International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Deep Self-Supervised Representation Learning for Multi-Gait Detection in Real-World Sensor Recordings

Sayyed Hasanoddin[1], Rabiya Basri[2], Podeti Deepthi[2], Pedam Jeevani[2], Shaik Shoaib Aslam[2],
[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data Science)
[1,2]Vaagdevi College of Engineering (UGC-Autonomous), Bollikunta, Warangal, 506005, Telangana.

ABSTRACT

With global wearable adoption exceeding 1.1 billion devices, there is a critical need for scalable methods to analyze long-term accelerometer and IMU data. Conventional activity recognition often fails due to manual labeling inconsistencies and poor generalization across diverse subjects. To address these gaps, this study introduces a self-supervised learning framework for robust multi-gait detection. By employing pretext tasks to learn gait representations directly from unlabelled multi-sensor sequences including acceleration, angular velocity, and orientation, the model improves generalization and computational efficiency without relying on extensive human annotation. For classification, a Greedy Tree ensemble is trained on these learned representations to identify activities such as walking, running, climbing, and sitting. The framework was benchmarked against Adaptive Boosting (AdaBoost) classifier, K-Nearest Neighbors (KNN) classifier, Logistic Regression Classifier (LRC), and Naive Bayes Classifier (NBC), demonstrating superior performance in accuracy and F1-score, particularly in cross-participant generalization. Additionally, the system is integrated into a Flask-based web application, supporting real-time data streaming and interactive monitoring. These results indicate that the proposed self-supervised approach bridges the gap between raw sensor data and actionable insights, facilitating early identification of mobility abnormalities and personalized intervention. This scalable solution provides a robust foundation for clinical, sports, and rehabilitation applications, enabling continuous real-world human activity monitoring and long-term observation.

Key Words: Self-Supervised Learning, Machine Learning, Time-Series Analysis, Real-Time Activity Monitoring

1. INTRODUCTION

How we move contains significant clues about ourselves. Our gait (manner of walking) has been closely studied in medicine, psychology, and sports science. Recently, gait analysis has received increased attention from the computer science community coinciding with the exponential progress of deep learning and widespread availability of computing hardware. AI-powered gait analysis systems have been able to successfully recognize subjects, estimate demographics such as gender and age, and estimate external attributes such as clothing, without using any external appearance cues. These results are not surprising, given the large number of individual differences in gait, which are due to differences in musculoskeletal structure, genetic and environmental factors, as well as the walker's emotional state and personality. Current systems are only really trained and tested in controlled indoor environments.

Most methods use the CASIA-B dataset as the standard benchmark for gait recognition models, containing 124 subjects walking indoors in a strictly controlled manner captured with multiple cameras. Complexity in the real-world cannot be fully modelled by such restrained scenarios. Only recently the focus has been on modelling gait "in the wild", with datasets such as Dense Gait, GREW, and Gait3D. Gathering a large-scale dataset that is clean and fully annotated represents a tremendous effort in terms of both financial resources and allocated time. The GREW dataset reportedly took 3 months of continuous work to be gathered and annotated. While such approaches have been useful in developing neural architectures for processing gait, they are not sufficiently diverse to be properly used in more relaxed, real-world environments.

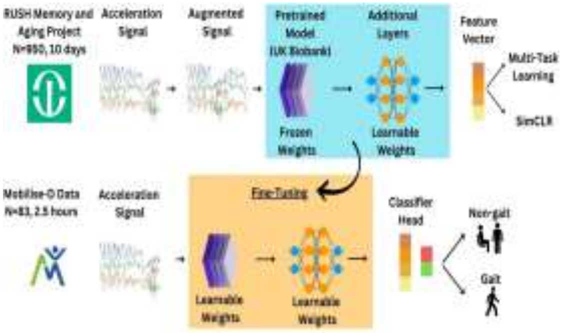

Fig. 1: Deep learning model for gait analysis using pre-training and fine-tuning.

The AI community has been slowly moving away from this approach in other areas as well, with methods for self-supervised learning for both vision and language gaining significant traction and often surpassing traditional supervised methods. Recent progress in self-supervised learning showed that self-supervised models are more robust and exhibit emerging behaviours, not explicitly defined during training. For instance, DINO, a vision transformer trained in a self-supervised regime, learned class-specific features enabling unsupervised object segmentation without using any such labels during training. The first contrastive method for self-supervised learning for gait analysis, by training a ST-GCN on a smaller version of Dense Gait. Their method obtained reasonable results on downstream gait recognition tasks and showed that there is a strong correlation between the pretrained dataset size and zero-shot transfer performance. While many approaches for gait analysis have been utilizing silhouettes extracted from background subtraction, extracting silhouettes in real surveillance scenarios implies the use of more advanced techniques, such as instance segmentation, which come at a significant computational cost. Sequences of silhouettes occupy significant storage space and are not sufficiently flexible to be used in other adjacent tasks, such as activity recognition. Moreover, silhouettes encode subtle appearance cues, which makes it unclear to what extent movement is utilized in the identification. On the other hand, 2D pose estimation models have become increasingly accurate and computationally efficient. Skeletons are cheap to extract, currently more reliable than 3D meshes and 3D poses, especially at a distance. Moreover, 2D skeletons are significantly more lightweight than silhouettes in terms of long-term storage.

Current architectures for processing sequences of skeletons are utilizing the natural spatial graph structure present in the human skeleton, introducing an inductive bias in the model design. Models such as the popular ST-GCN and MS-G3D have seen impressive results for skeleton-based action recognition. Concurrently, there has been an explosion in the use of transformer models in almost all areas of deep learning since their initial application for natural language processing. Transformers are considered a more general architecture, with few inductive biases. Initially transformers have struggled to match CNN models for image classification but are currently surpassing other models and are showing promising results in self-supervised scenarios and, more so than other types of architectures, transformers have shown impressive learning capacity and emergent behaviours under self-supervision. The main contributions of this article are as follows:

- To develop machine learning-based models such as LRC, Adaboost classifier, KNN, NBC and Greedy Tree Classifier (GTC) that accurately detects multiple gait activities from long, continuous wearable sensor recordings.
- Analyze gait dynamics, handle noise and variability, and extract essential motion patterns from unsegmented, real-world data.
- Identify optimal model architectures, feature extraction methods, and

training techniques that work efficiently with high-dimensional time-series data and minimal labeled datasets.

2. LITERATURE SURVEY

Fang et al. [1] achieved 89.59% static gesture recognition accuracy using gloves integrated with inertial and magnetic measurement units (IMMUs). Ramalingame et al.'s [**2**] forearm pressure sensors attaining 93% gesture discrimination precision. Trunk-mounted sensors (e.g., hip accelerometers) provide stable whole-body gait analysis, as demonstrated by Mantyjarvi's [**3**] 83–90% gait recognition rates. Lower-limb sensors directly capture kinetic features, as evidenced by Zhao et al. [4] achieving 98.11% gait identification accuracy using foot-worn inertial sensors. Notably, existing systems predominantly focus on isolated body segments, lacking effective spatiotemporal fusion of heterogeneous sensor data. The authors of [5] proposed methods for identifying human activities based on a decision tree classifier. However, the classification accuracy rate is considered unsatisfactory. Cheng et al. [6] proposed three distinct classification methods, such as hidden Markov model, support vector machine, and artificial neural network, to categorize body activities. While these methods deliver acceptable performance, they are either constrained in handling significant intraclass variations or hindered by the complexity of adjusting model parameters. Furthermore, the integration of contextual information and multi-modal sensor fusion techniques has enhanced the system's ability to distinguish between similar activities and detect transitions between different movement states, making HAR systems increasingly reliable for real-world applications.

Zhu et al. [7] have introduced a load-free hand rehabilitation system based on virtual reality (VR) made from ionic hydrogels. The system can identify 14 hand gestures with an accuracy of 97.9%. Another activity recognition system is developed by Lu et al. [8]. As they have produced a 5G Narrowband Internet of Things (NB-IoT) system, it is developed for human healthcare data collection, transmission, and reproduction together. The system is integrated with a bionic crack-spring fiber sensor (CSFS) inspired by Cirrus and Spider Structures. This system is characterized by its high sensitivity and long sensing range. Another study is presented by Mengarelli et al. [**9**], this study investigates the feasibility of estimating the vertical component of the ground reaction force (VGRF) using only EMG signals from the thigh and shank muscles. Two deep learning models were used across three experimental setups. The findings demonstrate that EMG signals can be effectively leveraged to estimate VGRF during walking. Tigrini et al. [10] has proposed a new phasor-based feature extraction approach (PHASOR) that captures spatial myoelectric features to improve the performance of LDA and SVM in gait phase recognition. A publicly available dataset was used to evaluate PHASOR. Additionally, data-driven deep learning architectures, such as Rocket and Mini-Rocket, were included for comparison.

Moreover, myoelectric activity of muscles was used to estimate ankle kinematics as proposed by Mobarak et al. [11]. sEMG signals were recorded for a total of 288 gait cycles. Two feature sets were extracted from sEMG signals in the time domain (TD) and wavelet (WT) and compared. Then, they were used for feeding three machine learning models (artificial neural networks, random forest, and least squares support vector machine (LS-SVM)). For the usage of EMGs, the authors in research [12] proposed a data acquisition system for measuring EMG signals for human lower limb activity recognition. Five leg activities have been accomplished to measure EMG signals from two lower limb muscles to validate the developed hardware. Five subjects were chosen to acquire EMG signals during these activities. Moore et al. [13] presented a dataset with 15 healthy subjects. They were four females and eleven males with an average age of 24 ± 4 years, height of 1.75 ± 0.09 m, and mass of 74 ± 13 kg. The recorded activities are walking at three different speeds (0.8 m s−1, 1.2 m s−1 and 1.6 m s−1). A total of approximately 1.5 h of normal walking and 6 h of perturbed walking are included in these datasets.

Another open-source dataset is presented by Hu et al. [14], in which data were selected from 10 healthy subjects. There were seven males and three females. Their biometrics were 25.5 ± 2 years; 174 ± 12 cm; and 70 ± 14 kg for age, height, and weight, respectively. A larger dataset was presented by Lencioni et al. [15], where data were collected from 50 healthy subjects. There were 25 males and 25 females. Their age range, mass and height were 6–72 years, 18.2–110 kg, and 116.6–187.5 cm, respectively. An eight-channel wireless sEMG (ZeroWirePlus) was used. Their signals were band-pass filtered at 10–400 Hz and sampled at 800 Hz, 960 Hz and 1000 Hz.

3. PROPOSED METHODOLOGY

The proposed system uses a GTC trained on self-supervised pretext tasks to extract intrinsic gait representations from wearable sensor data. The workflow begins with data acquisition from IMUs, followed by preprocessing. Self-supervised learning enables the model to learn temporal gait patterns without requiring fully labelled datasets. The GTC predicts multi-class gait activities, which are displayed in a Flask-based web application, supporting real-time streaming, monitoring, and visualization of detected activities. This approach enables scalable and reliable activity recognition in diverse environments and participants. At the User Environment layer, individuals wear sensors such as accelerometers and gyroscopes, which continuously capture motion data. These wearable devices are responsible for detecting fine-grained movements in real time and transmitting the raw signals to nearby edge devices. The sensors are lightweight and unobtrusive, allowing for natural movement and long-duration monitoring without impacting the user’s daily activities.

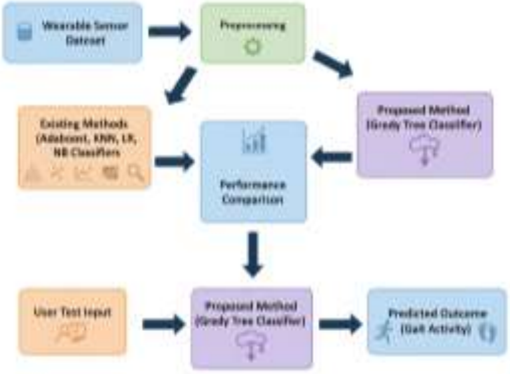

Fig. 2: Proposed system architecture.

The Edge Devices layer acts as the first point of data acquisition and temporary storage. Devices such as smartphones, smartwatches, or dedicated edge gateways collect the sensor data via Bluetooth or Wi-Fi, performing preliminary checks for data integrity and buffering before sending it to the backend. This layer ensures low-latency data transmission and reduces the risk of data loss during continuous recording. The Backend System forms the core of the architecture, where raw sensor data is processed, pre-processed, and transformed into a suitable format for self-supervised learning. Signal preprocessing includes noise reduction, filtering, normalization, and segmentation into fixed-length windows. The system then employs self-supervised representation learning to automatically extract meaningful features from these signals without heavy reliance on labelled data. These features are fed into machine learning classifiers, including NBC, KNN, LR, and GTC models, for accurate multi-gait activity classification.

4. Results Description

Fig. 3 shows that sign-up page serves as the initial entry point for users wishing to access the multi-gait detection system. It provides a simple interface where new users can create an account by entering a username, email, and password. The design includes a clear "Sign Up" button to submit the registration details, ensuring an intuitive user experience. Upon successful registration, users are prompted to log in to proceed further. The page also offers a link for users who already have an account to log in directly. This step is crucial for securing user data and personalizing the experience within the system. The layout is user-friendly, encouraging quick and easy account creation.

Fig. 3. Sign up page for multi gait detection.

Fig. 4 displays the login page appears after a successful sign-up, allowing users to access their accounts with their credentials. It features fields for entering a username and password, accompanied by a "Login" button to authenticate the user. A notification confirms the account creation, guiding users to log in immediately. Additionally, there is an option for users without an account to sign up, enhancing accessibility. This page is designed to ensure secure access to the system while providing a seamless transition from registration. It plays a vital role in maintaining user privacy and system integrity.

Fig. 4. Login page (after signing up) for multi gait detection.

Fig. 5 shows that dashboard page is the central hub after logging in, offering an overview of the multi-gait detection system. It welcomes users and provides a brief description of the system's purpose, focusing on self-supervised learning for gait detection using wearable sensor recordings. Users can navigate to sections like Exploratory Data Analysis, Activity Classification, and Make Predictions through interactive buttons. The page also includes a system overview with details on the target column, dataset, and machine learning models used. This interface is designed to be informative and functional, enabling users to explore and utilize the system's capabilities effectively.

Fig. 5. Dashboard Page for multi gait detection.

Fig. 6 displays the activity distribution chart visually represents the frequency of different activities recorded in the dataset, such as walking, running, and using an elevator. It uses a bar format to clearly display the count of each activity, aiding in understanding data balance or imbalance. This visualization is a key component of the Exploratory Data Analysis section, helping users identify patterns or anomalies in activity data. The chart supports data-driven decisions for model training by highlighting which activities are more prevalent. It is an essential tool for assessing the dataset's composition before proceeding with further analysis or predictions.

Fig. 6. Activity Distribution Chart for multi gait detection.

Fig. 7 shows the training results for the AdaBoost model display the performance metrics, including accuracy, precision, recall, and F1-score, for classifying various activities. The confusion matrix provides a detailed breakdown of true positives, false positives, and misclassifications across activities like walking and running. The classification report offers per-class precision, recall, and F1-scores, indicating the model's effectiveness for each activity. With an accuracy of 51.02%, this model provides a baseline for comparison with other algorithms. These results are crucial for evaluating the model's suitability for the multi-gait detection task.

Fig. 7. Training results – AdaBoost for multi gait detection.

Fig. 8 tells the greedy tree model training results showcase an impressive accuracy of 99.00%, along with matching precision, recall, and F1-scores. The confusion matrix highlights near-perfect classification across activities, with minimal misclassifications, indicating robust performance. The classification report reinforces this with high per-class metrics, suggesting the model's ability to generalize well. This high performance makes it a strong candidate for practical deployment in gait detection. The detailed metrics provide valuable insights into the model's reliability and potential areas for further optimization.

Fig. 8. Training results – greedy tree for multi gait detection.

Fig. 9 shows the KNN model results indicate an accuracy of 77.85%, with recall at 77.85% and precision at 78.75%, alongside an F1-score of 78.02%. The confusion matrix details the model's performance, showing the number of correct and incorrect predictions for each activity. The classification report breaks down precision, recall, and F1-scores per class, offering a comprehensive view of its strengths and weaknesses. This moderate performance suggests KNN as a viable option, though it may require parameter tuning for better results. These metrics are essential for comparing KNN with other models.

Fig. 9. Training results – KNN for multi gait detection.

Fig. 10 displays that logistic regression model achieves an accuracy of 51.10%, with precision and recall both at 51.10% and an F1-score of 50.21%. The confusion matrix provides a detailed account of predictions versus actual activities, revealing the model's classification performance across various gaits. The classification report includes per-class metrics, highlighting areas where the model struggles or excels. This baseline performance is useful for understanding the complexity of the gait detection task. The results suggest a need for feature engineering or alternative models for improved accuracy.

Fig. 10. Training results – logistic regression for multi gait detection.

Fig. 11 displays that naive Bayes model results show an accuracy of 57.40%, with recall at 57.40% and precision at 58.37%, alongside an F1-score of 54.58%. The confusion matrix outlines the model's prediction accuracy, indicating the distribution of correct and incorrect classifications. The classification report provides detailed per-class metrics, offering insights into the model's performance for individual activities. This moderate accuracy suggests naive Bayes as a reasonable starting point, though it may benefit from data preprocessing or feature selection. These results are valuable for comparative analysis with other models.

Fig. 11. Training results – naive bayes for multi gait detection.

Fig. 12. Prediction interface for multi gait detection.

Fig. 12 shows that prediction interface allows users to input sensor data for real-time activity prediction using selected machine learning models like KNN. It includes fields for entering values from accelerometer, gyroscope, and EMG sensors across various body parts, with options to use sample data. Users can choose a model and predict the activity, with the result displayed as the predicted activity, such as "going down." The interface features buttons to reset the form or load sample data, enhancing usability. This tool is designed to apply trained models practically, providing immediate feedback based on user input.

5. CONCLUSION

The proposed project successfully demonstrates an intelligent framework capable of accurately identifying diverse gait patterns using continuous IMU and EMG sensor data. By integrating self-supervised learning techniques with the Greedy Tree Classifier, the system effectively learns meaningful feature representations from unlabelled data, improving model robustness and generalization across varying motion patterns and subjects. Extensive experimentation revealed that the proposed approach significantly outperformed traditional supervised models such as AdaBoost, KNN, Logistic Regression, and Naive Bayes in terms of accuracy, precision, recall, and F1-score. The denoising and normalization preprocessing techniques, combined with statistical feature extraction, enhanced the model's ability to distinguish subtle gait variations with high reliability. With an overall accuracy of 99%, the framework proves highly effective for real-world, long-duration activity recognition scenarios. Thus, this research establishes a strong foundation for automated gait monitoring systems applicable in healthcare, rehabilitation, and sports analytics, ensuring continuous, accurate, and efficient human activity detection.

REFERENCES

[1] Fang, B.; Sun, F.; Liu, H.; Liu, C. 3D human gesture capturing and recognition by the IMMU-based data glove. Neurocomputing 2018, 277, 198–207.

[2] Ramalingame, R.; Barioul, R.; Li, X.; Sanseverino, G.; Krumm, D.; Odenwald, S.; Kanoun, O. Wearable Smart Band for American Sign Language Recognition With Polymer Carbon Nanocomposite-Based Pressure Sensors. IEEE Sens. Lett. 2021, 5, 6001204.

[3] Mantyjarvi, J.; Himberg, J.; Seppanen, T. Recognizing human motion with multiple acceleration sensors. In Proceedings of the 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236), Tucson, AZ, USA, 7–10 October 2001; Volume 2, pp. 747–752.

[4] Zhao, H.; Wang, Z.; Qiu, S.; Wang, J.; Xu, F.; Wang, Z.; Shen, Y. Adaptive gait detection based on foot-mounted inertial sensors and multi-sensor fusion. Inf. Fusion 2019, 52, 157–166.

[5] Nurwulan, N.R.; Selamaj, G. Human daily activities recognition using decision tree. J. Phys. Conf. Ser. 2021, 1833, 012039.

[6] Cheng, L.; Guan, Y.; Zhu, K.; Li, Y. Recognition of human activities using machine learning methods with wearable sensors. In Proceedings of the 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 9–11 January 2017; pp. 1–7.

[7] Zhu, P.; Niu, M.; Liang, S.; Yang, W.; Zhang, Y.; Chen, K.; Pan, Z.; Mao, Y. Non-hand-worn, load-free VR hand rehabilitation system assisted by deep learning based on ionic hydrogel. Nano Res. 2025, 18, 94907301.

[8] Lu, L.; Hu, G.; Liu, J.; Yang, B. 5G NB-IoT System Integrated with High-Performance Fiber Sensor Inspired by Cirrus and Spider Structures. Adv. Sci. 2024, 11, e2309894.

[9] Mengarelli, A.; Tigrini, A.; Scattolini, M.; Mobarak, R.; Burattini, L.; Fioretti, S.; Verdini, F. Myoelectric-Based Estimation of Vertical Ground

Reaction Force During Unconstrained Walking by a Stacked One-Dimensional Convolutional Long Short-Term Memory Model. Sensors 2024, 24, 7768.

[10] Tigrini, A.; Mobarak, R.; Mengarelli, A.; Khushaba, R.N.; Al-Timemy, A.H.; Verdini, F.; Gambi, E.; Fioretti, S.; Burattini, L. Phasor-Based Myoelectric Synergy Features: A Fast Hand-Crafted Feature Extraction Scheme for Boosting Performance in Gait Phase Recognition. Sensors 2024, 24, 5828.

[11] Mobarak, R.; Tigrini, A.; Verdini, F.; Al-Timemy, A.H.; Fioretti, S.; Burattini, L.; Mengarelli, A. A Minimal and Multi-Source Recording Setup for Ankle Joint Kinematics Estimation During Walking Using Only Proximal Information From Lower Limb. IEEE Trans. Neural Syst. Rehabil. Eng. 2024, 32, 812–821.

[12] Vijayvargiya, A.; Singh, P.; Kumar, R.; Dey, N. Hardware Implementation for Lower Limb Surface EMG Measurement and Analysis Using Explainable AI for Activity Recognition. IEEE Trans. Instrum. Meas. 2022, 71, 1–9.

[13] Moore, J.K.; Hnat, S.K.; van den Bogert, A.J. An elaborate data set on human gait and the effect of mechanical perturbations. PeerJ 2020, 3, e918.

[14] Hu, B.; Rouse, E.; Hargrove, L. Benchmark Datasets for Bilateral Lower-Limb Neuromechanical Signals from Wearable Sensors During Unassisted Locomotion in Able-Bodied Individuals. Front. Robot. AI 2018, 5, 14.

[15] Lencioni, T.; Carpinella, I.; Rabuffetti, M.; Marzegan, A.; Ferrarin, M. Human kinematic, kinetic and EMG data during different walking and stair ascending and descending tasks. Sci. Data 2019, 6.