objectives:

- To record high-quality indoor video data of human motions using a three-camera smartphone setup and extract accurate 2D joint positions using the MMPose framework [7].
- To convert extracted 2D joint positions into precise 3D pose representations using the SemGCN PoseLifter and compute essential kinematic features such as joint velocities, accelerations, and joint angles using NumPy and SciPy.
- To design and implement an LSTM-based neural network capable of classifying normal versus fatigued movements and quantitatively predicting fatigue levels.

4 Research Questions

1. How can 3D human pose be reliably extracted from motions such as walking and running in low-demand settings using smartphone-recorded videos?
2. How accurately can LSTM neural networks predict fatigue levels based on the kinematic features of human motion ?

Motivation: Human fatigue leads to subtle but measurable changes in movement such as shorter stride lengths, altered posture, and less stable motion dynamics. While these variations may go unnoticed by the human eye, they can serve as valuable indicators of physical strain and performance degradation. Accurately identifying such patterns has important applications across domains like sports training, fitness tracking, workplace ergonomics, and rehabilitation.

However, traditional fatigue assessment methods are often either subjective, require expensive motion capture systems, or involve wearable sensors that may be uncomfortable or intrusive. In contrast, advancements in computer vision now enable 3D human pose estimation using low-cost, consumer-grade equipment such as smartphones, making movement analysis more scalable and accessible in real-world environments.

This project explores reliable methods for extracting 3D pose data from common motions—such as walking and running using smartphone-recorded videos in low-demand settings. Building on this, we investigate how machine learning models, specifically Long Short-Term Memory (LSTM) networks can be trained on kinematic features derived from these poses to predict an individual's fatigue level. This approach enables non-intrusive, automated fatigue monitoring that can be deployed in everyday scenarios, paving the way for smarter training systems, real-time injury prevention tools, and health-oriented motion analysis solutions.

5 Methodology

This thesis follows a structured, experimental approach to develop and evaluate a deep learning-based system for classifying human motion into **normal** and **fatigued** states using smartphone-recorded videos. The methodology includes collecting and preprocessing video data, performing 2D and 3D pose estimation, extracting kinematic features, and training an LSTM neural network for classification and fatigue level prediction.

Research Question 1

How can 3D human pose be reliably extracted from motions such as walking and running in low-demand settings using smartphone-recorded videos?

The first step is to collect custom motion data instead of relying on existing datasets. The data collection process involves recording videos indoors using three smartphone cameras positioned at different angles. These recordings capture full-body human movements in both normal and fatigued states, enabling a comprehensive dataset suitable for training and evaluation.

Pose estimation is performed using MMPose, an open-source pose estimation toolkit known for its robustness across various camera angles and lighting conditions [7]. From the video frames, 2D joint keypoints are extracted using MMPose. These keypoints are then passed through the SemGCN model, a graph convolutional network (GCN) designed for lifting 2D poses into accurate 3D representations, effectively handling occlusions and depth ambiguities [2]. This portion of the overall pipeline ensures that 3D pose data can be reliably extracted using low-cost, consumer-grade equipment such as smartphones.

Research Question 2

How accurately can LSTM neural networks predict fatigue levels based on the kinematic features of human motion?

The extracted 3D pose data is converted into time-series sequences, including features such as joint velocities, accelerations, and joint angles, computed using Python libraries like NumPy and SciPy [6]. Each motion sequence is labeled as either normal or fatigued, with additional annotations.

These labeled sequences are then input into an LSTM (Long Short-Term Memory) neural network [8], which is specifically chosen for its ability to capture long-range temporal dependencies in sequential data. The model is trained using 4-fold cross-validation to ensure generalization and robustness. After completing the 4-fold cross-validation, the model's classification and regression outputs are stored in a structured CSV file for detailed evaluation. Each row of the file corresponds to one test sample and includes actual and predicted fatigue levels [4]. This format allows clear comparison between ground truth labels and the model's predictions, enabling further statistical analysis and visualization of results. These outputs are later used to compute overall classification accuracy and fatigue prediction metrics such as Mean Absolute Error (MAE).

The design of the neural network is guided by past research demonstrating the effectiveness of LSTMs for activity recognition and fatigue analysis [8, 11]. Insights from recent pose estimation studies also inform the feature extraction and preprocessing steps to improve classification performance [5, 6].

Research Methods and Measured Quantities

Research Question 1: *How can 3D human pose be reliably extracted from motions such as walking and running in low-demand settings using smartphone-recorded videos?*

This research question adopts an Experimental Research focusing on pose estimation, where videos recorded using a three-camera smartphone setup are processed using MMPose for 2D keypoint extraction, followed by SemGCN for lifting to 3D. The reliability of the extracted 3D poses is assessed using standard metrics in the field, such as Mean Per Joint Position Error (MPJPE), along with qualitative evaluations under varying indoor conditions. These measurements help establish the accuracy and robustness of the proposed pose estimation pipeline.

Research Question 2:*How accurately can LSTM neural networks predict fatigue levels based on the kinematic features of human motion?*

This question is approached through an Quantitative Experimental Research (classification approach) using supervised machine learning, where kinematic features derived from the 3D pose sequences—such as joint velocities, accelerations, and angles—are used as input to a Long Short-Term Memory (LSTM) neural network.
The model's performance is evaluated using two primary metrics:
Classification accuracy for distinguishing between normal and fatigued states, and Mean Absolute Error (MAE) for predicting fatigue levels. These metrics enable an objective assessment of the model's ability to capture subtle variations in human motion associated with fatigue.

Additional Methodological Considerations

- Ethical Aspects: Participants are informed about the purpose of the study, and written informed consent is obtained prior to data collection. Privacy is ensured through anonymized storage of data and the option to withdraw at any point [1].
- Reliability and Validity: To ensure reliability, the system is tested under various indoor lighting conditions and camera configurations. For validity, results are cross-checked against relevant benchmarks and prior works on 3D human pose estimation and fatigue analysis [3].
- Data Storage and Structure: All 2D and 3D pose data is organized in structured JSON files, maintaining clarity and accessibility for downstream processing and future research.

This methodological framework supports the thesis objectives and ensures the system developed is accurate, ethical, and practically applicable in real-world scenarios like sports training, rehabilitation, and ergonomics.

6 Expected Outcomes

The key outcomes of this study include:

- Accurate Movement Classification: Developing a reliable system using MMPose and LSTM to distinguish between normal and fatigued movements effectively.
- Structured Dataset: Creating an organized, accessible dataset containing 3D joint positions, velocities, and accelerations for future research [3].
- Robust Performance Evaluation: Assessing classification accuracy through metrics such as overall accuracy, fatigue-level MAE.
- Extendable Framework: Designing a flexible and user-friendly pipeline adaptable for analyzing various types of human motion in future studies.
- The classification and fatigue prediction results will be stored in a structured CSV file for each sample, enabling detailed evaluation and comparison between actual and predicted outcomes .

This research aims to deliver a practical and reliable approach to analyzing human motion, beneficial in healthcare, sports, and rehabilitation contexts.

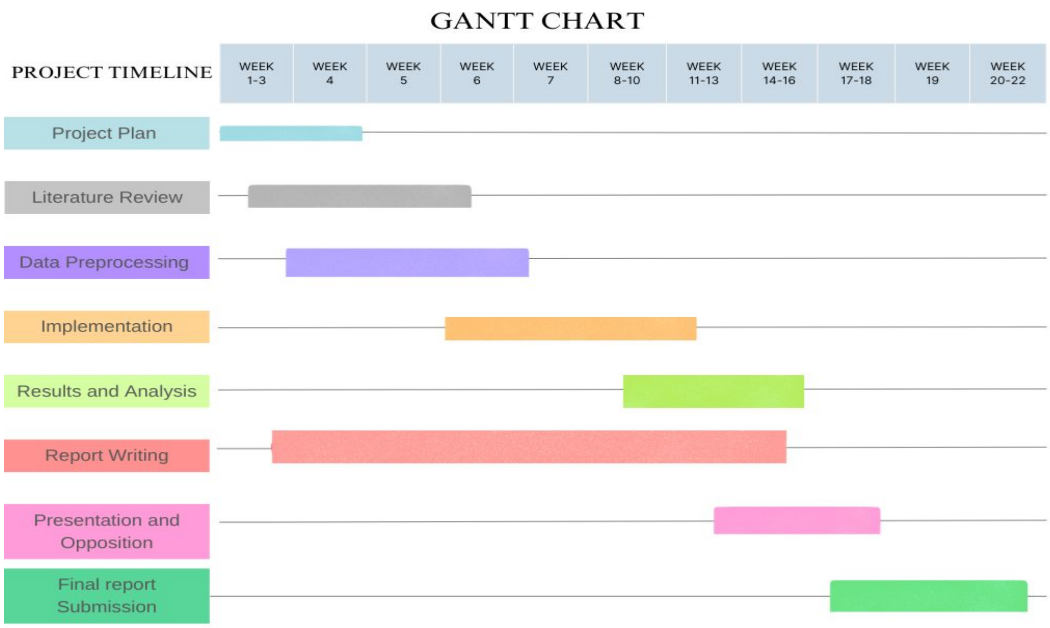

Figure 1: Time and activity plan

7 Time and activity plan

7.1 Supervision plan

We will update our supervisor every week to share our progress. If we have any quick questions, we will send an email for clarification. For tasks that need detailed discussion or guidance, we will set up an online or in-person meeting as needed. This regular communication will help us stay on

8 Limitations and Risk Management

8.1 Limitations

Despite aiming for robust performance, our approach faces several constraints:

1. Reliance on Pre-Recorded Videos:
   - The system is not designed for real-time capture and depends on previously recorded data.
2. Single-Person Focus:
   - Classifications are optimized for one subject at a time, making multiple-person [3] or overlapping movements difficult to handle.
3. Environmental Variations:
   - Fluctuations in lighting and camera placement can introduce noise or synchronization issues.
4. Generalization Challenges:

- The model may struggle with completely novel or unanticipated movement patterns, requiring additional training data and fine-tuning [9].

5. Computational Demands:
   - Processing 3D pose sequences can be resource-intensive, especially on standard hardware.

8.2 Risk Management

To address these limitations and reduce potential risks, the following strategies are employed:

1. High-Quality Data Collection:
   - Capture clear videos with accurate labeling to minimize errors during pose estimation.
2. Robust Testing of Environmental Conditions:
   - Evaluate the system under various lighting setups and camera placements to ensure consistent accuracy [3].
3. Preventing Overfitting:
   - Incorporate data augmentation and diverse movement scenarios for better generalization [9].
4. Computational Optimization:
   - Streamline algorithms to maintain efficiency on standard hardware.
5. Ethical Compliance:
   - Obtain informed consent, secure participant data, and respect privacy and withdrawal rights.

Risk	**Severity**	**Likelihood**	**Mitigation Strategy**
Inaccurate pose extraction due to poor lighting	High	Medium	Test under varied lighting; use image preprocessing.
Motion misclassification	High	Medium	Refine features, expand data, and fine-tune LSTM.
Limited generalization	Medium	High	Use diverse datasets and data augmentation.
High real-time load	High	Low	Optimize algorithms, use efficient models.
Participant privacy	High	Low	Informed consent, data anonymization, ethical guidelines.
Varying camera angles	Medium	Medium	Standardize placement; use multi-angle data.

Table 1: Risk Plan for the Motion Classification Framework

References

[1] Anonymous, “Human pose estimation: Ethical and privacy considerations.” [Online]. Available: https://insights.raconteur.net/human-pose-estimation

[2] bkcamp, “bkcamp/semgcn,” Apr. 2023, original-date: 2023-04-18T12:30:09Z. [Online]. Available: https://github.com/bkcamp/semgcn

[3] P. W. Code, “3d human pose estimation | papers with code.” [Online]. Available: https://paperswithcode.com/task/3d-human-pose-estimation

[4] H. Kim, S. Lee, D. Lee, S. Choi, J. Ju, and H. Myung, “Real-time human pose estimation and gesture recognition from depth images using superpixels and svm classifier,” *Sensors*, vol. 15, 2015. [Online]. Available: https://doi.org/10.3390/s150612410

[5] F. Koleini, M. U. Saleem, P. Wang, H. Xue, A. Helmy, and A. Fenwick, “BioPose: Biomechanically-accurate 3D Pose Estimation from Monocular Videos,” Jan. 2025, arXiv:2501.07800 [cs]. [Online]. Available: http://arxiv.org/abs/2501.07800

[6] W. Liu, Q. Bao, Y. Sun, and T. Mei, “Recent advances in monocular 2d and 3d human pose estimation: A deep learning perspective,” 2021, arXiv:2104.11536 [cs]. [Online]. Available: http://arxiv.org/abs/2104.11536

[7] OpenMMLab, “Mmpose: Openmmlab pose estimation toolbox and benchmark.” [Online]. Available: https://github.com/open-mmlab/mmpose

[8] F. J. Ordóñez and D. Roggen, “Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition,” *Sensors*, vol. 16, no. 1, p. 115, 2016.

[9] X. Wang, Y. Mi, and X. Zhang, “3d human pose data augmentation using generative adversarial networks for robotic-assisted movement,” *Frontiers in Neurorobotics*, 2024. [Online]. Available: https://doi.org/10.3389/fnbot.2024.00001

[10] M. Zago, C. Sforza, I. Pacifici, V. Cimolin, M. Galli, and C. Condoluci, “Applications of pose estimation in human health and performance evaluation: A scoping review,” *Frontiers in Bioengineering and Biotechnology*, vol. 9, 2021. [Online]. Available: https://doi.org/10.3389/fbioe.2021.722745

[11] Z. Zhang, Y. Luo, and L. Lin, “Driver fatigue detection method based on human pose information entropy,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, 2022. [Online]. Available: https://doi.org/10.1109/TITS.2021.3053301