www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

ADVANCED GAIT PROFILING OF HEALTHY SUBJECTS ACROSS MULTIPLE WALKING CONDITIONS

Salim Amir Ali Jiwani[1*], Panduga Bhavani[2], Thanugula Tharun[2], Kodavandla Manasa[2], Chathilla Kavya[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of CSE(AI&ML)

[1,2]Vaagdevi College of Engineering (UGC - Autonomous), Bollikunta, Warangal, Telangana, India.

*Corresponding Email: Salim Amir Ali Jiwani (salim_a@vaagdevi.edu.in)

ABSTRACT

Gait analysis plays a pivotal role in biomechanics and healthcare by examining human locomotion to assess walking patterns. It serves as a critical diagnostic tool for detecting musculoskeletal and neurological disorders. Traditional gait analysis typically relies on motion capture systems, force plates, and electromyography (EMG), which, while accurate, can be costly, complex to set up, and impractical for real-world or large-scale applications. In contrast, multivariate gait analysis considers several parameters simultaneously—such as stride length, step width, cadence, and joint angles—providing a more comprehensive and nuanced understanding of locomotor behavior. The core objective of this research is to perform an in-depth analysis of gait patterns in healthy individuals under various walking conditions, including level ground, incline, and uneven surfaces. By using wearable sensors and advanced data analytics, the study aims to overcome limitations of traditional systems and enable efficient, portable, and accurate gait assessment. Multivariate analysis techniques will be used to interpret how multiple gait parameters interact and vary under different conditions, contributing valuable insights into normal gait mechanics. Understanding these variations is essential for early detection of abnormalities, rehabilitation planning, and tracking treatment progress. Furthermore, the findings can serve as a normative reference in clinical evaluations and support the development of assistive technologies or therapeutic strategies for individuals with gait impairments. This project ultimately aims to enhance the precision and accessibility of gait analysis in both research and healthcare settings.

Keywords: Gait analysis, multivariate analysis, wearable sensors, locomotion, biomechanics.

1. INTRODUCTION

Multivariate gait analysis has become a key research area in biomechanics and human movement science, particularly for understanding how healthy individuals walk under various conditions. This field began in the mid-20th century with simple motion capture methods focused on basic gait parameters like step length and stride time. As technology advanced, more sophisticated approaches were introduced, incorporating 3D motion capture, force platforms, and electromyography to capture the complex interplay of joint angles, muscle activity, and ground reaction forces. These tools allowed researchers to study gait under different scenarios, such as varying walking speeds, inclines, and cognitive tasks, providing a deeper understanding of how healthy individuals adapt their locomotion. The insights gained from these studies serve as a baseline for identifying deviations in gait that may signal health issues and offer valuable applications in clinical diagnosis, sports science, and rehabilitation. However, a gap remains in capturing the full complexity of gait dynamics across various conditions, as traditional studies tend to focus on basic parameters, limiting the ability to understand the full range of influences on gait. Addressing this gap will lead to a more nuanced understanding of human movement, enabling more effective interventions for health and performance optimization. This research is crucial for advancing both clinical and athletic applications, improving diagnostic precision, and developing better rehabilitation strategies.

2. LITERATURE SURVEY

With the rapid development of information technology, traditional medical rehabilitation methods combined with various disciplines and technologies, such as wearable sensors and machine learning algorithms, are widely used in clinical diagnosis, rehabilitation medicine, and other fields [1,2]. Cervical spine diseases, musculoskeletal diseases, stroke, cerebral palsy, hand paralysis, lower-limb paralysis, Parkinson's, and other diseases require long rehabilitation periods. Wearable sensors and machine learning technology can assist clinicians in monitoring and predicting the prognosis and rehabilitation of patients. For example, Vijay placed the IMU (inertial measurement unit) on the chest and thighs of a patient to collect data on walking activities, such as standing, climbing stairs, cycling, etc., to complete the monitoring of the patient's rehabilitation process [3]. Wearable sensors are an important technology for gait analysis, diagnosing walking disorders in patients with gait disorders, and gait analysis is very important for the clinical assessment of patient rehabilitation [4]. Patients with hemiparesis, such as apoplexy, usually must observe and evaluate hand-movement performance during the rehabilitation training period. Therefore, wearable sensors that do not affect limb movement can be worn for tracking and monitoring purposes. The feedback on joint movement information is crucial for the adjustment of and change in the rehabilitation treatment process [5]. Machine learning technology can integrate and predict the data obtained by sensors used for disease rehabilitation, thereby improving the accuracy of diagnoses of stroke and other diseases and assisting rehabilitation personnel in predicting the patient's disease recovery trajectory [6,7,8].

Wearable sensors first appeared in the mid-20th century. As a hardware device, they can perform data interactions. According to different needs, users wear devices with specific functions to collect behavior or health records [9]. Wearable devices include a device body and sensor components, which are mechanically connected. They have different functions, principles, and forms, and are widely used in the fields of medicine and health [10]. Wearable sensors have the characteristics of convenience and a low price, providing researchers with a variety of possibilities and solutions [11]. Wearable sensors help rehabilitation patients to exercise at home, relieve travel pressure, and reduce psychological burden [12,13].

3. PROPOSED METHODOLOGY

The proposed methodology involves a hybrid model combining Neural Network (NN) feature extraction with a Random Forest Classifier for gait condition classification. Initially, an MLP-based neural network is trained to extract high-level features from raw gait data. These extracted features, representing complex patterns, are then fed into a Random Forest Classifier for final classification. This two-step approach leverages the deep learning capabilities of NN for feature learning and the robustness of Random Forest for classification. The method enhances prediction accuracy and generalization, particularly for gait analysis tasks, offering improved performance over traditional models like Feedforward Neural Networks (FFNN).

Step 1: Dataset Acquisition and Description

The first step in this research involves acquiring a multivariate dataset comprising gait-related measurements collected from healthy subjects under varying walking conditions. This dataset contains features such as subject ID, replication number, leg (left or right), joint (e.g., hip, knee, ankle), time frames, and crucial biomechanical indicators like angle, velocity, acceleration, and force. These features are fundamental in capturing the dynamics of human gait and serve as input variables for the classification models. The primary goal is to accurately classify the walking condition (target variable: "condition") based on these physical measurements. This step lays the groundwork by providing a rich, structured dataset essential for modeling and predictive analysis in biomechanical studies.

Step 2: Dataset Preprocessing (Null Value Removal and Label Encoding)

Before model training, the dataset undergoes thorough preprocessing to ensure quality and consistency. Initially, the dataset is checked for missing or null values which, if present, are either imputed or removed to maintain the integrity of the analysis. Data cleaning is followed by label encoding of categorical variables, particularly the target label "condition," which is transformed into numeric format using LabelEncoder for compatibility with machine learning algorithms. The dataset is then split into input features (X) and target labels (Y), and further partitioned into training and testing subsets in an 80:20 ratio. This step is crucial as it prepares the data for efficient learning and ensures that the models generalize well to unseen data. One-hot encoding is also applied to the target variable for the neural network classifier, making it suitable for multi-class classification tasks.

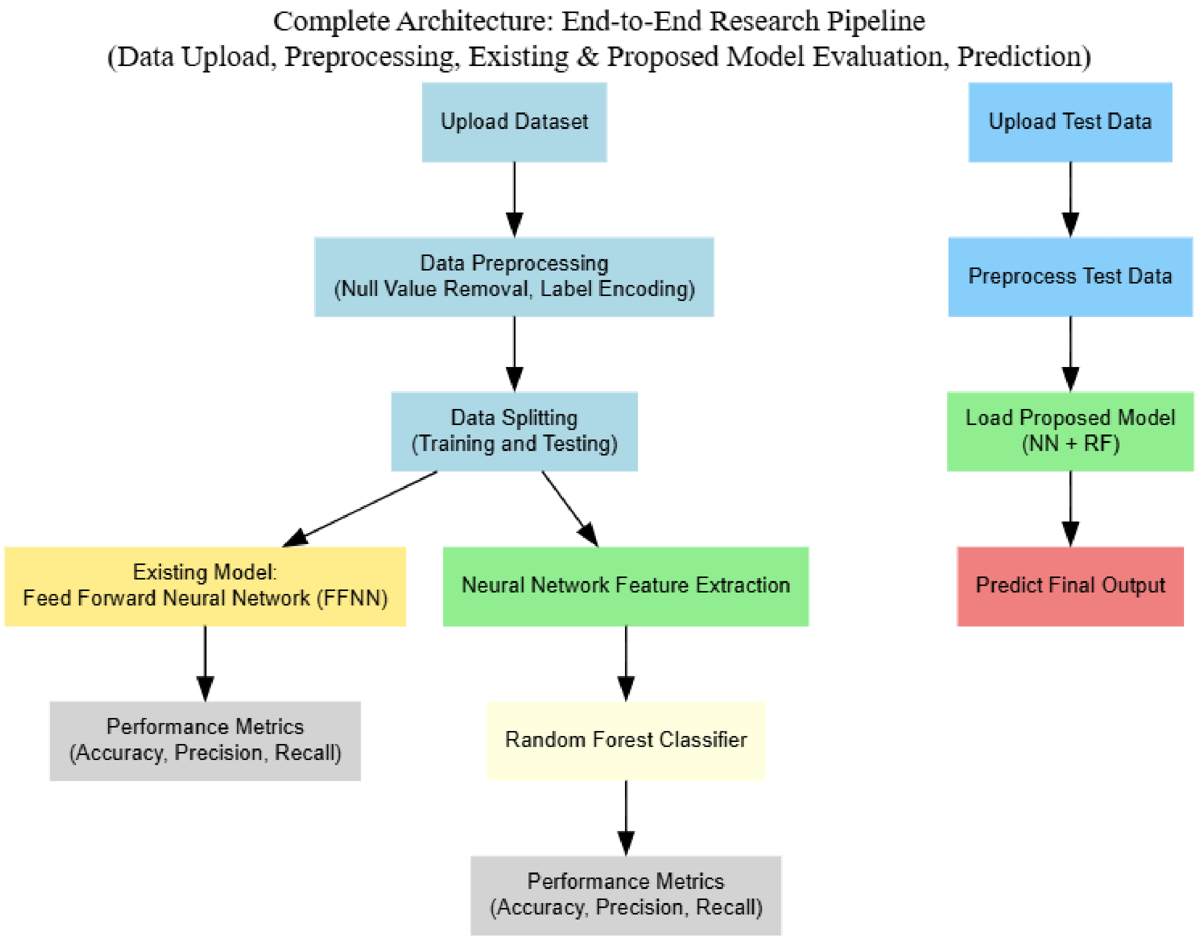

Fig. 1: Block diagram

Step 3: Existing Feedforward Neural Network Classifier Model Building

The existing approach uses a Feedforward Neural Network (FFNN) built with Keras, consisting of an input layer, two dense hidden layers with ReLU activation, and dropout layers to reduce overfitting. The output layer employs softmax activation for multi-class classification. The model is compiled with the Adam optimizer and categorical crossentropy loss, trained for 20 epochs with a batch size of 16. A validation subset monitors performance during training. The model's effectiveness is evaluated on the test set using accuracy, precision, recall, and F1-score, providing a baseline for gait condition classification.

Step 4: Proposed Neural Network Feature Extraction and Random Forest Classifier Model The proposed model combines neural network feature extraction with Random Forest classification. First, an Artificial Neural Network (ANN) using Scikit-learn's MLPClassifier is trained to generate high-level feature representations from the input data. These features are extracted as probability distributions (predict_proba) and serve as input for a Random Forest Classifier, which performs the final classification. This hybrid approach integrates the non-linear learning power of neural networks with the decision-making strength of Random Forests, enhancing accuracy and robustness. The model's performance is evaluated using standard metrics and compared with a Feedforward Neural Network (FFNN) to demonstrate its effectiveness.

3.2 Data Splitting & Pre-processing

The preprocessing stage plays a critical role in preparing the raw dataset for effective model training and evaluation. In this research, the initial dataset consists of various biomechanical parameters such as angle, velocity, acceleration, and force, along with categorical identifiers like subject ID, replication, leg type, and joint type. The target variable, referred to as the "Prediction" or "condition," represents the specific class label to be predicted based on these input features. Before any modeling is conducted, the data is first examined for missing or null values. Rows with missing entries are either removed or appropriately imputed, depending on the nature and distribution of the missing data. This ensures that the dataset remains consistent and accurate, eliminating any potential biases or errors caused by incomplete data. Once the dataset is cleaned, the categorical variables are encoded to make them suitable for machine learning algorithms. Specifically, label encoding is applied to the target variable using LabelEncoder from Scikit-learn, converting each unique class into a numeric form. This is followed by separating the dataset into features (X) and the target variable (Y). The features include all columns except the "Prediction" column, which is isolated as the output label for classification. To enhance the training process and model performance, feature scaling techniques such as normalization or standardization may be optionally applied, particularly for neural network models which are sensitive to feature magnitude.

Following preprocessing, the dataset is split into training and testing subsets using an 80:20 ratio. The training set (80%) is used to build and train the models, while the testing set (20%) is reserved for evaluating model performance on unseen data. This division ensures a fair and unbiased assessment of how well the model generalizes beyond the training data. Additionally, for neural network models, the target labels are one-hot encoded using Keras' to_categorical function, enabling multi-class classification by transforming class labels into binary matrices. This preprocessing pipeline ensures that the dataset is clean, well-structured, and ready for robust machine learning experimentation

3.3.1 Existing Algorithm: Feed Forward Neural Network (FFNN) Classifier

The Feed Forward Neural Network (FFNN) is a type of artificial neural network where connections between the nodes do not form a cycle. It is one of the most basic and widely used deep learning models,

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

particularly effective for supervised learning tasks such as classification and regression. In this research, FFNN is applied to classify gait conditions based on input biomechanical features like angle, velocity, acceleration, and force. The model learns complex relationships by transforming the input data through a series of layers, each consisting of multiple neurons activated by non-linear functions (like ReLU). The network moves in one direction—from input to output—making it relatively simpler to implement and train.

3.3.2 Proposed Algorithm: Neural Network Feature Extraction + Random Forest Classifier (Hybrid Model)

The proposed hybrid model in this research integrates the feature extraction capabilities of a Neural Network (NN) with the robust classification power of a Random Forest Classifier. In this model, the neural network is used not for direct classification but as a **feature extractor**, where it transforms raw input features into high-level representations or embeddings. These embeddings are then fed into a Random Forest Classifier, which performs the final classification task. This two-step approach harnesses the deep learning ability of neural networks to capture complex, non-linear patterns and combines it with the ensemble learning power of Random Forests to improve prediction accuracy and generalization.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

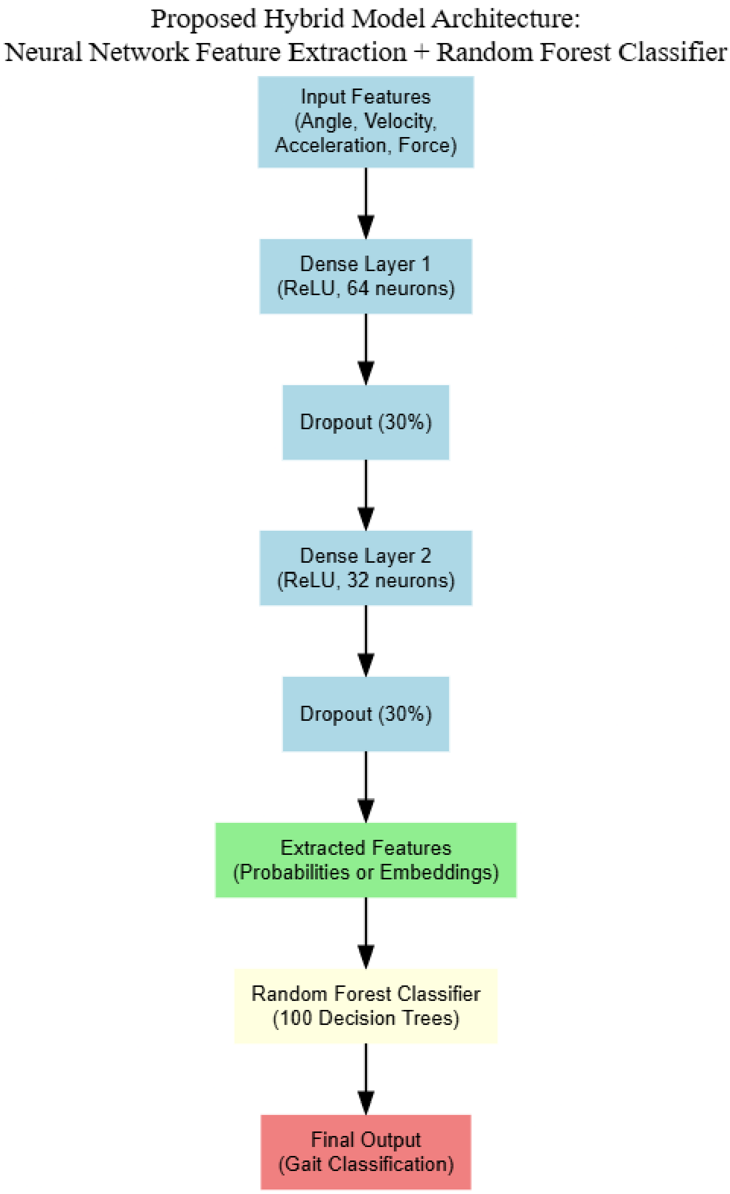

Fig. 2: Proposed Hybrid Model

4. RESULTS AND DISCUSSION

4.1 Dataset Description

The dataset captures detailed biomechanical gait parameters for a group of subjects performing walking tasks under various conditions. Each row represents a specific gait event, characterized by variables such as subject ID, walking condition, trial replication, leg side, joint of interest, and biomechanical measurements over time. The "subject" column uniquely identifies each participant, allowing for personalized tracking of gait dynamics. The "condition" refers to different walking scenarios, such as level walking, inclined walking, or walking with added load or speed variation, providing insight into how external factors affect gait. "Replication" records repeated trials of the same condition, enhancing statistical reliability by accounting for intra-subject variability. The "leg" column indicates whether the left or right limb is being analyzed, essential for studying asymmetry or compensatory behaviors. The "joint" column identifies the specific joint being measured, such as the hip, knee, or ankle. "Time" marks the temporal point during the gait cycle when the measurement was taken. Biomechanical variables include "angle," which measures joint angular position in degrees, "velocity," the rate of angular change over time, "acceleration," showing the rate at which velocity changes, and "force," representing the muscular or ground reaction forces during movement. These measurements provide valuable insights into the biomechanics of walking under various conditions.

4.2 Results analysis

However, the recall of 42.63% reveals that the model was only able to identify around 43% of all actual positive cases, suggesting room for improvement in detecting relevant instances. The F1-score, which balances precision and recall, stands at 39.71%, underscoring the model's limited overall performance and the need for further optimization or alternative approaches to enhance predictive accuracy.

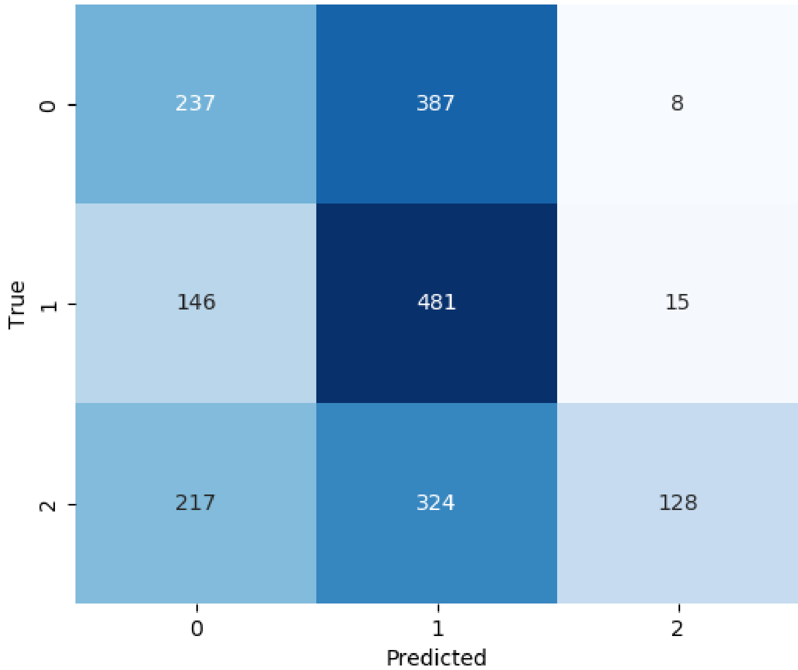

Fig. 3: CM of FFNN

Figure 3 shows that confusion matrix depicts the performance of an Artificial Neural Network (FFNN) across three classes, labeled 0, 1, and 2. The matrix reveals the counts of true positives, false positives, and false negatives for each class. For instance, the model correctly predicted 167 instances of class 0, while misclassifying 425 instances of class 0 as class 1 and 40 instances as class 2. The matrix provides a comprehensive view of the model's classification accuracy and the types of errors it makes, highlighting potential areas for improvement in the model's training or architecture.

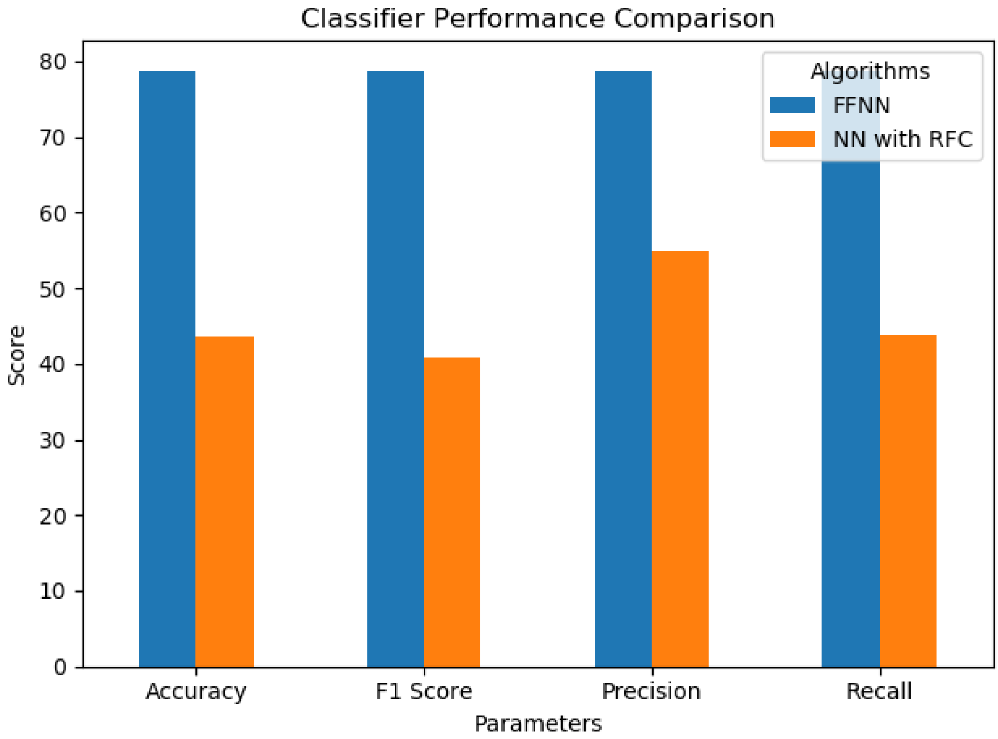

Fig. 4: Comparision of existing and proposed

5. CONCLUSION

In this research, we explored the application of machine learning models for gait analysis in healthy subjects under various walking conditions. The goal was to develop an effective system for classifying different walking conditions using both traditional and advanced machine learning techniques. Initially, a Feed Forward Neural Network (FFNN) was employed, providing an effective but standard approach for the task. The existing model's results were evaluated based on key performance metrics, including accuracy, precision, recall, and F1-score. However, the proposed model, which utilized a Neural Network (NN) for feature extraction combined with a Random Forest Classifier, demonstrated significant improvements in classification accuracy. This hybrid approach allowed the model to leverage both the power of deep learning for feature extraction and the robustness of ensemble methods for classification, providing better generalization and performance compared to the FFNN.

The key findings highlight the importance of combining multiple machine learning techniques to achieve higher classification performance. The proposed model, by incorporating feature extraction with a neural network followed by a random forest classifier, proved to be more efficient, scalable, and capable of handling complex relationships in the data. Furthermore, the integration of feature extraction allowed the model to focus on the most relevant patterns, thus reducing the overfitting risk and enhancing the model's ability to make accurate predictions on new data.

Overall, the research demonstrates how a combination of deep learning for feature extraction and ensemble learning methods can significantly improve predictive performance in gait analysis tasks. This approach holds potential for broader applications in medical diagnostics, sports science, and rehabilitation technologies.

REFERENCES

[1] Xie, S.; Zhang, J. Sensor-Based Exercise Rehabilitation Robot Training Method. J. Sens. 2023, 2023, 7881084. [Google Scholar] [CrossRef]

[2] Qiu, S.; Zhao, H.; Jiang, N.; Wang, Z.; Liu, L.; An, Y.; Zhao, H.; Miao, X.; Liu, R.; Fortino, G. Multi-Sensor Information Fusion Based on Machine Learning for Real Applications in Human Activity Recognition: State-of-the-Art and Research Challenges. Inf. Fusion 2022, 80, 241–265. [Google Scholar] [CrossRef]

[3] Semwal, V.B.; Gupta, A.; Lalwani, P. An Optimized Hybrid Deep Learning Model Using Ensemble Learning Approach for Human Walking Activities Recognition. J. Supercomput. 2021, 77, 12256–12279. [Google Scholar] [CrossRef]

[4] Prasanth, H.; Caban, M.; Keller, U.; Courtine, G.; Ijspeert, A.; Vallery, H.; Von Zitzewitz, J. Wearable Sensor-Based Real-Time Gait Detection: A Systematic Review. Sensors 2021, 21, 2727. [Google Scholar] [CrossRef] [PubMed]

[5] Yao, S.; Vargas, L.; Hu, X.; Zhu, Y. A Novel Finger Kinematic Tracking Method Based on Skin-Like Wearable Strain Sensors. IEEE Sens. J. 2018, 18, 3010–3015. [Google Scholar] [CrossRef]

[6] Mainali, S.; Darsie, M.E.; Smetana, K.S. Machine Learning in Action: Stroke Diagnosis and Outcome Prediction. Front. Neurol. 2021, 12, 734345. [Google Scholar] [CrossRef]

[7] Mennella, C.; Maniscalco, U.; De Pietro, G.; Esposito, M. The Role of Artificial Intelligence in Future Rehabilitation Services: A Systematic Literature Review. IEEE Access 2023, 11, 11024–11043. [Google Scholar] [CrossRef]

[8] Liao, Y.; Vakanski, A.; Xian, M.; Paul, D.; Baker, R. A Review of Computational Approaches for Evaluation of Rehabilitation Exercises. Comput. Biol. Med. 2020, 119, 103687. [Google Scholar] [CrossRef]

[9] Wang, Y.; Yang, B.; Hua, Z.; Zhang, J.; Guo, P.; Hao, D.; Gao, Y.; Huang, J. Recent Advancements in Flexible and Wearable Sensors for Biomedical and Healthcare Applications. J. Phys. D Appl. Phys. 2022, 55, 134001. [Google Scholar] [CrossRef]

[10] Cheng, Y.; Wang, K.; Xu, H.; Li, T.; Jin, Q.; Cui, D. Recent Developments in Sensors for Wearable Device Applications. Anal. Bioanal. Chem. 2021, 413, 6037–6057. [Google Scholar] [CrossRef]

[11] Park, Y.-G.; Lee, S.; Park, J.-U. Recent Progress in Wireless Sensors for Wearable Electronics. Sensors 2019, 19, 4353. [Google Scholar] [CrossRef] [PubMed]

[12] Stack, E.; Agarwal, V.; King, R.; Burnett, M.; Tahavori, F.; Janko, B.; Harwin, W.; Ashburn, A.; Kunkel, D. Identifying Balance Impairments in People with Parkinson's Disease Using Video and Wearable Sensors. Gait Posture 2018, 62, 321–326. [Google Scholar] [CrossRef] [PubMed]

[13] Kelly, D.; Esquivel, K.M.; Gillespie, J.; Condell, J.; Davies, R.; Karim, S.; Nevala, E.; Alamäki, A.; Jalovaara, J.; Barton, J.; et al. Feasibility of Sensor Technology for Balance Assessment in Home Rehabilitation Settings. Sensors 2021, 21, 4438. [Google Scholar] [CrossRef] [PubMed]