International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 10 (2026) 124–138

Paper Type: Original Article

An Experimental Study on Pose Estimation Models and Data Preparation Strategies for Human Posture Assessment Pipeline

Basant Adel [1,*], Ahmad Salah [2], Mahmoud A. Mahdi [1], and Heba Mohsen [3]

[1] Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, Zagazig, Egypt; Basant.anany020@fci.zu.edu.eg; mamahdi@zu.edu.eg.

[2] Department of Information Technology, College of Computing and Information Sciences, University of Technology and Applied Sciences, Ibri, Oman; ahmad.salah@utas.edu.om.

[3] Department of Computer Science, Faculty of Computer and Information Technology, Future University in Egypt, New Cairo, Egypt; hmohsen@fue.edu.eg.

Received: 01 Nov 2025 **Revised:** 18 Jan 2026 **Accepted:** 17 Feb 2026 **Published:** 18 Feb 2026

Abstract

Human posture assessments play a crucial role in improving quality of life and performance, and even in preventing injuries; however, real-world applications still face many challenges. Multiple people per image, varying camera angles, and varied person orientations are recognized as existing obstacles in academia. This paper gives a structured exploration of four development trials completed while developing a comprehensive posture- analysis pipeline. Firstly, a representation of a comparative analysis of five popular human datasets: 3.6M, MPII, COCO, CMU Panoptic, and Human Eva. Secondly, four different human pose estimation models—Open Pose, MediaPipe, MMPose RTMO-I, and YOLO- NAS-POSE—were evaluated against predefined criteria after testing on the same sample of images. Thirdly, data preprocessing techniques—class imbalance handling, feature engineering, and splitting order—were applied on a custom-built and labeled dataset for the view classification module. Finally, a Unity-generated avatar dataset with normal, conditional normal, and abnormal human posture representations. The results of those four trials guided the design choices for the final pipeline. RTMO-I was selected as the preferred multi-person HPE model, where accuracy was chosen over latency. None of the five datasets used as image datasets; instead, a custom dataset was built from the COCO dataset annotation files and labeled with custom rules. Class imbalance was managed internally by the classification model; feature engineering was left out for simplicity, as that distance-related features does not improve the performance of our model, and formal splitting of the dataset before preprocessing steps was preserved. Finally, low-contrast, standing-still side-view avatars were generated to test elbow and knee posture for verification. These findings emphasize practical considerations in designing real-world posture-analysis systems. Additionally, they indicate the benefit of iterative methodological trials in refining both model selection and dataset design.

Keywords: Human Pose Estimation, Motion Analysis, Data Preprocessing, View Classification, Abnormal Human Posture Dataset.

1 | Introduction

How a person stands, moves, or performs specific tasks can be the difference between good performance and getting injured. That is not limited to athletic performance and sport analysis but workplace ergonomics [1],

Corresponding Author: Basant.anany020@fci.zu.edu.eg

walking down the streets, and daily household activities as well [2]. Therefore, reliable tools that can interpret, understand, and analyze human posture have become increasingly important. A promising direction is the technology of estimating the human posture known as Human Pose Estimation (HPE), which can deliver fast and objective feedback of the human pose. The significantly increasing demand made human pose estimation an essential and widely adopted technique for analyzing body posture and movement [3]. Human pose estimation refers to the process of identifying key body joints from input data that can be 2D or 3D images, recorded videos, or live streams.

Machine learning, deep learning, and other cutting-edge techniques are used to detect different body parts or joints, where each model produces its own set of body landmarks. Popular datasets of human postures are also now annotated and publicly available for training yet more HPE models. Joints-like hips, shoulders, elbows, and knees are detected by those models and represented by their two-dimensional space location in a skeleton-like representation. Models differ in their ability to detect single and multiple people per image, 2D or 3D images or videos, and their ability to detect fine details like torso landmarks. Because HPE is a rich domain in academia, it offers a flexible foundation for more advanced forms of movement analysis [4, 5]. HPE models are allowed to track how the joints are moving relative to each other via a central concept called human keypoints. This is the representation of the coordinates of anatomical landmarks on the human body. Angles can be computed, movement patterns can be measured, and potential deviations from the standard normal positions can be detected once the keypoints are detected. These detailed insights from the HPE models are compatible with the fact that human motion is multidimensional, interactive, and highly nonlinear. Despite the strength of those modern HPE models and posture assessment techniques, several challenges still exist.

A key remaining challenge is reliable multi-person detection, as in many real-world scenarios, such as industrial safety settings, a single instructor cannot monitor every worker's posture at the same time. HPE systems that can analyze multiple people in parallel within one frame, therefore, offer a practical solution. Additionally, different camera positions and a person's orientation led to different information being obtained and, as a result, different judgments and actions accordingly. Therefore, the selection of the HPE model and the consideration of such variations must be addressed by any practical posture assessment systems.

Our broader research [6], aimed to address gaps in the field, which ultimately resulted in a complete posture analysis pipeline. The pipeline integrates multi-person human pose estimation, a novel view classification module, and full-body angle-based posture assessment. The full system is presented in detail in a separate publication. In this article, we highlight four key developmental trials that shaped the creation of that final pipeline; those are

1. Human Posture Dataset Selection: The selection process of one of the publicly available human posture datasets.
2. HPE Model Selection and Evaluation: The selection and evaluation of state-of-the-art HPE models that fulfill our focus.
3. Data Preprocessing for View Classification: Investigation of multiple data preprocessing techniques and how they affect the performance of the view classification model.
4. Danger Dataset Creation Process: The creation process of a synthetic abnormal/extreme poses dataset for multiple scenarios pipeline testing.

Despite significant advances in human pose estimation, deploying reliable posture assessment systems in real-world scenarios remains challenging due to strong dependencies on view orientation, data preparation strategies, and the interaction between pose estimation and downstream classification components. In this work, we investigate these challenges through an experimental study that examines how design choices across the pipeline influence the reliability of a single-camera, view-aware posture assessment system. Rather than proposing a single new algorithmic contribution, this study aims to systematically analyze and justify the practical design decisions that shape the performance and robustness of such pipelines.

Accordingly, this study is guided by the following research question: How do different design choices across the posture assessment pipeline—particularly dataset selection, data preparation strategies, view labeling, and model selection—affect the reliability and robustness of a single-camera, view-aware posture assessment system?

This paper is organized as follows: Section 2 reviews related studies on human pose estimation advancements, data preprocessing techniques with human posture assessment, and synthetic dataset generation for posture assessment. Section 3 presents the methodology adopted in a series of experiments conducted during the development of the proposed pipeline, each followed by its respective results. Finally, Section 4 concludes the paper with key findings, overall contributions, and future work potential areas.

2 | Related Work

This work presents a series of experimental trials conducted during the development process of a human posture assessment pipeline with novel contributions like the view classification model and the abnormal joint angles avatar Unity dataset. Therefore, this section presents recent studies and approaches related to the main components of the sub-domains that are covered next in Section 3. The related literature is organized to correspond with these stages as follows:

1. Human Pose Estimation Advancement
2. Data Preprocessing Techniques with Human Posture Assessment
3. Synthetic Datasets Generation for Posture Assessment

2.1 | Human Pose Estimation Advancement

In the Methodology and Results, section 3.2 presents the process of selecting the most suitable HPE model as one of the four steps in the posture assessment pipeline. A practical analysis was conducted by testing various scenarios, including single- and multi-person images, different views, and camera orientations. Accordingly, related studies that present comparative analyses of HPE models and techniques are discussed here.

Two-dimensional human posture estimation models have gained a lot of attention and are now being discussed extensively in academia. Therefore, authors in [7] provided a starting point for newcomers and directed researchers to find new models by examining the methodology and architectural shortcomings of previous studies. This work has examined and compared fourteen human pose estimation models with various comparison factors. Models' backbone architecture, like ResNet, Mask CNN, and HRNet-W32, and the number of people per image, single or multi-person, were summarized for all models, alongside other factors like the dataset used, loss functions, bottom-up or top-down, and evaluation metrics.

The survey study [8] discusses the remaining issues with deep learning-based human pose estimation methods, including occlusion, depth ambiguities, and a lack of training data. An evaluation of current deep learning-based methods for 2D and 3D posture estimation from over 200 research papers since 2014 was conducted. The purpose of this study was a methodical examination and comparison of these methods according to their inference methods and input data. A summary and discussion of the studied approaches' quantitative performance comparisons on widely used datasets are provided. The difficulties, usages, and potential avenues for future study are discussed by the authors as well.

2.2 | Data Preprocessing Techniques with Human Posture Assessment

In the area of wearable sensor-based human posture detection, the authors in [9] conducted a comparative study of classification algorithms and data preprocessing methods. Written accelerometers were used to test the NN, SVM, and Bayes algorithms over different-sized sets and subsets of testing. Additionally, their experimental results, various feature selection meth- odds, scaling strategies, and plurality voting mechanisms were proposed to enhance the overall system's robustness and dependability. Similarly, the effects of various combinations of frequently used data preparation techniques on the ability to classify gait patterns were examined in [10].

Authors addressed the issue that is there are very few recommended practices or detailed comparisons of different preprocessing methodologies and how they affect classification performance. Three-dimension ground reaction forces (GRFs) were recorded using a public available dataset of one day trails of forty-two participants. Combinations of preprocessing steps: GRF filtering, time derivative, time normalization, data reduction, weight normalization and data scaling were analyzed. Different classification models were used to conduct this analysis. The findings show that while some combinations had little impact on the performance, filtering GRF data and supervised data reduction (such as Principal Components Analysis) boosted the machine-learning classifiers' prediction performance.

2.3 | Synthetic Datasets Generation for Posture Assessment

The rapid development of posture assessment techniques highlighted the importance of datasets that capture wider ranges of human motion. Scenarios like non-standard, motion-related, and extreme postures. Recent studies have explored synthetic dataset generation techniques. For instance, authors in [11] address the lack of advancement in 3D human posture estimation applications connected to monocular sports. They identified the absence of a dataset that records quick, high-acceleration athletic motions as the gap in the current human pose databases. They proposed AthletePose3D, a large-scale dataset intended to capture complicated, fast-moving biomechanical movements that tend to be missed by traditional datasets. Using 1.3 million frames and 165k postures from 12 sports-related motions in the dataset, state-of-the-art 2D and 3D pose estimation models were assessed; the results demonstrate an increase in accuracy.

Similarly, HmPEAR, a multimodal dataset designed to advance 3D human pose estimation and human action recognition in unconstrained outdoor settings, was introduced in [12]. Over 300,000 synchronized frames of RGB images, LiDAR point clouds, and motion capture-derived 3D human posture annotations encompassing 25 subjects in ten scenarios are available in the dataset, with It also includes 40 types of everyday human behaviors, yielding over 6,000 annotated action clips. The authors worked to fill the gap in the literature by focusing in the out- door environments and cross-task learning. Good annotation accuracy was demonstrated, while highlighting the dataset's challenges for existing approaches, the results also show significant improvements in performance when combining modalities, indicating that pose estimation and action recognition tasks are mutually reinforcing.

2.4 | Research Gap and Study Relevance

The reviewed studies provide strong foundations for this work; for instance, comparative HPE surveys offer comprehensive evaluations of architectures and datasets, guiding model choice in our trials. Preprocessing research demonstrates how filtering, normalization, imbalance handling, and feature selection affect posture-related classification. Synthetic datasets such as AthletePose3D and HmPEAR contribute large-scale, high-quality motion data across sports or outdoor settings.

However, abnormal or unsafe joint configurations are not primarily represented by these datasets; they instead focus on natural or athletic movements. Also, the accuracy requirements of multi-person, multi-view posture assessment scenarios are also not a focus of previous HPE evaluations. This article, on the other hand, details real-world experiments carried out during pipeline development. Experimentations like choosing a model that prioritizes accuracy over latency like RTMO-I overYOLO-NAS-POSE. In addition, building the Unity-based dataset that records extreme joint angles, which despite its small size, fills a niche and lays the foundation for further advancements.

3 | Methodology and Results

In this section, the experimental workflow and outcomes of several trials are represented. During the development of the human posture assessment pipeline, those trials have been conducted where some affected the pipeline construction directly, and some have not contributed at all. For each stage, the applied methods are described first, followed by a summary of the obtained results and observations

3.1 | Dataset Primary Selection

At the early stage of this work, the objective was to identify a publicly available human image dataset of reasonable size that could serve as a foundation for running a human pose estimation (HPE) model. The plan was to extract 2D keypoints from the dataset and subsequently process them to assess postural conditions in terms of potential ergonomic risks. Several well- known datasets were considered and reviewed in terms of their scale, person-level annotations, and suitability for 2D pose-based posture analysis. The existence of keypoints annotations, the diversity of the postures, and the multi-person scenes were the elements of the criteria of the selection, which are shown in Table 1

Table 1. Comparison of publicly available HPE datasets considered during the early dataset selection stage.

Dataset	Size	Single/Multi Person	Keypoints	Suitability
3.6M [13]	3.6 million frames	Multi-person	18	Medium
MPII [14]	25K images	Single-person annotations	16	High
COCO [15]	200K+ images	Multi-person	17	High
CMU Panoptic [16]	100K+ images	Single-person	Limited	Low
HumanEva [17]	11 subjects (motion sequences)	Single-person	15	Low

Results:

After a deeper investigation, it became clear that none of the existing datasets directly provided the view orientation (front, back, side) required for the proposed pipeline. Since this orientation was critical for mapping postures to specific Range of Motion (ROM) guidelines, it was necessary to extend the available data. Consequently, a custom dataset was constructed using the COCO dataset annotations as a foundation, with additional manually defined view labels generated through a set of custom rule-based labeling criteria. The construction and replication of the custom COCO-derived dataset follow the same pipeline and annotation logic described in our previously published study [6], where the dataset generation process is fully detailed. COCO was selected as the base dataset not only due to its scale, but primarily because it provides keypoint visibility flag annotations per joint, which enabled the construction of view labels required for training the view classification model.

This newly created dataset offered the required alignment between human pose data and the intended posture assessment framework. While public datasets such as MPII and COCO were sufficient for early experimentation with pose estimation models, the custom-labeled dataset ultimately became the core resource for training and evaluating the final view classification model in later stages of this work

3.2 | Model Selection

The goal in this stage was to accurately identify full-body keypoints by a human pose estimation (HPE) model. Also, detecting various human poses from different viewing angles. We selected and tested four different models based on their reported accuracy, supported keypoints, and ability to handle multi-person scenarios. These include:

- **OpenPose**[18] Estimation (based on pre-trained single-person model)
- **MediaPipe**[19] Pose (Google lightweight single-person estimator)
- **YOLO-NAS-Pose**[20] (multi-person real-time detector)
- **MMPose RMOT-I**[21] (multi-person, high-accuracy model)

Each model was evaluated on a diverse collection of images that collectively represented different scenarios including:

- Frontal clear-person images
- Side-view person images
- Back-view person images
- Multi-person scenes

While not all images were identical across models, each set covered similar types of cases to allow a fair assessment of model performance across varied viewpoints and scene compositions.

The main evaluation criteria focused on:

i. Detection accuracy.
ii. Model latency.
iii. Ability to estimate multiple person postures per image.

Default weights and configurations were used on the same hardware setup to ensure a consistent comparison for all the models. For visualization and further analysis, sample results from each model were recorded and compared in the following section.

Table 2. Comparison of Human Pose Estimation Models Evaluated in Trial 1

Model	Single / Multi	Accuracy	Latency	Selected
OpenPose	Single-person	Low-Medium	High	No
MediaPipe Pose	Single-person	Low-Medium	Low	No
YOLO-NAS-Pose	Multi-person	Medium-High	Low	No
MMPose RTMO-I	Multi-person	High	Medium-Low	Yes

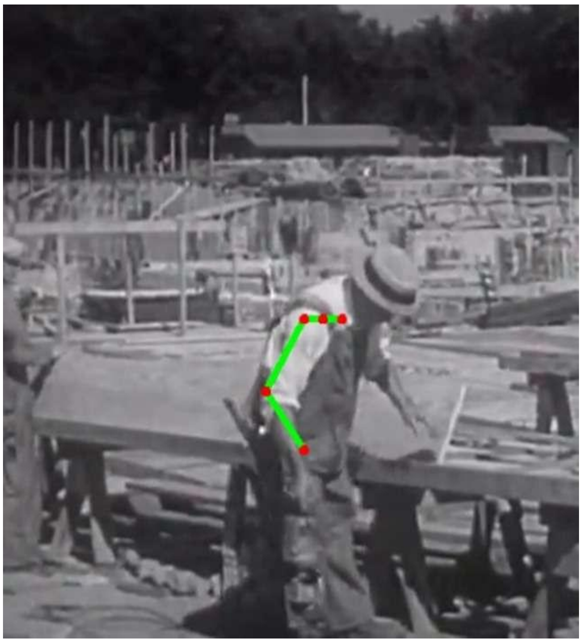
Results:

Table 2 represents the performance of the four selected models in terms of single/multi-person detection ability, accuracy, and latency. Visual confirmation of the results reported in the table is provided in two scenarios:

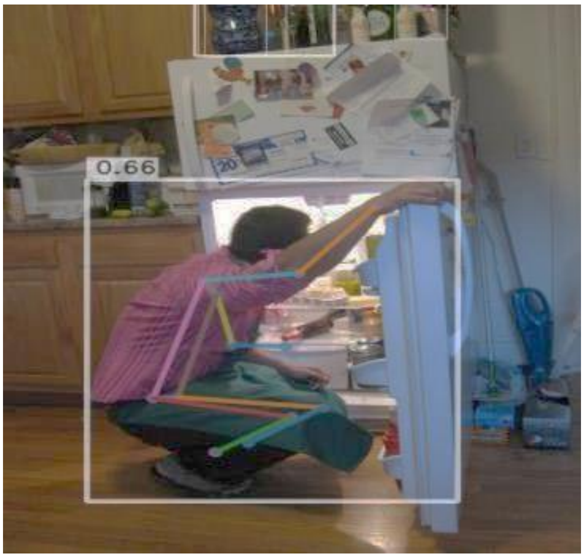
1. Single-Person Side View Scenes.
2. Multi-Person Different Views Scenes.

To aid interpretation, Figures 1 and 2 show representative qualitative examples, whereas quantitative evaluation was conducted separately using standardized image sets. As shown in Figure 1, MediaPipe and OpenPose generate only partial or less reliable key- point configurations, while YOLO-NAS-Pose and MMPose RTMO-I produce more complete skeleton overlays, with MMPose offering the most stable and anatomically consistent detec- tions. All models are further assessed in multi-person scenarios in Figure 2. In these exam- ples, OpenPose creates merged or entangled keypoints across various individuals, resulting in distorted skeletons, while MediaPipe detects only one person and ignores the others. YOLO- NAS-Pose and MMPose RTMO-I, on the other hand, both successfully detect multiple people; however, YOLO-NAS-Pose also generates false-positive bounding boxes and lower-confidence estimations (e.g., confidence score of 0.66), while MMPose provides more precise keypoint lo- calization and cleaner person separation. When taken as a whole, these visual results show why MMPose was chosen for the pipeline and validate the quantitative trends shown in the table.

a. MediaPipe Pose – Side View



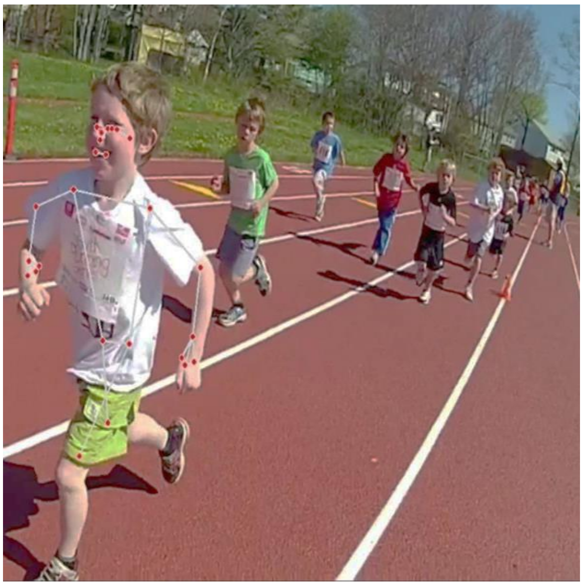
b. OpenPose – Side View



c. YOLO-NAS-Pose – Side View

d. MMPose RTMO-I – Side View

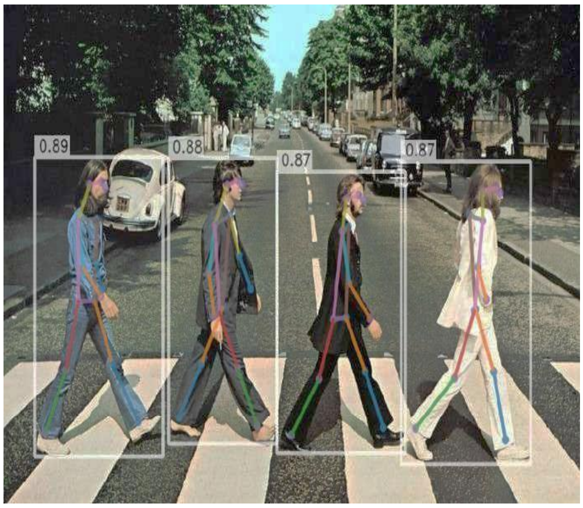
Figure 1. Side-view comparison of skeleton predictions generated by four HPE models (all samples from the MPII dataset).



a. MediaPipe Multi-Person Skeleton Overlay

b. OpenPose Multi-Person Skeleton Overlay

c. MMPose RTMO-I Multi-Person Skeleton Overlay



d. YOLO-NAS-Pose Multi-Person Skeleton and Bounding Boxes Overlay

Figure 2. Evaluated multi-person scenes using four human pose estimation models.

It should be noted that the latency and accuracy comparisons reported in this study are primarily qualitative and intended to support model selection within the proposed pipeline rather than to serve as a formal benchmarking analysis. Comprehensive numerical latency benchmarks and statistical confidence intervals are therefore considered beyond the scope of this work.

3.3 | Data Preprocessing for View Classification

Following the extraction of 2D Keypoints using the selected HPE model, the next step involved preparing the data for training the custom view classification models. The preprocessing phase aimed to enhance data quality, address class imbalance, and evaluate whether the data splitting strategy or feature engineering could influence the model's performance.

Three major factors were examined:

- Splitting strategy: Two strategies were compared—Split before Preprocessing and Split after Preprocessing. In the first case, the dataset was divided into training and testing subsets prior to handling missing values and class imbalance, ensuring a completely un- seen test distribution. In the second, the entire dataset was preprocessed first, then split, ensuring more consistent scaling and balancing.

- Balancing technique: Since the dataset exhibited noticeable class imbalance, with the "unknown" and "front" views dominating the distribution, two imbalance handling approaches were tested. The Advanced Balancing strategy combined oversampling of minority classes and under-sampling of the majority using techniques such as SMOTE and RandomUnderSampler. In contrast, the Model Balanced configuration relied on internal class weighting provided by the classifiers (e.g., class weight='balanced' in Random Forest [22, 23] and XGBoost [24, 25].

- Feature engineering: Additional geometric features were introduced based on distances between anatomically relevant keypoints. For example, the Euclidean distance between the two shoulder joints was hypothesized to help distinguish between front and side views, as the apparent distance decreases when the person is rotated sideways.

To systematically assess the impact of these design choices, four configurations were evaluated as summarized in Table 3.

Table 3. Configurations used during data preprocessing for the view classifier

Configuration	Split Strategy	Balancing Method	Feature Engineering (Distance)
Config 1	Split After	Advanced Balancing	used
Config 2	Split After	Model Balanced	Not used
Config 3	Split After	Advanced Balancing	Not used
Config 4	Split Before	Advanced Balancing	Not used

Results:

Figure 3 illustrates the class-wise F1-scores obtained across the four preprocessing configurations, while Figure 4 summarizes the test set metrics for the same configurations. The F1 score was used as the primary evaluation metric to capture the balance between precision and recall for each class, given that the dataset was inherently imbalanced. All view classification models were optimized using RandomizedSearchCV (Randomized Search Cross Validation) with 5-fold cross-validation. The folds reported in Figures 3 and 4 correspond to the internal cross-validation folds used during hyperparameter tuning, and all reported results reflect the best-performing hyperparameter configurations, not default model settings.

To further account for imbalance effects, the Matthews Correlation Coefficient (MCC) was also computed. MCC provides a more reliable overall measure of classification performance under unequal class distributions, as it considers true and false positives and negatives across all classes. This makes it especially valuable for assessing the consistency and generalization of the model's predictions beyond accuracy or F1 alone.

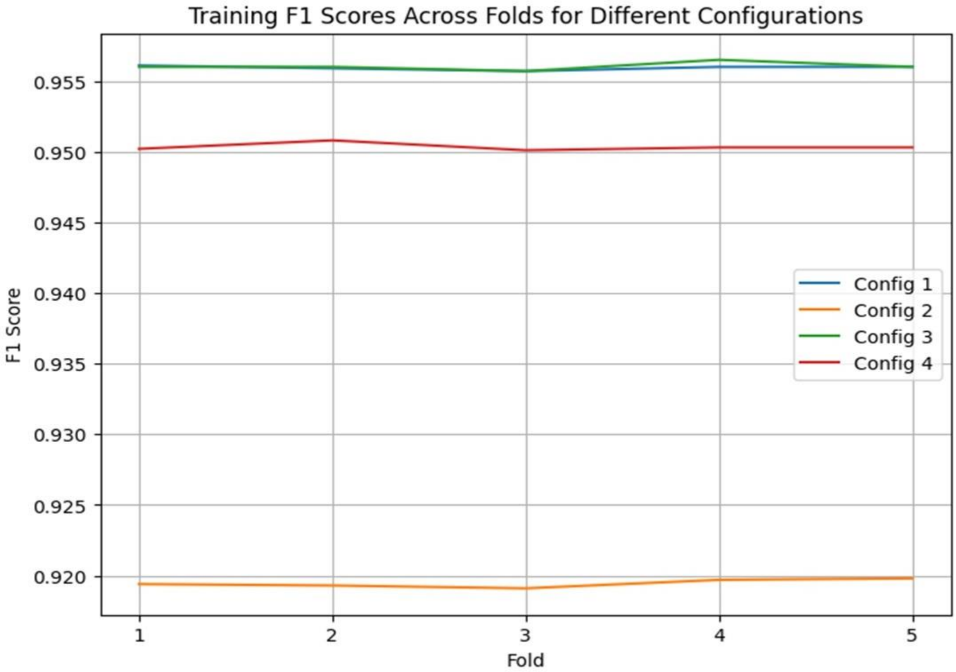


Figure 3. Class-wise F1 scores for the four preprocessing configurations.

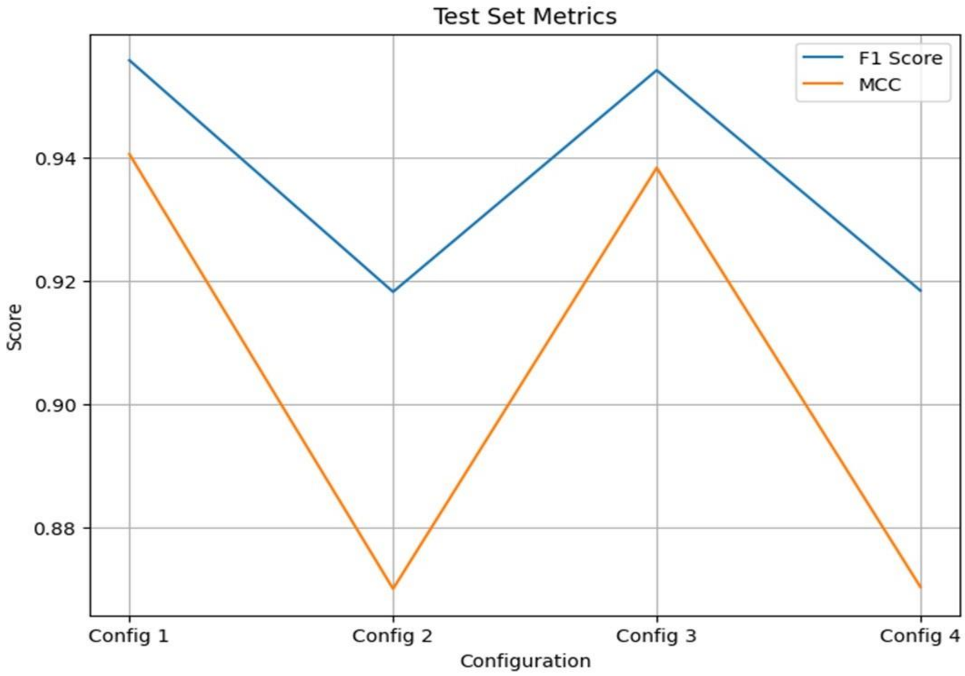


Figure 4. Overall F1 and MCC test metrics for the four preprocessing configurations.

From the results, all configurations that adopted the Split after Preprocessing strategy achieved higher F1 and MCC values, indicating more stable feature scaling and better handling of minority classes. However, despite the improved numerical performance, this approach is methodologically sound, as preprocessing prior to data splitting can lead to information leakage from the test set into the training process. Therefore, the Split before preprocessing configuration was considered the scientifically appropriate setup for subsequent experiments, even if it yielded slightly lower scores.

From the results, configurations adopting the Split after Preprocessing strategy achieved higher F1 and MCC values, suggesting more stable feature scaling and improved handling of minority classes. However, despite the improved numerical performance, this setup is methodologically unsound, as performing preprocessing before data splitting introduces information leakage from the test set into the training process. Consequently, these results are considered misleading. For this reason, the Split before Preprocessing configuration was selected for all subsequent experiments, as it represents the scientifically appropriate and unbiased evaluation protocol, even though it resulted in slightly lower performance scores. This configuration better reflects real-world deployment conditions, where preprocessing parameters are learned exclusively from the training data and applied to previously unseen samples, thereby providing a more reliable estimate of model generalization.

Additionally, the introduced distance-based features delivered by the feature engineering phase which were evaluated across all tested view classification models and preprocessing configurations showed no consistent performance improvement compared to raw Keypoints coordinates alone. This suggests that the selected classifiers were able to implicitly capture joint-distance relationships. Accordingly, feature engineering was not further explored in this work, as the baseline performance was already satisfactory for the intended task.

3.4 | Testing Pipeline and Synthetic Dataset Preparation

To evaluate the robustness of the proposed pipeline, including both pose estimation and view classification components, a controlled testing environment was developed. Because naturally occurring images depicting abnormal or extreme joint angles are difficult to obtain online, a custom synthetic dataset was generated using the Unity engine (version 6000.0.36f1).

An avatar from the Unity gallery was selected and positioned in the side view. The left knee joint was chosen for simulation due to its clear rotational axis and relevance to ergonomic evaluation. The custom ApplyInjury() function was used to force extension beyond the normal physiological range (from −90° to −120° on the X-axis) while keeping Y/Z rotations fixed at 0°. This procedure yielded 28 synthetic images with left knee angles varying from 0° (neutral) to approximately 54°, representing normal, conditionally normal, and abnormal postures.

For general testing, an additional 23 real images of edge-cases were collected from online resources, covering all three primary views (front, back, and side) and multiple joint movements including flexion, extension, adduction, abduction, and lateral bending. Some images were horizontally flipped to test the classifier's ability to differentiate mirror-symmetric poses. In addition to the synthetic and challenging edge-case images, the models were evaluated on a larger set of over 100 images representing typical postures, ensuring broader coverage and robustness.

Visual Contrast Evaluation. Two contrast settings were tested to examine whether the classifier's performance depended on visual contrast between the subject and background. In the first setup, the avatar was rendered with light clothing and a light background (low contrast), and in the second, with dark clothing and a light background (high contrast). Both configurations were correctly classified, indicating robustness to contrast variations (see Figure 5).

Pose Variation Evaluation. Further experiments included atypical or uncommon body configurations to test model generalization. As shown in Figure 7, cases with unusual arm positions (e.g., arms crossing or raised asymmetrically) led to occasional misclassifications or detection failures, particularly when limbs occluded the torso. However, the model successfully detected and classified views for challenging scenarios such as sleeping postures or front-facing surveillance-style images.

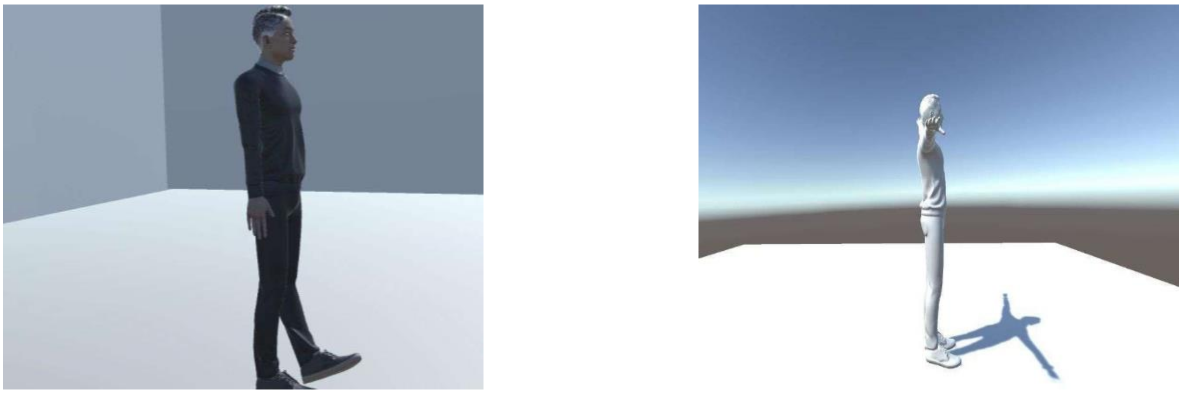
Table **4** summarizes the classification outcomes across all tested scenarios, indicating where the model succeeded or failed to identify the view correctly.

Results:

The evaluation of the proposed pipeline was conducted using both synthetic and real images, as described in the previous section. Each image was assessed in terms of its inclusion in the dataset, whether the view classification component correctly identified the pose, and whether the human pose estimation (HPE) model successfully detected the keypoints. Table 4 summarizes the outcomes for the representative figures included in this study. These results provide an overview of the pipeline's performance under varying contrast conditions and pose variations

Table 4. Summary of classification and pose estimation results for selected images.

Figure	Used in Dataset	View Classification Correct	HPE Successful
Figure 5(a) High Contrast Setup	No	Yes	Partially
Figure 5(b) Low Contrast Setup	Yes	Yes	Partially
Figure 7(a) sleeping pose	No	Yes	Yes
Figure 7(b) Surveillance-like captured position	No	Yes	No
Figure 7(a) Back View dangerous knee pose	No	Yes	No
Figure 7(b) front view dangerous elbow pose	No	Yes	No



a. High Contrast Setup

b. Low Contrast Setup

Figure 5. Contrast variation experiments for the synthetic Unity avatar. Both configurations were correctly classified as side view.

a. side sleeping position

b. Surveillance-like captured position

Figure 6. Pose variation testing using synthetic and real images. The model failed in some occluded-arm cases but succeeded in classifying side and front views for complex postures

a. Back View dangerous knee pose

b. front view dangerous elbow pose

Figure 7. Pose variation testing using synthetic and real images. The model failed in some occluded-arm cases but succeeded in classifying side and front views for complex postures.

4 | Conclusion and Future Work

This paper presented a structured series of four development trials carried out during the creation of a human posture analysis pipeline through systematic experimentation. First, five popular human pose datasets underwent a comparative study, and the most suitable one, the COCO dataset, along with its annotation files, was selected to generate our custom-built and labeled dataset for the novel view classification module. Subsequently, the ability of various Human Pose Estimation (HPE) models to detect full-body keypoints across varied viewpoints in multi-person scenarios was evaluated, and the most suitable model was identified based on accuracy and consistent keypoint extraction.

Following, dataset preparation trials were addressed for the purpose of solving challenges like class imbalance, order of train-test splitting, and feature engineering on the performance. Finally, different scenarios were tested during the development of a Unity avatar danger posture dataset, including color contrast, and viewpoint for testing the overall pipeline performance with normal, abnormal, and conditional normal scenarios. Overall, the insights gained from these trials highlight practical considerations in building reliable pose-based posture-analysis systems and provide potential work areas to fill in the gap in the academia.

Despite the gained insights that helped in enhancing our published pipeline, several aspects still require further investigation. A potential area for future studies is finding an HPE model that is suitable for real-time multi-person applications, where both higher accuracy and lower latency are required. Additionally, training or fine-tuning HPE models on abnormal or non- standard postures can highly improve HPE applications in workplace ergonomics and sport accident scenarios. Further work is also needed in the area of data preprocessing; researchers may want to consider designing a standard preprocessing pipeline tailored for pose-based applications, where more stable and accurate results can be achieved. In addition, future work will focus on developing more sophisticated view labeling strategies to better capture intermediate and ambiguous orientations (e.g., three-quarter and in-between views), and on reducing potential bias introduced by rule-based labeling.

Although in our original posture assessment pipeline, the view classification achieved promising overall accuracy, additional improvements may be obtained by continuing to improve dataset entries, labeling techniques, and data preprocessing. Moreover, the danger posture dataset could be further enriched by incorporating greater diversity in lighting conditions, body appearances, genders, skin tones, camera orientations, and multi-person interactions, and further extend all the scenarios on joint-level analysis. A collective of full-body joints diverse scenarios can be assembled in one general-purpose human pose estimation dataset, and a joint- specific ones can also be a point of contribution. Collaboration with domain experts would further enhance the validity, usability, and impact of these datasets, opening new avenues for both synthetic data generation and biomechanically-informed human pose research.

Author Contribution

Conceptualization, B.A. and A.S.; Methodology, B.A. and A.S.; Software, B.A.; Validation, A.S. and M.M.A.; Formal Analysis, B.A.; Investigation, B.A. and A.S.; Resources, A.S.; Data Curation, B.A. and M.M.A.; Writing—Original Draft Preparation, B.A.; Writing—Review and Editing, A.S., M.M.A., and H.M.; Visualization, B.A., A.S. and M.M.A.; Supervision, A.S., M.M.A., and H.M.; Project Administration, A.S. All authors have read and agreed to the published version of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The dataset, annotation files, and code used in this study will be made available upon request to the corresponding author for research purposes.

Conflicts of Interest

The author declares that there are no conflicts of interest related to the content or publication of this research.

References

[1] U. Deshpande, S. Araujo, S. Deshpande, et al., “A review of machine learning techniques for ergonomic risk assessment based on human pose estimation,” Discover Artificial Intelligence, vol. 5, Oct. 2025. DOI: 10.1007/s44163-025-00566-5.

[2] J. Stenum, K. M. Cherry-Allen, C. O. Pyles, R. D. Reetzke, M. F. Vignos, and R. T. Roemmich, “Applications of pose estimation in human health and performance across the lifespan,” Sensors, vol. 21, no. 21, 2021, ISSN: 1424-8220. DOI: 10.3390/s21217315. [Online]. Available: https://www.mdpi.com/1424-8220/21/21/7315.

[3] U. Özbalkan and Ö. Turna, “Advancements in human pose estimation: A review of key studies and findings till 2024,” Academic Platform Journal of Engineering and Smart Systems, vol. 13, pp. 94–107, Sep. 2025. DOI: 10.21541/apjess.1588025.

[4] Z. Gao, J. Chen, Y. Liu, Y. Jin, and D. Tian, “A systematic survey on human pose estimation: Upstream and downstream tasks, approaches, lightweight models, and prospects,” Artificial Intelligence Review, vol. 58, p. 68, 2025. DOI: 10.1007/s10462-024-11060-2

[5] X. Yan, B. Liu, and G. Qu, “Efficient monocular human pose estimation based on deep learning methods: A survey,” IEEE Access, vol. 12, pp. 72 650–72 661, 2024. DOI: 10. 1109/ACCESS.2024.3399222

[6] Adel, B.; Salah, A.; Mahdi, M.A.; Mohsen, H. View-Aware Pose Analysis: A Robust Pipeline for Multi-Person Joint Injury Prediction from Single Camera. AI **2026**, 7, 7. https://doi.org/10.3390/ai7010007

[7] T. L. Munea, Y. Z. Jembre, H. T. Weldegebriel, L. Chen, C. Huang, and C. Yang, “The progress of human pose estimation: A survey and taxonomy of models applied in 2nd human pose estimation,” IEEE Access, vol. 8, pp. 133 330–133 348, 2020. DOI: 10. 1109/ACCESS.2020.3010248

[8] C. Zheng, W. Wu, C. Chen, et al., “Deep learning-based human pose estimation: A survey,” ACM Comput. Surv., vol. 56, no. 1, Aug. 2023, ISSN: 0360-0300. DOI: 10. 1145/3603618. [Online]. Available: https://doi.org/10.1145/3603618

[9] S. Zhang, D. Monekosso, and P. Remagnino, “Data pre-processing and model selection strategies for human posture recognition,” in 2018 11th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP), 2018, pp. 1–DOI: 10.1109/CSNDSP.2018.8471858

[10] J. Burdack, F. Horst, S. Giesselbach, I. Hassan, S. Daffner, and W. I. Schöllhorn, “Systematic comparison of the influence of different data preprocessing methods on the performance of gait classifications using machine learning,” Frontiers in Bioengineering and Biotechnology, vol. 8, p. 507 468, 2020. DOI: 10.3389/fbioe.2020.00260. [Online]. Available: https://doi.org/10.3389/fbioe.2020.00260

[11] C. Yeung, T. Suzuki, R. Tanaka, Z. Yin, and K. Fujii, Athletepose3d: A benchmark dataset for 3d human pose estimation and kinematic validation in athletic movements, 2025. arXiv: 2503.07499 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2503.07499

[12] Yitai Lin, Zhijie Wei, Wanfa Zhang, Xiping Lin, Yudi Dai, Chenglu Wen, Siqi Shen, Lan Xu, and Cheng Wang. 2024. HmPEAR: A Dataset for Human Pose Estimation and Action Recognition. In Proceedings of the 32nd ACM International

Conference on Multimedia (MM '24). Association for Computing Machinery, New York, NY, USA, 2069–2078. https://doi.org/10.1145/3664647.3681055

[13] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 7, pp. 1325–1339, Jul. 2014

[14] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2014

[15] T.-Y. Lin, M. Maire, S. Belongie, et al., Microsoft coco: Common objects in context, 2015. arXiv: 1405.0312[cs.CV]. [Online]. Available: https://arxiv.org/ abs/1405.0312

[16] H. Joo, H. Liu, L. Tan, et al., "Panoptic studio: A massively multiview system for social motion capture," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3334–3342

[17] L. Sigal, A. O. Balan, and M. J. Black, "Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," Int.J. Comput. Vision, vol. 87, no. 1–2, pp. 4–27, Mar. 2010, ISSN: 0920-5691. DOI: 10.1007/s11263-009-0273-6. [Online]. Available: https://doi.org/10. 1007/s11263-009-0273-6

[18] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, Openpose: Realtime multi- person 2d pose estimation using part affinity fields, 2019. arXiv: 1812.08008[cs.CV]. [Online]. Available: https://arxiv.org/abs/1812.08008

[19] C. Lugaresi, J. Tang, H. Nash, et al., Mediapipe: A framework for building perception pipelines, 2019. arXiv: 1906.08172[cs.DC]. [Online]. Available: https:// arxiv.org/abs/1906.08172

[20] S. Aharon, O. Louis-Dupont, O. Masad, et al., Super-gradients, 2023. DOI: 10.5281/ZENODO . 7789328. [Online]. Available: https://zenodo.org/records/7789328

[21] T. Jiang, P. Lu, L. Zhang, et al., Rtmpose: Real-time multi-person pose estimation based on mmpose, 2023. arXiv: 2303.07399 [cs.CV]. [Online]. Available: https:// arxiv.org/abs/2303.07399

[22] Scikit-learn developers, Random forest classifier documentation, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (visited on 03/25/2025).

[23] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[24] XGBoost developers, Xgboost documentation, 2025. [Online]. Available: https://xgboost.readthedocs.io/en/latest/ (visited on 03/25/2025).

[25] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '16, San Francisco, California, USA: ACM, 2016, pp. 785–794, ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939785.