Artificial Intelligence Applications And Innovations Vol. 2 No. 1 (2025)

Received 17 August 2025
Revised 19 February 2026
Accepted 16 May 2026
Published online 1 January 2025

Modeling Ankle Joint Dynamics Based on Inertial Sensor Data Using Machine and Deep Learning

A. Naeimi-Sadigh[1*], A. Shirazi[1]

[1] Department of Computer Science, Semnan University, Semnan, Iran

*** Corresponding author email address**: naeimi@semnan.ac.ir

Article Info

Article type:
Original Research

How to cite this article:
Naeimi-Sadigh, A., & Shirazi, A. (2025). Modeling Ankle Joint Dynamics Based on Inertial Sensor Data Using Machine and Deep Learning. *Artificial Intelligence Applications and Innovations, 2*(1), 65-73.
https://doi.org/10.61838/jaiai.2.1.5

ABSTRACT

In recent years, human motion analysis using inertial measurement units (IMUs) has emerged as a lightweight, low-cost, and portable alternative to advanced biomechanical equipment such as optical motion capture systems or force plates. IMUs enable the recording of motion data in real-world and daily settings, making them powerful tools for medical, sports, and rehabilitation applications. This study proposes a compact and intelligent framework for maodeling ankle joint dynamics using only a single IMU mounted on the right ankle, with the dual aim of classifying movement activities and estimating selected dynamic parameters. Data were obtained from the publicly available HuGaDB dataset, comprising accelerometer and gyroscope signals sampled at 100 Hz. The signals were segmented using sliding windows, and 36 statistical features were extracted from each segment. Significant differences between classes were examined using the Kruskal–Wallis test, and the most relevant features were selected with the SelectKBest algorithm. Three modeling approaches were evaluated: Gaussian Process Regression, Random Forest, and a hybrid Convolutional Neural Network with Long Short-Term Memory. The CNN+LSTM model achieved the highest classification accuracy at 96.41%, outperforming the other models, while Random Forest reached 90.5% accuracy using the top 10 selected features. GPR produced satisfactory results for continuous parameter estimation. Training time, resource consumption, and model size were also assessed. The findings demonstrate that high accuracy in human motion analysis can be achieved using data from a single IMU. The proposed framework holds promise for clinical, sports, and rehabilitation contexts, as well as for developing lightweight and personalized wearable systems.

Keywords: *Ankle joint dynamics modeling, IMU sensor, machine learning, deep learning, classification, human motion, CNN-LSTM network*

1. Introduction

Human motion analysis is a fundamental and widely applied topic across various fields such as sports science, physiotherapy, biomedical engineering, rehabilitation, prosthesis and exoskeleton design, human–machine interaction, and even digital health. The ability to accurately understand joint movements, especially in the lower limbs, plays a crucial role in diagnosing movement disorders, planning treatments, preventing injuries, optimizing athletic performance, and developing wearable technologies. Among the key joints, the ankle joint has been studied less extensively than the knee and hip, despite its vital role in maintaining balance, generating propulsion, and ensuring stability during walking or running.

Dysfunction in this joint can lead to serious injuries or a reduced quality of life.

In the past, precise motion analysis was only possible using advanced and expensive systems such as optical cameras, force plates, and electromyography (EMG). Although these tools provide high accuracy, they require a laboratory setting, controlled conditions, and specialized operators, limiting their portability and everyday use. To overcome these limitations, inertial measurement units (IMUs)—consisting of accelerometers and gyroscopes—have been introduced as inexpensive, lightweight, and field-friendly tools [1]. IMUs can measure linear acceleration, angular velocity, and, in some cases, magnetic fields, and can be easily mounted on various joints of the body.

With the rapid adoption of IMUs in motion studies, the main challenge has shifted toward intelligent analysis of raw sensor data. Such data are typically time series, often noisy, and sensitive to variations in movement conditions. Extracting meaningful information such as torque estimation, activity classification, or disorder detection requires advanced data analysis algorithms. Machine learning, and deep learning in particular, have become powerful tools for processing such complex time-series data [2].

Previous studies have shown that classical machine learning models such as Random Forest and Gaussian Process Regression (GPR), as well as deep learning models like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, can effectively classify movements and predict kinetic and kinematic variables from IMU data [3, 4]. For example, Wang et al. employed LSTM, CNN, and RNN models to estimate lower-limb joint torques with high accuracy [5], while Shi et al. combined Random Forest with automated feature extraction to predict knee and ankle motion parameters effectively [6]. Recurrent architectures such as NARX have also been used to forecast joint forces in nonlinear data [7].

Moreover, hybrid CNN+LSTM architectures, capable of simultaneously capturing spatial and temporal patterns, have shown great success in processing multidimensional IMU data [8]. Public datasets such as HuGaDB provide opportunities to evaluate models under more natural and diverse human activities [9]. In recent research, advanced architectures such as Transformers have been explored for modeling long-term dependencies in time sequences [10], while Graph Neural Networks (GNNs) have been applied to represent skeletal structures and analyze joint relationships [11]. The integration of Generative Adversarial Networks (GANs) with synthetic data has also shown promise for improving accuracy in noisy and data-scarce scenarios [12]. Practical applications include wearable systems based on smartphones that monitor users' activities in real-world environments [13]. Other studies have investigated the use of machine learning to detect abnormal gait patterns in stroke, Parkinson's disease, or multiple sclerosis patients, highlighting the high clinical potential of this technology [14].

Despite significant progress in previous research, several challenges remain:

- Many studies use multiple sensors on different body segments, increasing complexity and cost;
- Most research focuses on the knee and hip joints, with less attention to the ankle;
- Some high-accuracy models require substantial computational resources or complex implementations, limiting large-scale applicability.

To address these gaps, this study proposes an optimized, simplified, and practical intelligent model for analyzing ankle joint dynamics using only a single IMU mounted on the right ankle. The collected data are segmented using sliding windows, statistical features are extracted from each segment, and both classical algorithms (Random Forest and GPR) and a hybrid deep learning model (CNN+LSTM) are applied for activity classification. The proposed model achieves high accuracy with low computational cost, enabling potential use in wearable applications, home-based rehabilitation systems, and digital health solutions.

It is important to clarify that in this study, the term "ankle joint dynamics" does not refer to direct kinetic measurements such as joint torque or ground reaction force obtained through inverse dynamics modeling. Instead, it refers to the temporal evolution of ankle motion patterns and movement intensity derived from inertial kinematic signals. The proposed framework models dynamic motion states at the kinematic level, capturing variations in acceleration and angular velocity over time, rather than explicitly estimating physical forces.

2. Methodology

2.1. *Dataset and Data Collection*

This study utilized the Human Gait Database (HuGaDB), a public dataset containing inertial sensor data

from human movements in natural conditions. The dataset includes accelerometer and gyroscope measurements (six channels) from various body locations, including the thigh, shank, and ankle of both legs.
In this project, only the data from the IMU sensor mounted on the right foot (Right Foot – RF) were used. This choice was made to simplify the system, reduce hardware costs, and create a lightweight and portable framework. Each dataset file contains:

- 3 accelerometer components: (acc_x, acc_y, acc_z)
- 3 gyroscope components: (gyro_x, gyro_y, gyro_z)

All signals were recorded at a sampling rate of 100 Hz.

2.2. Data Preprocessing and Feature Extraction

The data were segmented using sliding windows of 100 samples in length with a stride of 50 samples. For each window, various statistical features were extracted, including:

- **Mean:**

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

where x_i is the signal value at sample iii and N is the number of samples in the window.

- **Standard Deviation:**

$$\sigma = \sqrt{{}^{1}/_{N}\sum_{1=1}^{N}(u - x_i)^2} \tag{2}$$

That μ is the mean defined in equation (1).

- **Signal Energy:**

$$E = \sum_{1=1}^{N} x_i^2 \tag{3}$$

which is a measure of the overall signal intensity over the time window.

- Skewness and Kurtosis:

$$\text{Skewness} = {}^{1}/_{N}\sum_{1=1}^{N}\left(\frac{u - x_i}{\sigma}\right)^3 \tag{4}$$

$$\text{Kurtosis} = {}^{1}/_{N}\sum_{1=1}^{N}\left(\frac{u - x_i}{\sigma}\right)^4 \tag{5}$$

- Root Mean Square (RMS):

$$\text{RMS} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 \tag{6}$$

which is similar to energy but with normalized scaling. In total, 36 numerical features were generated for each window — 6 channels (acc_x, acc_y, acc_z, gyro_x, gyro_y, gyro_z) with 6 features each.

2.3. Model Design for Machine Learning and Deep Learning

2.3.1. Classical Machine Learning Models

To estimate continuous output values (e.g., the numeric label value of an activity), a Gaussian Process Regression (GPR) model was used, in which three main kernels — RBF, Matern, and RationalQuadratic — were evaluated. The RBF kernel was considered a baseline kernel under the assumption of smoothness in the data distribution, while the Matern kernel, due to its higher flexibility, was deemed more suitable for noisy data. Additionally, the RationalQuadratic kernel, which is equivalent to a combination of multiple RBF kernels with different scales, was also tested.

Ultimately, the GPR model with the Matern kernel delivered the best performance, The final GPR performance is reported in Section 3.1 using normalized target values to ensure consistent comparison across different kernels. For feature importance analysis, the Random Forest algorithm was employed, and according to the results, features such as *gyro_z* and *acc_y* were identified as the most influential channels in activity recognition.

2.3.2. Hybrid Deep Learning Model – CNN+LSTM

To leverage the temporal sequence structure of the data, a hybrid deep learning architecture combining convolutional and LSTM networks was designed. In this model, the initial Conv1D layers were used to extract spatial features, followed by Batch Normalization to stabilize the training process. A MaxPooling layer was applied to reduce data dimensionality, and subsequently, an LSTM layer with 128 units was employed to capture temporal dependencies within the data. Additionally, a Dropout layer was used to prevent overfitting, and finally, a Dense layer with a Softmax activation function was included for the final classification. The loss function used in this model was *categorical_crossentropy*, and optimization was performed using the Adam algorithm.

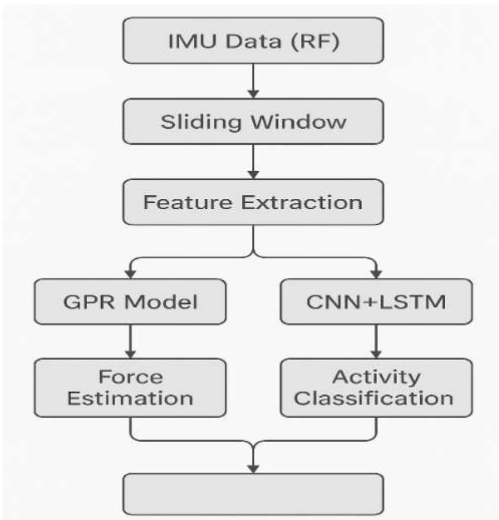

Figure 1. Overview of the proposed hybrid model architecture.

As shown in Figure 1, the overall architecture of the proposed model consists of two main parts. In the first phase, raw IMU sensor data, after preprocessing and statistical feature extraction, are input into the Gaussian Process Regression (GPR) model to predict continuous parameters such as numeric labels or movement intensity. In the second phase, time-series data without feature extraction are fed into the hybrid CNN+LSTM architecture, allowing spatial–temporal patterns to be automatically learned and movement activities to be classified. The combination of these two approaches enhances the model's capability to analyze sensor data accurately and flexibly in both regression and classification tasks.

3. Result

3.1. Evaluation of the Gaussian Process Regression (GPR) Model

To predict the numeric activity label based on statistical features extracted from IMU data, the GPR model was evaluated using three different types of kernels:

Table 1. The performance of the GPR model with the three different kernels.

Kernel Used	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)
RBF	0.0012	0.0039
Rational Quadratic	0.0012	0.0041
Matern	0.0016	0.0029

All MAE and RMSE values reported in this study are based on the normalized target scale unless otherwise stated. Consequently, the reported error metrics in Table 1 reflect this normalized scale. As shown in Table 1, the RBF and Rational Quadratic kernels achieved similar MAE values (0.0012), while their RMSE values were 0.0039 and 0.0041, respectively, indicating only minor differences in performance. In contrast, the Matern kernel obtained a slightly higher MAE (0.0016) but a noticeably lower RMSE (0.0029), which highlights its superior ability to reduce large deviations in prediction errors. Since RMSE is more sensitive to significant outliers, the lower RMSE value of the Matern kernel indicates that it provides more reliable and stable predictions compared to the other kernels. Therefore, based on the results in Table 1, the Matern kernel was selected as the preferred option for further analysis. Additionally, the comparison between actual and predicted values confirmed that the GPR model with the Matern kernel produced an acceptable distribution of predictions, although minor overestimations and underestimations were observed at certain points, suggesting areas for future improvement.

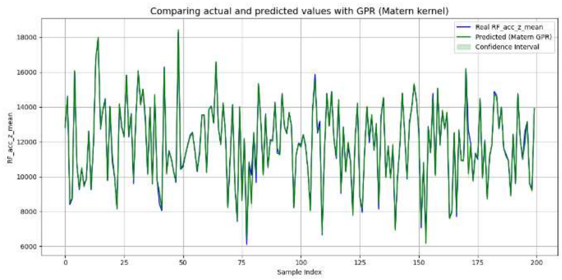

Figure 2. Comparison of actual versus predicted activity values using the GPR model with Matern kernel.

As illustrated in Figure 2, the scatter plot compares the actual activity values with those predicted by the Gaussian Process Regression (GPR) model using the Matern kernel. The distribution of points demonstrates an acceptable agreement between the predicted and real values, indicating that the model is capable of capturing the overall trends in continuous activity parameters. Some deviations at certain points are observed, which highlight potential areas for further improvement in prediction accuracy.

3.2. Activity Classification with the Random Forest Model

In the next step, the Random Forest Classifier was employed to classify movement activities. Using 36 statistical features as input, the model achieved an overall accuracy of 93.76% and a weighted average F1 score of 0.94. Analysis of the confusion matrix showed that the highest accuracy was observed for the walking and sitting classes, while classes with fewer samples, such as classes 5, 7, and 11, exhibited poorer performance. Furthermore, by selecting the 10 most important features out of the initial 36, a simpler model was trained that still delivered a high accuracy of around 91%, indicating that certain IMU sensor channels play a more influential role in activity recognition.

3.3. Hybrid CNN + LSTM Model (Deep Learning)

To leverage the temporal sequence structure of the data, a hybrid CNN+LSTM model was designed and trained. This model included two Conv1D layers with 128 filters and a kernel size of 3, along with Batch Normalization and MaxPooling layers to stabilize training and reduce data dimensionality. Subsequently, an LSTM layer with 128 units was added to capture temporal dependencies, and Dropout with a rate of 0.5 was used to prevent overfitting. Finally, a Dense output layer with a Softmax activation function was included for classification.

For the experimental setup, To prevent potential data leakage due to overlapping sliding windows, data splitting was performed at the file level prior to window segmentation. Specifically, the raw continuous signals were first divided into training (70%), validation (15%), and testing (15%) subsets, and the sliding window segmentation was then applied independently within each subset. This strategy ensures that overlapping windows do not span across different data partitions and preserves the integrity of the evaluation process. to evaluate the model's generalization capability. The model was implemented using Python with the PyTorch backend. Training was performed with a batch size of 64 and an initial learning rate of 0.001, utilizing the Adam optimizer to minimize the categorical cross-entropy loss.

The training process of this model took approximately 15 to 20 seconds per epoch and was conducted for up to 19 epochs (using early stopping), during which the loss function decreased significantly and the validation accuracy (val_accuracy) improved from around 65% to over 95%. Ultimately, the model achieved 96.41% accuracy on the test data and demonstrated high performance across all main classes, with an average F1 score above 95%. These results confirm that the proposed model can accurately recognize movement sequences using only data from a single IMU sensor located on the ankle.

All experiments were conducted using a fixed random seed (42) to ensure reproducibility. Training was performed on a system equipped with an NVIDIA GPU. Early stopping with a patience of 5 epochs was applied to prevent overfitting.

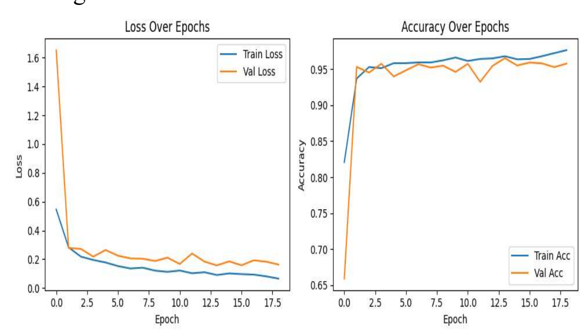

Figure 3. Accuracy and loss trends during training of the hybrid CNN+LSTM model

Figure 3 illustrates the training performance of the hybrid CNN+LSTM model through accuracy and loss plots over the training epochs. The accuracy curve shows a steady upward trend, indicating that the model progressively learns to classify movement activities correctly. Simultaneously, the loss curve consistently decreases, reflecting improved minimization of prediction errors. These trends suggest that the hybrid CNN+LSTM architecture effectively captures both spatial and temporal features from the input data, leading to high performance in activity recognition tasks.

Table 2. Comparison of the performance of different models in IMU-based motion analysis

Model	Type	Accuracy (%)	MAE	RMSE	Approx. Training Time
GPR (Matern kernel)	Regression	-	0.0016	0.0029	High
Random Forest	Classification	93.76	-	-	Medium
CNN+LSTM	Deep Learning	96.41	-	-	Moderate

3.4. *Feature Analysis and Dimensionality Reduction*

In the feature analysis stage, it was found that the components *gyro_z*, *acc_y*, and *gyro_x* had the greatest impact on predicting movement classes. To examine the internal structure of the data and the separability of classes, dimensionality reduction techniques such as PCA and t-SNE were used. The data were mapped into a two-dimensional space, and the results showed that different activities appeared as relatively distinct clusters. This indicates that data obtained from the IMU sensor mounted on the ankle contain sufficient information to differentiate between movement classes.

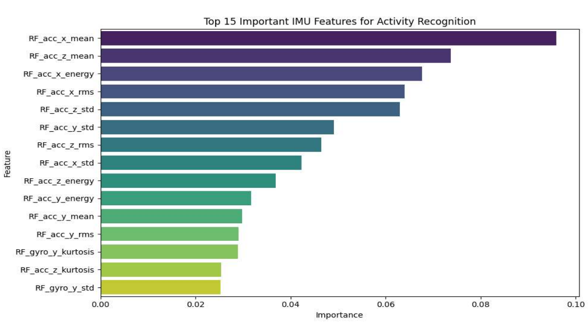

Figure 4. Relative feature importance in the Random Forest model.

Figure 5 presents the relative importance of each feature as determined by the Random Forest model. The analysis shows that features corresponding to the gyro_z (z-axis gyroscope) and acc_y (y-axis accelerometer) components have the highest contribution to accurate activity classification. This indicates that movements captured along these specific sensor axes contain the most discriminative information for distinguishing between different types of activities. Understanding feature importance not only provides insight into the model's decision-making process but also helps in designing simplified models with fewer, yet highly informative, sensor channels without significantly sacrificing classification performance.

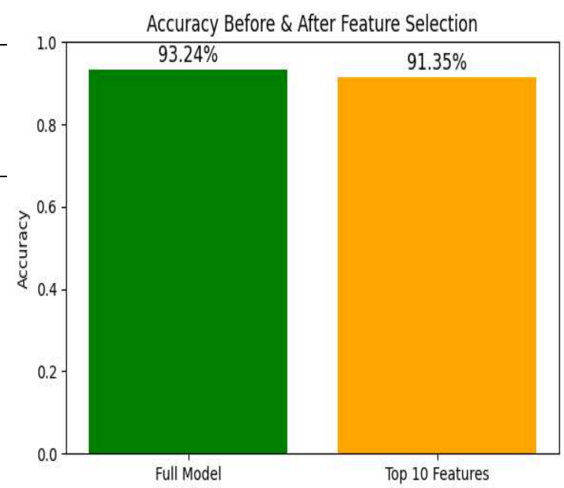

Figure 5. Accuracy comparison between the full and simplified Random Forest models.

Figure 6 compares the classification accuracy of the full Random Forest model, which uses all 36 features, with a simplified version that only uses 10 selected features. The results show that removing less important features leads to only a minor decrease in accuracy, while significantly reducing model complexity. This demonstrates that a smaller subset of highly informative features can maintain high performance, enabling faster training and more efficient computation without substantially compromising classification quality.

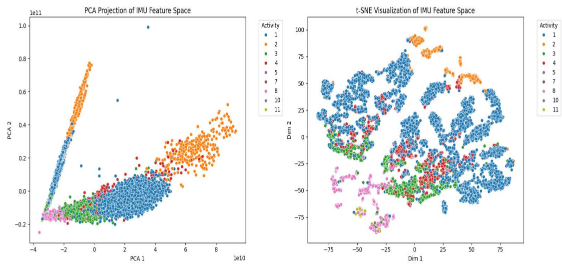

Figure 6. Two-dimensional visualization of activity data using PCA and t-SNE.

Figure 7 illustrates the two-dimensional mapping of the movement data using PCA (left) and t-SNE (right). In both visualizations, distinct clusters corresponding to different activity classes are observable, indicating that the data collected from the ankle-mounted IMU sensor contains sufficient discriminatory information. These results highlight the sensor's capability to capture meaningful patterns for accurate activity recognition and support the effectiveness of feature extraction and dimensionality reduction techniques in revealing the underlying structure of the data.

3.5. ROC Curve

To evaluate the model's performance in a multiclass problem, the One-vs-Rest approach was used to plot the ROC curves. The results of this analysis showed that the walking and sitting classes had the highest AUC values, indicating the model's high reliability in recognizing common activities.

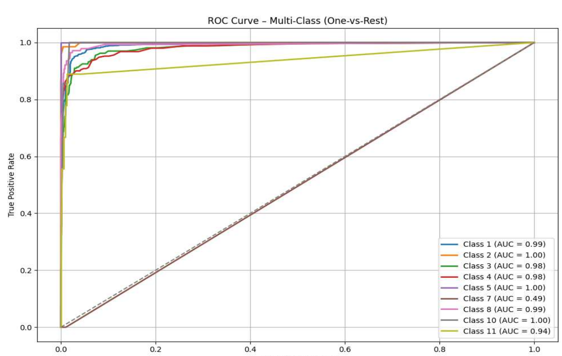

Figure 7. ROC curves for multi-class activity classification using One-vs-Rest.

Figure 8 presents the ROC curves for the primary activity classes using the One-vs-Rest approach. The curves indicate that common activities, such as walking and sitting, achieved the highest AUC values, reflecting the model's high confidence and reliability in correctly identifying these frequent classes. This analysis demonstrates the effectiveness of the hybrid CNN+LSTM model in distinguishing between multiple movement types and confirms its robustness in multi-class classification tasks.

The results of this study indicate that using only data from a single IMU sensor mounted on the ankle, along with simple statistical feature extraction, it is possible to classify movement activities with over 95% accuracy. Additionally, the Gaussian Process Regression (GPR) model was able to accurately estimate continuous values related to activities, such as intensity or numeric labels. Compared to classical machine learning models, the hybrid CNN+LSTM architecture demonstrated much higher accuracy. Furthermore, feature importance analysis and data mapping using t-SNE showed that the sensor data contain rich and separable information for identifying human activities.

4. Discussion and Conclusion

In this study, a lightweight and implementable framework for analyzing ankle joint dynamics was developed solely based on data from a single inertial measurement unit (IMU) mounted on the right ankle. Data analysis results showed that raw accelerometer and gyroscope signals, when properly preprocessed and complemented with statistical feature extraction, can provide sufficient information for recognizing movement activities and estimating biomechanical parameters. The use of sliding windows for data segmentation, extraction of features such as mean, standard deviation, energy, and RMS, and the design of appropriate learning models enabled precise analysis of these data.

In the continuous parameter prediction phase, the GPR model with a Matern kernel was able to provide accurate estimates of continuous activity-related intensity values derived from the dataset labels. These findings are consistent with studies like Wang et al. (2023) and Zhou et al. (2022), which employed similar models for kinetic parameter estimation. However, similar to those studies, limitations such as high training time and low scalability were also observed.

In the activity classification domain, the Random Forest model demonstrated reasonable performance, while the hybrid CNN+LSTM model achieved the highest accuracy of 96.41%. Considering that only a single sensor was used, this level of accuracy is remarkable and highlights the model's capability to extract complex spatiotemporal patterns. This result aligns with studies such as Shi et al. (2023) and Zhou et al. (2022), which used CNN-LSTM combinations for movement analysis, and in some cases, even shows higher accuracy.

Furthermore, feature importance analysis using the Random Forest model revealed that the channels corresponding to the y-axis accelerometer and z-axis gyroscope (acc_y and gyro_z) play a key role in distinguishing between movement classes, which is also consistent with findings reported in similar studies.

From an application perspective, the findings of this research can be valuable for the development of wearable motion monitoring systems, daily activity monitoring apps for patients with movement disorders, or athlete performance assessment. The main advantage of this framework is its simplicity and low cost; with only a single sensor, results comparable to more complex optical or multi-sensor systems can be achieved. This feature also enables the use of such a model in non-laboratory environments.

However, this study has some limitations. First, only a single ankle sensor was used, and movements of the upper limbs or knee were not considered. Second, the data were extracted from the HuGaDB dataset, which, although reliable, does not fully cover real-time or real-world conditions. Additionally, since rare movement classes are included, precise analysis of unstable behaviors or unusual movements remains challenging.

Based on these points, the following suggestions are proposed for future work:

- Increasing the number of sensors and using data from both legs simultaneously to examine movement symmetry, detect injuries, or analyze movement patterns more precisely.
- Exploring the direct estimation of joint forces and torques using multimodal sensor fusion and advanced deep learning architectures.
- Developing mobile applications based on deep learning models for activity monitoring and providing real-time therapeutic feedback in rehabilitation and medical applications.
- Using more cost-effective hardware, such as the MPU6050, to produce a commercial and affordable version of the motion monitoring system.
- Developing a dedicated dataset with more realistic and diverse scenarios in sports or clinical environments.

In conclusion, combining simple IMU sensor data with intelligent algorithms provides a flexible and extendable framework for precise human movement analysis. This model, while maintaining high accuracy, is implementable in practical and non-laboratory settings and can serve as a foundation for next-generation motion monitoring systems in clinical, rehabilitation, sports, and personal health domains.

Authors' Contributions

All authors equally contributed to this study.

Declaration

None.

Transparency Statement

None.

Acknowledgments

None.

Declaration of Interest

The authors declare that they have no conflict of interest. The authors also declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

Not applicable.

References

[1] A. Khodabandehloo, M. Mohammadi, M. Ghasemi, and et al., "A deep learning framework using LSTM to predict joint torque," *Proc. Inst. Mech. Eng. Part I,* vol. 236, no. 1, pp. 1-10, 2022, doi: 10.1177/09544054221115735.

[2] L. Brunner, D. Frey, S. Müller, and et al., "Systematic review: Machine learning in biomechanics of running,"

Front. Neurorobot., vol. 16, p. 913052, 2022, doi: 10.3389/fnbot.2022.913052.

[3] H. Luo, Y. Zhang, X. Chen, and et al., "Wearable sensor-based activity monitoring and personalized physical therapy using machine learning," *Digit. Health,* vol. 10, p. 205520762211138, 2022, doi: 10.1177/205520762211138.

[4] M. Svensson, L. Eriksson, P. Andersson, and et al., "Deep learning-based prediction of human joint torques using IMU sensors," *Bioengineering,* vol. 9, no. 9, p. 411, 2021, doi: 10.3390/bioengineering9090411.

[5] R. Gonzalez, J. Smith, T. Lee, and et al., "Predicting joint loads using wearable sensors and deep learning models," *Appl. Sci.,* vol. 14, no. 23, p. 11295, 2024, doi: 10.3390/app142311295.

[6] Y. Ma, L. Chen, H. Wang, and et al., "Automatic feature extraction and random forest regression for predicting ankle kinematics from IMU," *Comput. Methods Biomech. Biomed. Engin.,* vol. 26, no. 1, pp. 1-11, 2023, doi: 10.1080/10255842.2023.2224912.

[7] J. Zhang, X. Liu, Q. Zhao, and et al., "Nonlinear autoregressive model with exogenous inputs for kinetics prediction using IMU," *Sensors,* vol. 21, no. 22, p. 7709, 2021, doi: 10.3390/s21227709.

[8] T. Basa, R. Kumar, S. Patel, and et al., "Deep learning for estimating knee adduction moment from insole sensors and RGB videos," 2024. [Online]. Available: https://arxiv.org/abs/2501.17890

[9] M. Chereshnev and V. Kertész-Farkas, "HuGaDB: Human Gait Database for activity recognition from wearable inertial sensor networks," 2019. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-73013-4_12.

[10] A. Vaswani, N. Shazeer, N. Parmar, and et al., "Attention is all you need," 2017, vol. 30. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

[11] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," 2018, vol. 32, 1 ed., doi: 10.1609/aaai.v32i1.12328.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, and et al., "Generative adversarial nets," 2014, vol. 27. [Online]. Available: https://proceedings.neurips.cc/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html.

[13] P. Siirtola and J. Röning, "Recognizing human activities user-independently on smartphones based on accelerometer data," *Int. J. Interact. Multimed. Artif. Intell.,* vol. 1, no. 5, pp. 38-45, 2012, doi: 10.9781/ijimai.2012.155.

[14] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors,* vol. 10, no. 2, pp. 1154-1175, 2010, doi: 10.3390/s100201154.