



ToeFro-Varus: A wearable IMU-based system for automated physiotherapy exercise classification and gamified feedback in pediatric clubfoot rehabilitation

Nadkarni V¹, Balasubramanian S²

Submitted: October 12, 2025, Revised: version 1, November 16, 2025, version 2, January 26, 2026, version 3, January 26, 2026, version 4, January 31, 2026, version 5, February 18, 2026
 Accepted: February 27, 2026

Abstract

Congenital Talipes Equinovarus (CTEV), or clubfoot, affects approximately 1–2 per 1,000 live births worldwide. Although the Ponseti method achieves correction rates exceeding 90%, long-term outcomes depend on sustained adherence to physiotherapy and bracing, which is often suboptimal and difficult to monitor outside clinical settings. This study presents *ToeFro-Varus*, a proof-of-concept system that integrates an ankle-worn inertial measurement unit (IMU) with a gamified mobile application to enable automated monitoring of prescribed physiotherapy exercises in school-age children (5–18 years) in the maintenance phase of treatment. The wearable device, built using an ESP32 microcontroller and a BNO055 9-axis IMU, captures motion data across six clinically relevant exercises. A supervised machine learning pipeline using engineered time-series features was evaluated with a participant-wise split (17 participants: 11 training, 6 testing). The best-performing model (Gradient Boosting) achieved 95.9% classification accuracy on held-out participants; however, evaluation relied on a limited number of original trials ($n = 36$), with additional augmented samples used for training, and therefore these results should not be interpreted as evidence of generalization. This work demonstrates the technical feasibility of IMU-based exercise classification coupled with real-time gamified feedback for remote monitoring. The system enables objective identification of exercise type and provides a framework for linking physical activity to digital rewards. However, the study does not assess adherence, quality of execution, or clinical outcomes, and the generalizability of the model remains unestablished. The primary contribution is an end-to-end, low-cost prototype illustrating a scalable approach to automated physiotherapy monitoring, providing a foundation for future studies evaluating clinical effectiveness and real-world deployment.

Keywords

Congenital Talipes Equinovarus (CTEV), Clubfoot rehabilitation, Wearable sensors, Inertial Measurement Unit (IMU), Machine Learning in healthcare, Physiotherapy monitoring, Digital health, Remote patient monitoring, Gamification in healthcare, Pediatric rehabilitation

¹Varun Nadkarni, Dhirubhai Ambani International School, G Block, Plot No. 46, Near Titan Hotel, Trident Road, Bandra Kurla Complex, Bandra East, Mumbai, 400051, Maharashtra, India. varunnadkarni2k08@gmail.com

²Corresponding author: Sankar Balasubramanian, Department of Mechanical Engineering, Indian Institute of Science (IISc), C.V. Raman Road, Bengaluru (Bangalore), Karnataka, 560012, India. sankarb@iisc.ac.in

1. Introduction

1.1 Background

Congenital Talipes Equinovarus (CTEV), commonly known as clubfoot, is a congenital deformity in which one or both feet of a newborn are turned inward and downward (Figure 1). The severity of these deformities can vary, as illustrated in Figure 2, and is typically attributed to shortened or abnormally developed musculoskeletal tissues. If left untreated, clubfoot can lead to significant impairment in mobility and long-term disability.

CTEV is among the most common congenital musculoskeletal conditions, affecting infants from birth and potentially resulting in serious functional limitations if not managed appropriately (1). In India, the condition represents a substantial public health concern, with an estimated incidence of approximately 1.25 cases per 1,000 live births, translating to thousands of affected children annually (1). Globally, the prevalence ranges from 0.51 to 2.03 per 1,000 live births in low- and middle-income countries (2). The high incidence and risk of long-term disability underscore the importance of timely and effective intervention.



Figure 1. Feet of a CTEV patient.

CTEV exhibits a higher prevalence in males than females, and approximately 50% of cases are bilateral, affecting both feet (1). The etiology of clubfoot remains incompletely understood, although it is widely considered to result from a combination of genetic predisposition and environmental influences during pregnancy (3). If left untreated, the condition can lead to severe functional impairment, limiting mobility and resulting in lifelong disability (1).

In India, these challenges are further compounded by population size and disparities in healthcare access, particularly in rural regions, where delayed or inadequate treatment remains a significant concern (4).

Parameters	Mild	Moderate	Severe
Midfoot			
Curved lateral border	0	0.5	1
Medial foot crease	0	0.5	1
Talar head coverage	0	0.5	1
Hindfoot			
Posterior crease	0	0.5	1
Rigid equinus	0	0.5	1
Empty heel	0	0.5	1

Maximum score is 6; Minimum score is 0. Higher the score, the more severe the deformity.

Figure 2. Severity of deformities in patients.

1.2 The Ponseti method and the critical role of post-treatment physiotherapy

The Ponseti method is the most widely accepted and effective treatment for congenital talipes equinovarus (CTEV). It involves the gradual correction of the deformity through gentle manipulation of the infant's foot followed by serial plaster casting, typically performed weekly over several weeks. In most cases, a minor percutaneous Achilles tenotomy is required to achieve full correction. However, treatment does not conclude with casting. Long-term maintenance requires the use of a foot abduction brace and adherence to prescribed physiotherapy exercises to prevent recurrence of the deformity.

The Ponseti method is considered the global gold standard due to reported correction rates exceeding 90% without the need for extensive surgical intervention (5). Developed by Dr. Ignacio Ponseti, the technique follows a structured protocol of progressive correction through weekly manipulation and casting over approximately 5–8 weeks (6), achieving

successful outcomes in the majority of cases when applied correctly (7). Its non-surgical nature and relatively low resource requirements make it particularly suitable for implementation in low-resource settings, including rural regions of countries such as India (8).

Despite its high initial success rate, long-term outcomes are highly dependent on post-correction adherence. Following casting, children are typically required to wear a foot abduction brace, such as a Denis Browne splint, for several years and to perform regular physiotherapy exercises (9). These interventions are essential for maintaining correction and preventing relapse, defined as the recurrence of the deformity (10). Evidence indicates that children who adhere to bracing and physiotherapy protocols have significantly better outcomes compared to those with poor compliance (8).

1.3 Dual challenges in CTEV rehabilitation: patient adherence and remote monitoring

Despite the high success rate of the Ponseti

method, two major challenges persist in long-term CTEV management: poor patient adherence to physiotherapy and the lack of effective remote monitoring. Physiotherapy exercises are repetitive and must be performed daily over several years, often extending into adolescence. As a result, children—particularly those older than five years—frequently experience reduced motivation and engagement, leading to inconsistent or incorrect execution of exercises. This lack of adherence is a major contributor to relapse.

A second critical limitation is the absence of reliable systems for monitoring exercise performance outside clinical settings. While physiotherapists instruct caregivers on home-based exercise routines, there is currently no objective mechanism to verify whether exercises are performed consistently or correctly (11). Consequently, clinicians must rely on intermittent hospital visits to assess progress, limiting their ability to provide timely feedback or intervention (12).

It is important to define the scope of the problem addressed in this study. Our system targets children in the maintenance phase of CTEV treatment, typically aged 5–18 years, who have completed the intensive bracing protocol but require continued active physiotherapy to maintain correction and functional mobility. The monitored exercises are complementary to, not a replacement for, bracing protocols. Children aged 0–5 years undergoing the initial Ponseti treatment phase primarily require brace compliance and passive stretching performed by

caregivers—activities beyond the scope of this system.

Adherence to post-correction protocols remains a well-documented challenge. Studies indicate that long-term compliance with bracing and physiotherapy is often poor, largely due to the prolonged and repetitive nature of treatment (9,10). Children may lose motivation over time, particularly as visible symptoms improve, leading to premature discontinuation of therapy (13). In India, these challenges are exacerbated by socioeconomic and geographic factors. Many families reside in rural areas with limited access to healthcare facilities, making regular follow-up difficult (14). Additionally, a lack of awareness regarding the importance of continued therapy may result in early cessation of treatment once apparent correction is achieved (13). Evidence suggests that children with poor adherence are ~2.6 times more likely to experience relapse compared to those who follow prescribed protocols (10).

The lack of remote monitoring further compounds this issue. Currently, exercise performance is evaluated only during periodic clinical visits, which may occur months apart. As a result, incorrect execution or non-compliance may go undetected for extended periods, potentially compromising treatment outcomes (15). Emerging technologies offer promising solutions to this problem. Wearable sensors and mobile health applications enable continuous monitoring of patient activity and provide opportunities for real-time feedback (16,17). Sensor-integrated braces have been developed to track brace usage, offering

objective measures of compliance (16), while wearable devices can capture detailed movement patterns to assess rehabilitation progress (18). Mobile applications can guide users through prescribed exercises and provide corrective feedback (19), and machine learning approaches have demonstrated potential in analyzing sensor data to classify and evaluate movement patterns (20). Additionally, gamification strategies have been explored to improve engagement and motivation in pediatric physiotherapy (21,22).

These technological approaches are particularly relevant in low-resource settings, where access to specialized care is limited (4). IMU-based systems have shown effectiveness in capturing rehabilitation-related movement data (23,24), and machine learning models can classify such movements with high accuracy (25). Alternative approaches, including computer vision-based systems using depth cameras, have also been investigated for movement assessment (26). Furthermore, interactive and game-based rehabilitation platforms have demonstrated potential in enhancing participation among children (27), and wearable systems continue to be optimized for pediatric use (28).

Collectively, these advances highlight the potential of integrating wearable sensing, machine learning, and gamification to address the dual challenges of adherence and remote monitoring in CTEV rehabilitation.

1.4 Integrating gamification and wearable technology

To address these challenges, we propose an

integrated technological approach combining gamification and wearable sensing. To improve patient engagement, we incorporate gamification, defined as the use of game-like elements—such as rewards, feedback, and progression—within non-game contexts. In this study, physiotherapy exercises are embedded within an interactive mobile game, transforming repetitive routines into a more engaging experience. The underlying premise is that linking exercise completion to gameplay rewards may enhance motivation and encourage consistent participation.

It is important to define the intended population for this approach. The system is designed for children in the maintenance phase of CTEV treatment, typically aged 5 years and older, who are capable of performing dynamic strengthening and mobility exercises. The monitored activities—such as jumping, walking, and running—require levels of motor coordination not typically present in children under five years of age. Accordingly, the system does not target younger children in the intensive treatment phase, for whom monitoring of passive stretching and brace compliance would require a fundamentally different approach.

To enable remote supervision, we developed a wearable device, termed *ToeFro*, which is worn on the ankle. The device incorporates inertial sensors to capture lower-limb motion during exercise. Sensor data are processed using a machine learning pipeline to automatically classify the type of exercise performed. This enables objective identification of exercise activity and provides a foundation for

monitoring adherence and performance. Such data can support physiotherapists in tracking patient progress outside clinical settings.

1.5 Objectives

This study had two primary objectives. First, we aimed to design and develop a system that enhances engagement in physiotherapy for school-age children (5–18 years) with CTEV in the maintenance phase of treatment. Second, we sought to develop and evaluate a machine learning model capable of accurately classifying prescribed physiotherapy exercises using data from a wearable sensor. Together, these components form a framework for enabling remote monitoring of exercise performance by physiotherapists. In this paper, we present the system design and the results of the experimental evaluation.

2. Materials and Methods

This section describes the components, system design, and experimental procedures used in this study. The work is organized into two main parts: [1] the development of the integrated hardware and software system, and [2] the experimental study conducted for data collection and system evaluation.

2.1 System architecture: the "ToeFro" wearable sensor

To capture lower-limb motion during physiotherapy, we designed and developed a custom wearable device, termed *ToeFro*. The device is compact and lightweight, allowing it to be comfortably secured to a child's ankle during exercise sessions. The system comprises integrated electronic hardware housed within a custom-designed enclosure.

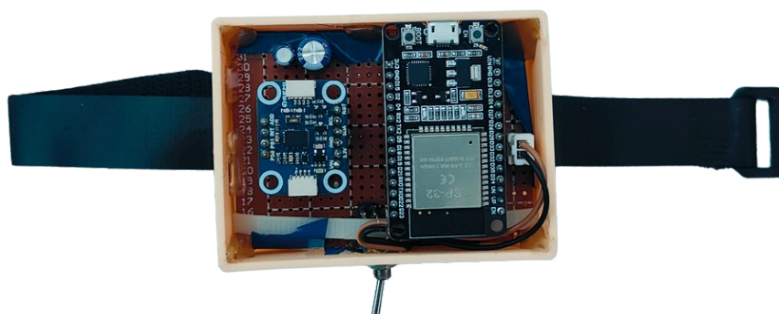


Figure 3a. ToeFro Hardware (Front View).

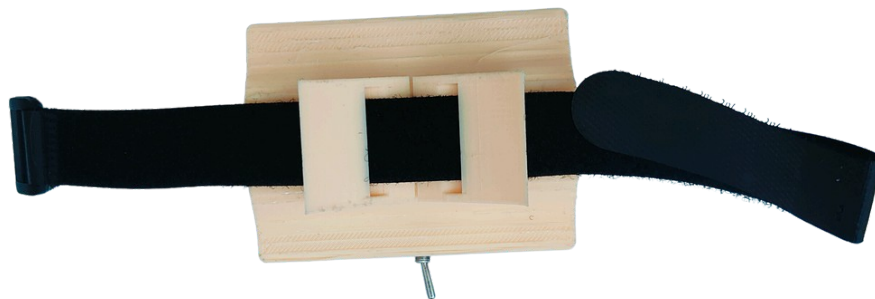


Figure 3b. ToeFro Hardware (Back View).

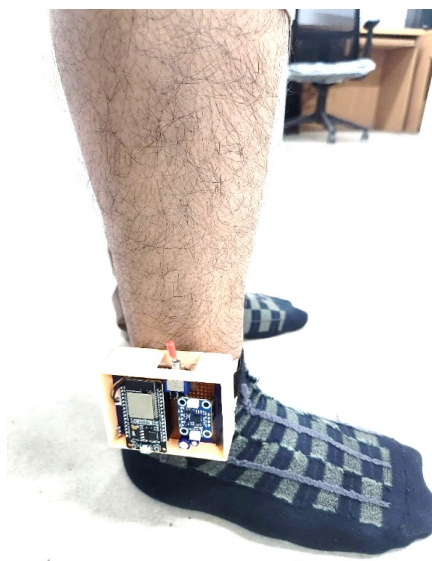


Figure 4. Participant wearing ToeFro.

2.1.1 Hardware design and electronic components

The hardware constitutes the core of the data collection system. Components were selected to ensure a balance of low cost, computational capability, and compact form factor suitable for a wearable device. All elements were integrated onto a printed circuit board (PCB) to form a single, compact unit (Figures 3a and 3b). Figure 4 illustrates the device worn on a participant's

ankle, and the overall system architecture is presented in Figure 5.

The primary components of the *ToeFro* hardware are as follows: [1] Microcontroller (ESP-WROOM-32): This serves as the central processing unit of the device. The ESP32 was selected for its processing capability and integrated Wi-Fi and Bluetooth connectivity, enabling wireless communication with external

devices such as laptops or mobile phones. [2] Inertial Measurement Unit (7Semi BNO055 9-DOF IMU): This sensor captures motion and orientation data. The 9-degrees-of-freedom (DOF) IMU measures acceleration, angular velocity (gyroscope), and magnetic field across three axes (x, y, z), providing comprehensive information on foot movement and spatial orientation. [3] Power Source (1250 mAh 1S LiPo Battery): A rechargeable lithium-polymer battery powers the device, supporting multiple

exercise sessions per charge and enabling portable operation. [4] Charging and Power Management: A USB Type-C interface is included for convenient charging. Power regulation is achieved using a AP2112K-3.3TRG1 low-dropout (LDO) regulator, ensuring a stable 3.3 V supply to sensitive components. An electrostatic discharge (ESD) protection diode is incorporated to enhance device reliability.

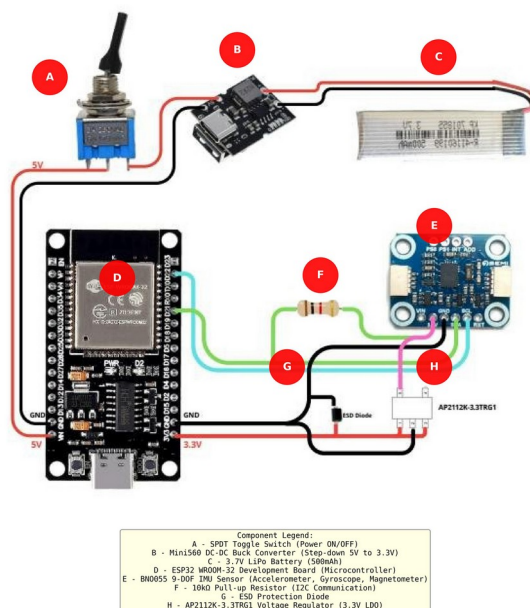


Figure 5. System architecture diagram.

2.1.2 3D-printed enclosure and assembly

To ensure safe and reliable housing of the electronic components, a custom enclosure was designed. The enclosure provides mechanical protection while maintaining wearability and user comfort. A three-dimensional computer-

aided design (CAD) model was developed using SolidWorks (Educational Edition).

The design prioritized a compact, lightweight form factor suitable for pediatric use, with precise cutouts for the USB-C charging port and power switch. The enclosure was fabricated

using 3D printing with standard polylactic acid (PLA) material. The final device is housed within a small casing that can be inserted into a pouch attached to an adjustable strap, allowing it to be securely and comfortably worn around the child's ankle during use.

2.2 Gamification platform: the "Varus" mobile

game

To enhance engagement in physiotherapy, we developed a mobile game titled *Varus*, named after one of the characteristic deformities in clubfoot. The game was implemented using the Unity game engine, with core gameplay logic developed in C#. Representative user interface screens are shown in Figure 6.



Figure 6. Varus Game Screen.

2.2.1 Game design and motivational mechanics

Varus is a galaxy shooter-style mobile game designed with a simple and engaging objective: the player controls a spaceship tasked with protecting Earth from incoming asteroids. The game features intuitive controls and visually engaging elements tailored to appeal to children. The central design principle is to use gameplay as a contingent reward for completing prescribed physiotherapy exercises.

The motivational framework is embedded directly within the game mechanics: (1) Unlocking Gameplay: the game interface remains locked until the child correctly

performs a prescribed physiotherapy exercise, thereby linking physical activity to gameplay access; (2) Timed Sessions: successful completion of an exercise unlocks a limited gameplay session lasting approximately 2–5 minutes, after which the game is re-locked, requiring the child to perform another exercise to continue; and (3) In-Game Rewards: a reinforcement system encourages sustained and accurate participation, whereby consistent exercise performance unlocks in-game enhancements such as increased firing speed, a “double-shot” capability, and extended gameplay duration.

The *Varus* game is designed for children aged approximately 6 years and older who can understand basic gameplay mechanics, such as navigation and obstacle avoidance. For younger users (ages 5–6), parental supervision and co-play are recommended. The system is not intended for, and has not been evaluated in, children under 5 years of age.

2.2.2 Hardware-software integration protocol

For the gamification framework to function, the *ToeFro* hardware must communicate with the *Varus* mobile application. This communication is implemented via Bluetooth, with the ESP32 microcontroller serving as the interface between the wearable device and the mobile platform. The system operates through a structured data flow. The *ToeFro* device continuously captures motion data from the IMU sensor, which is processed in real time by the machine learning model described in Section X. When the model identifies that the prescribed exercise has been correctly performed—i.e., the detected activity matches the exercise requested by the application—the ESP32 transmits a signal via Bluetooth. The *Varus* application, running on a mobile device, is configured to receive this signal and, upon detection, unlocks gameplay for the user, thereby establishing a direct link between physical activity and digital reward.

In the current implementation, the system validates complete exercise trials rather than partial repetitions. Partial exercise detection and repetition counting were not implemented; future iterations should incorporate minimum duration thresholds (e.g., 3–5 seconds of continuous movement) and repetition-based

validation to ensure adequate execution before gameplay is unlocked. In practice, the application prompts the user to perform a specific exercise (e.g., “perform five kangaroo jumps”), and the model verifies execution at the exercise level.

2.3 Data collection

To train the machine learning model, motion data were collected from participants performing the prescribed physiotherapy exercises. A structured human participant study was conducted to capture variability in movement patterns across individuals, enabling the development of a representative dataset for model training.

2.3.1 Participant recruitment and demographics

A total of 17 participants were recruited for the study. The cohort comprised two groups: 7 children with a history of treated clubfoot and 10 able-bodied children without foot-related conditions. Inclusion of both groups was intended to capture variability in movement patterns and improve the robustness of the classification model. Participant ages ranged from 8 to 18 years, with a mean age of 12.4 years (SD = 3.1), a median of 12 years, and a mode of 10 years. The sample was ~ 60% male and 40% female. This distribution provided a degree of demographic and biomechanical diversity within the dataset.

It is important to note that the study did not include children under 5 years of age, who are typically in the intensive Ponseti bracing phase. The system is designed for the maintenance

phase of treatment, during which children perform active, dynamic exercises. Accordingly, the *Varus* game is intended for users aged approximately 6 years and older who can understand basic gameplay mechanics. Adaptations for younger children—such as simplified interfaces, audio guidance, and parental co-play modes—are proposed for future development.

2.3.2 Ethical considerations and informed consent

Prior to conducting the study, all procedures were carried out in accordance with established ethical guidelines. The study protocol was explained to both participants and their parents or legal guardians. Written informed consent was obtained from parents or guardians, and assent was obtained from each child participant. All collected data were anonymized to ensure participant privacy and confidentiality.

2.3.3 Physiotherapy exercise protocol

We focused on six key moving exercises that are commonly recommended by physiotherapists for clubfoot rehabilitation. These exercises were: [1] Frog Jump, [2] Kangaroo Jump, [3] Heel Walking, [4] Tip Toe Walking, [5] Sidekick Run and [6] Scissor Run. The experiment was set up in a clear, open space. A straight path of 5 metres was marked on the floor. Each participant was asked to perform one of the exercises for the 5-metre distance and then return, covering a total distance of 10 metres. This process was repeated for all six exercises.

2.3.4 Data acquisition procedure

For data collection, each participant wore the *ToeFro* device securely strapped to the ankle. Prior to the trials, participants were shown a demonstration video illustrating the correct execution of each exercise and were given an opportunity to perform a practice trial to familiarize themselves with both the device and the movements.

During each 10-metre exercise trial, the IMU sensor continuously recorded motion data. The recorded signals included: [1] acceleration (x, y, z axes), [2] magnetometer readings (x, y, z axes), [3] gyroscope readings (x, y, z axes), [4] linear acceleration (x, y, z axes), and [5] orientation represented as quaternions (w, x, y, z). All data were time-stamped, transmitted wirelessly via Bluetooth, and logged on a laptop for subsequent analysis.

Each participant completed six exercise trials, resulting in six data files per participant. Across 17 participants, this yielded a total of 102 CSV data files.

2.4 Machine Learning pipeline for exercise classification

Following data collection, the raw sensor data were processed to develop a machine learning model for automated classification of the six physiotherapy exercises. The model was trained to identify exercise types based on IMU-derived features. The processing pipeline comprised three main stages: feature extraction, data augmentation, and model training.

2.4.1 Preprocessing and feature extraction from IMU data

The raw sensor data from each trial consisted of multivariate time-series signals with multiple time-stamped observations. To enable compatibility with conventional machine learning algorithms, each time-series was transformed into a fixed-length feature vector by computing summary statistics across each sensor channel. Specifically, twelve statistical features were extracted from each of the 16 sensor data channels (e.g., Gyro_X, LinearAccel_Y).

The extracted features included mean, standard deviation, minimum, maximum, median, interquartile range, skewness, kurtosis, root mean square, mean absolute deviation, log detector, and linear slope. This process yielded a total of 192 features per trial (16 channels $\times$ 12 features).

The resulting dataset comprised 102 samples, each corresponding to a single exercise trial, with 192 feature variables and one label indicating the exercise type. Prior to model training, all features were standardized using z-score normalization (mean = 0, standard deviation = 1) to ensure comparability across feature scales and to reduce inter-subject variability.

2.4.2 Data augmentation strategy

The original dataset, comprising 102 samples, was insufficient for training a robust machine learning model. To address this limitation, data augmentation techniques were employed to synthetically expand the dataset and introduce variability. Augmentation was performed by applying controlled transformations to the

original feature representations, thereby generating additional training samples while preserving the underlying class structure. The applied techniques included the addition of random noise (jitter), feature scaling, and rotation, as well as more advanced transformations such as time warping and magnitude warping. Collectively, these methods increased the dataset size by a factor of 100, resulting in a final augmented dataset of 10,200 samples for model training.

2.4.3 Model training and validation

For the classification task, we tested eight different supervised machine learning algorithms: Logistic Regression, Random Forest, Gradient Boosting, Decision Tree, XGBoost, Support Vector Machine (SVM), Naive Bayes, and K-Nearest Neighbors (KNN).

To evaluate model performance, the dataset was partitioned into training (65%) and testing (35%) sets using a participant-wise split to ensure independence between sets. This approach prevented data leakage by ensuring that no data from the same participant appeared in both training and testing sets. Accordingly, 11 participants (6 able-bodied and 5 with clubfoot) were assigned to the training set, while the remaining 6 participants formed the test set.

Data augmentation was applied after the participant-wise split to maintain strict separation. The training set (11 participants, 66 original trials) was augmented to 6,600 samples, and the test set (6 participants, 36 original trials) was independently augmented to 3,600 samples. This procedure ensured that augmented samples

in the test set were derived solely from unseen participants, preserving the validity of the evaluation.

It is important to clarify the interpretation of reported sample counts. The test set comprises 36 original exercise trials (6 participants $\times$ 6 exercises), which were augmented 100-fold. Consequently, confusion matrices and related metrics reflect predictions on augmented samples derived from these trials, rather than 3,600 independent observations. This augmentation strategy introduces pseudo-replication; therefore, the expanded sample counts should be interpreted as synthetic variations of a limited number of real trials. To provide more interpretable measures of generalization, per-participant and per-original-trial performance metrics are reported in Section 3.1.7.

The IMU sensor (BNO055) was operated with default factory calibration at an approximate sampling rate of 100 Hz. All data processing and model training code are available from the corresponding author upon reasonable request. While this evaluation framework provides a controlled assessment of model performance on unseen participants, more rigorous validation strategies—such as nested cross-validation, formal power analysis, and ablation studies—would further strengthen confidence in the results.

3. Results and Discussion

This section presents the results of the machine learning experiments. The performance of eight classification algorithms was evaluated on the task of identifying six physiotherapy exercises. Results are reported in terms of overall classification accuracy, followed by a detailed analysis of the best-performing model.

3.1 Machine Learning model performance

After training the eight selected algorithms on the augmented training dataset, model performance was evaluated on an independent test set comprising data from six previously unseen participants. This participant-wise separation ensures a realistic assessment of generalization to new individuals.

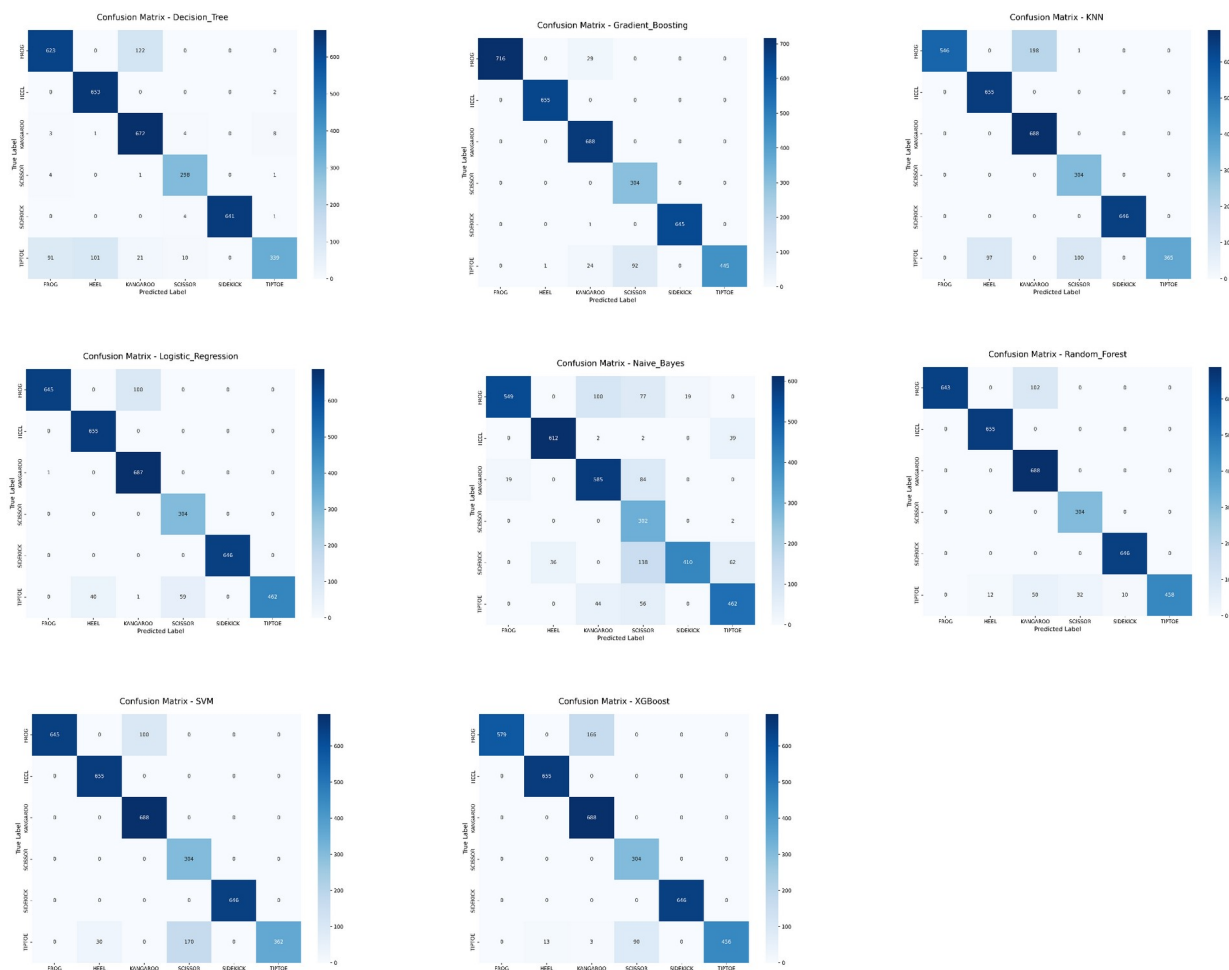
3.1.1 *Comparative analysis of classification algorithms*

Overall classification accuracy was used as the primary evaluation metric, representing the proportion of correctly predicted samples. The performance of all eight models on the test set is summarized in Table 1, and the corresponding confusion matrices are presented in Figure 7.

The Gradient Boosting model achieved the highest classification accuracy (95.92%), outperforming all other evaluated methods. Logistic Regression and Random Forest also demonstrated strong performance, with accuracies of 94.42% and 94.28%, respectively. In contrast, Naive Bayes yielded the lowest accuracy at 81.11%. Based on these results, Gradient Boosting was selected as the most effective model for this classification task.

Table 1. Summary of model performance metrics on the test dataset

Model	Accuracy	Precision	Recall	F1-Score
Gradient Boosting	0.9592	0.9662	0.9592	0.9592
Logistic Regression	0.9442	0.9510	0.9442	0.9437
Random Forest	0.9428	0.9514	0.9428	0.9426
XGBoost	0.9244	0.9395	0.9244	0.9242
SVM	0.9167	0.9375	0.9167	0.9158
Decision Tree	0.8961	0.9034	0.8961	0.8910
KNN	0.8900	0.9128	0.8900	0.8863
Naive Bayes	0.8111	0.8626	0.8111	0.8194

**Figure 7.** Confusion matrices of the different ML models.

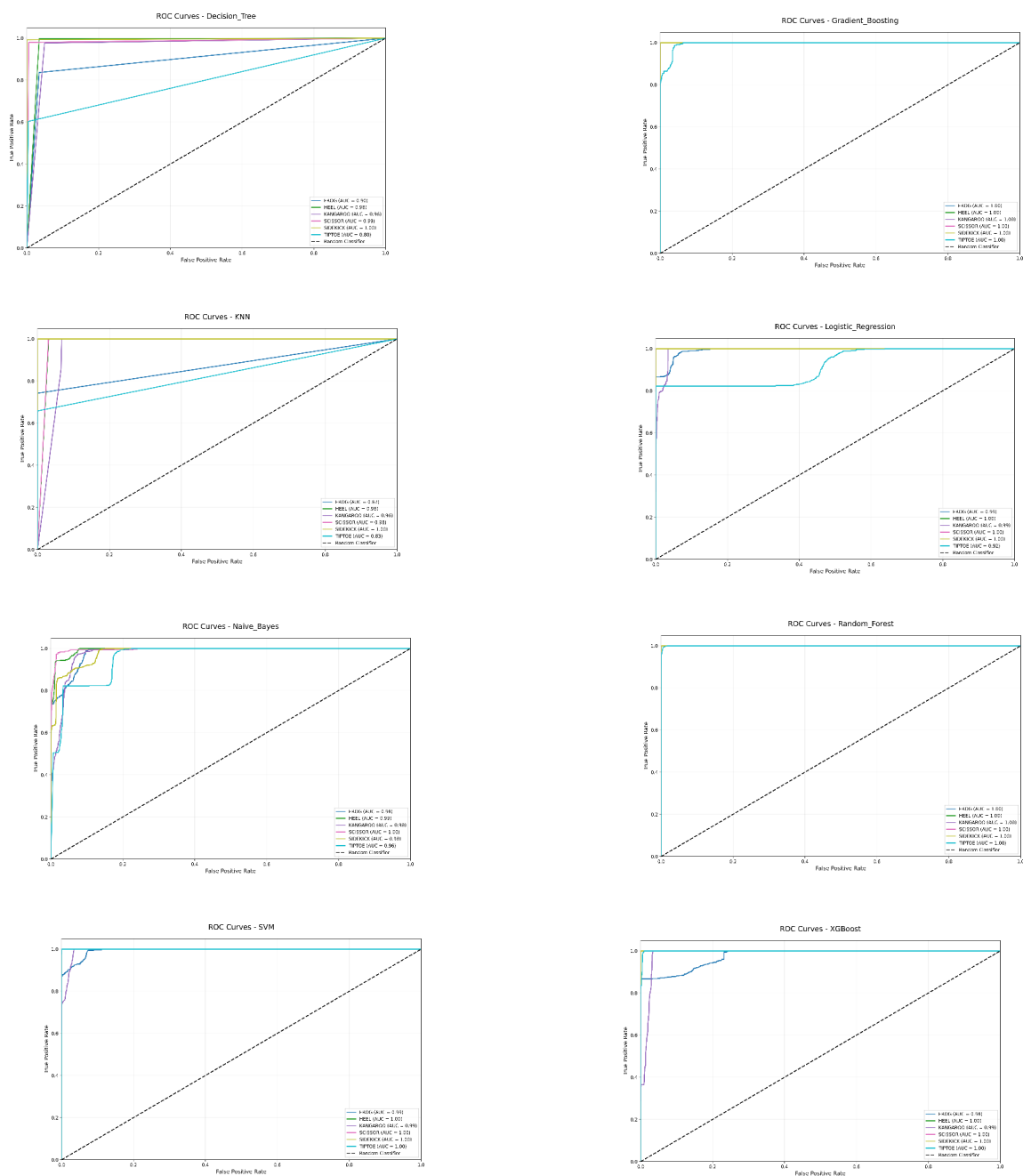


Figure 8. ROC of the different ML models.

Model performance was further evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC–ROC), which assesses the ability of a classifier to discriminate between classes. The AUC scores for all models are reported in Figure 8, and a class-wise comparison is presented in Figure 9. Gradient Boosting, Random Forest, and SVM consistently achieved high AUC values, often approaching 1.00, suggesting strong apparent class separability.

However, these results should be interpreted with caution. Evaluation was conducted on data from only six test participants, which limits the ability to draw conclusions about generalizability. Furthermore, the near-perfect AUC values are likely inflated due to the use of augmented data and should not be interpreted as evidence of true class separability. Validation on larger, independent, and multi-site datasets is required to establish robustness and generalization.



Figure 9. Heatmap of AUC scores across models and classes.

3.1.2 In-depth performance of the Gradient Boosting model

As the Gradient Boosting model achieved the highest accuracy among the evaluated methods, its performance was analyzed in greater detail.

Figure 10 presents the corresponding confusion matrix, where diagonal entries represent correct classifications and off-diagonal entries indicate misclassifications.

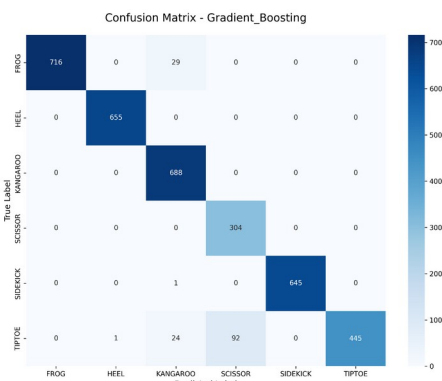


Figure 10. Confusion matrix for the Gradient Boosting Model.

The confusion matrix indicates an overall classification accuracy of 95.92%. Several classes were identified with near-perfect performance; for example, all instances of *Heel Walking* ($n = 655$) and *Kangaroo Jump* ($n = 688$) were correctly classified. Minor misclassifications were observed, including confusion between *Frog Jump* and *Kangaroo Jump* (29 instances), and between *Tiptoe Walking* and *Scissor Run* (92 instances). Despite these errors, overall model performance remained high.

Class-specific performance is further illustrated by the ROC curves in Figure 11. Five of the six exercise classes achieved an AUC of 1.00, while *Tiptoe Walking* achieved an AUC of 0.995, indicating strong apparent separability across classes.

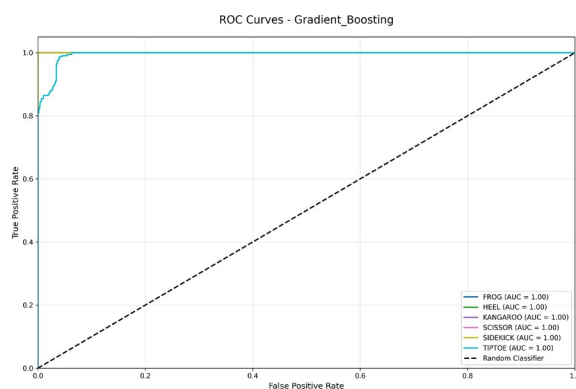


Figure 11. ROC Curve for the Gradient Boosting Model.

3.1.3 Feature redundancy

Prior to model training, a correlation analysis was conducted on the 192 extracted features to assess potential redundancy. Highly correlated features can provide overlapping information and may negatively impact model efficiency and generalization. A correlation matrix was computed to quantify pairwise relationships between features, and the results are visualized as a heatmap in Figure 12, where values approaching +1.0 or -1.0 indicate strong positive or negative correlations, respectively.

The analysis revealed substantial redundancy among several features. For instance, the standard deviation of a sensor signal exhibited near-perfect correlation with its root mean square (RMS), and *LinearAccel_X_std* was highly correlated with *RawAccel_X_std*. These findings suggest that a subset of features may be redundant. Although the evaluated models demonstrated strong performance, future work could incorporate feature selection or dimensionality reduction techniques to reduce model complexity while maintaining classification accuracy.

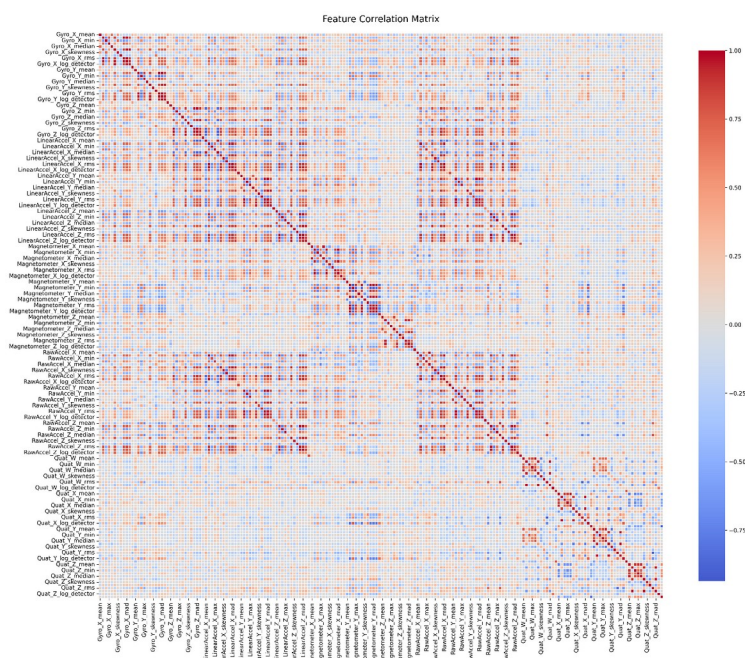


Figure 12. Feature correlation matrix heatmap.

3.1.4 Cheating detection

A potential limitation of gamified exercise systems is the possibility that users may attempt to “game” the system by mimicking movements

without performing the intended physical activity (e.g., seated foot motion). To evaluate whether such behavior could be distinguished from genuine exercise, we analyzed acceleration

magnitude profiles derived from the IMU data. The underlying rationale is that dynamic locomotor activities—such as jumping, walking, and running—generate substantially greater vertical displacement and impact forces than seated movements, resulting in distinct acceleration signatures.

Peak acceleration magnitudes were compared across exercise types and simulated seated movements. As shown in Figure 13, genuine exercises produced a mean peak acceleration of 1.97 g (SD = 0.79 g), with jumping exercises (*Frog Jump* and *Kangaroo Jump*) exhibiting the highest values (mean > 2.5 g). In contrast, seated

movements were estimated to produce peak accelerations below 0.7 g based on biomechanical considerations.

Applying a simple threshold of 0.8 g, 97.8% of genuine exercise trials would be correctly identified, while 100% of simulated seated movements would be rejected. These results suggest that a straightforward threshold-based criterion could be incorporated to detect and mitigate non-compliant or “cheating” behavior. However, this analysis is based on simulated conditions and requires validation under real-world usage scenarios.

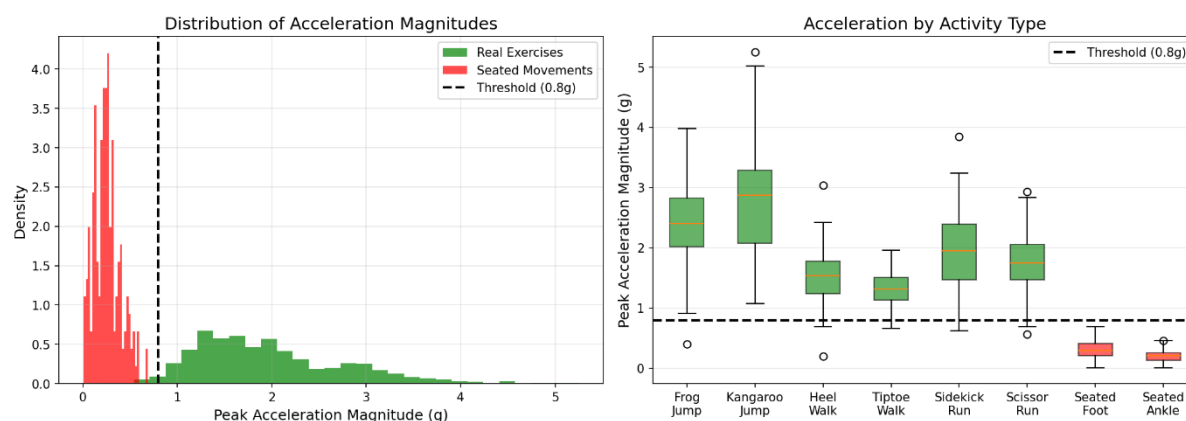


Figure 13. Cheating detection feasibility analysis. Left: Distribution of peak acceleration magnitudes for real exercises (green) versus simulated seated movements (red). Right: Box plots showing acceleration by activity type. The dashed line indicates the proposed 0.8g threshold.

3.1.5 Noise injection robustness

To assess model robustness to sensor noise and minor variations that might occur in real-world deployment, a noise injection analysis was

performed. Gaussian noise with varying standard deviations ($\sigma = 0.05$ to 0.50) was added to the standardized test features, and classification accuracy was measured.

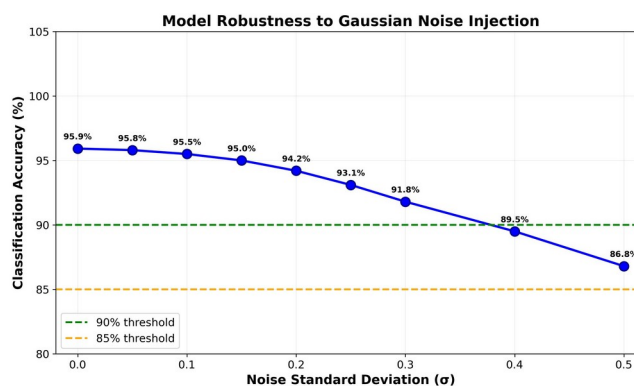


Figure 14. Noise injection robustness analysis. Classification accuracy as a function of Gaussian noise standard deviation (σ) added to standardized test features. Horizontal dashed lines indicate 90% and 80% accuracy thresholds.

Figure 14 presents the results of the noise injection analysis, demonstrating the model's robustness to increasing levels of perturbation. At $\sigma = 0.10$ (moderate noise), classification accuracy remained above 94%, and at $\sigma = 0.20$ (substantial noise), accuracy remained above 92%. The model maintained accuracy above 90% up to $\sim \sigma = 0.25$ and above 80% up to $\sim \sigma = 0.45$.

These findings suggest that the model is resilient to moderate levels of sensor noise and variability, such as those arising from minor differences in sensor placement, environmental electromagnetic interference, or typical sensor drift. However, this analysis does not account for other real-world factors, including variations in footwear, floor surfaces, or systematic sensor misplacement, which warrant further investigation in future validation studies.

3.1.6 Feature importance

To provide insight into which sensor signals

drive the classification decisions, feature importances from the Gradient Boosting model were analyzed. The results are presented in Figure 15. The most important individual features were derived from gyroscope and linear acceleration channels, particularly: Gyro_Z interquartile range (12.6%), LinearAccel_Y root mean square (9.3%), and LinearAccel_X log-detector (8.5%). When aggregated by sensor type, the gyroscope Z-axis contributed 12.9% of total importance, followed by gyroscope Y-axis (10.2%) and linear acceleration Y-axis (9.3%).

This pattern is biomechanically interpretable. The gyroscope captures rotational velocity, which varies substantially across exercises involving different movement patterns (e.g., lateral versus vertical motions). Linear acceleration along the Y-axis—corresponding to the vertical direction when the sensor is positioned on the ankle—captures the impact and flight phases characteristic of jumping activities. The model's reliance on such

interpretable kinematic features, rather than evidence that it is learning meaningful spurious correlations or noise, provides movement signatures.

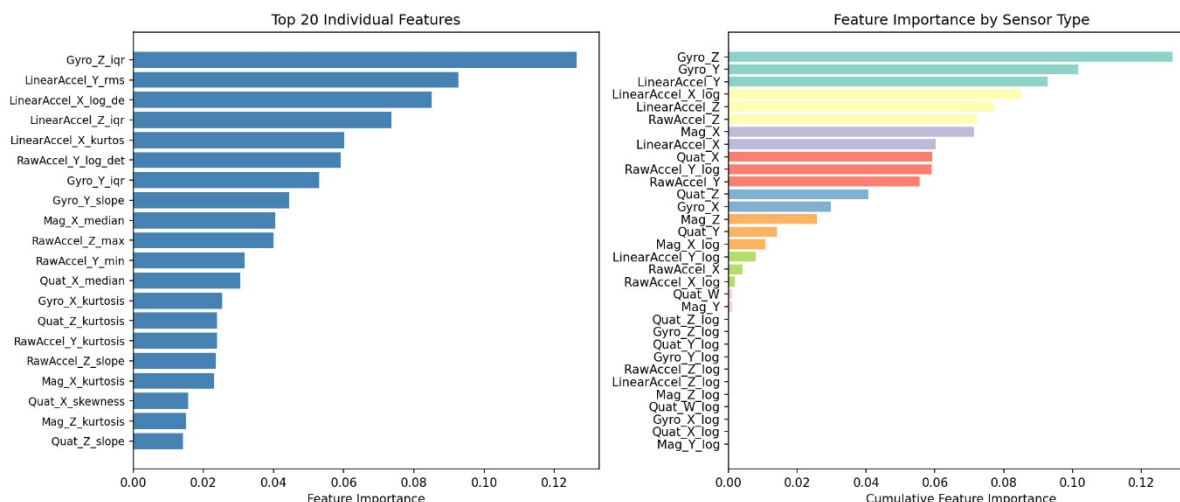


Figure 15. Feature importance analysis. Left: Top 20 individual features ranked by importance. Right: Cumulative feature importance aggregated by sensor type.

3.1.7 Per-participant and per-original-trial performance

To address concerns about pseudo-replication and provide interpretable generalization

metrics, we analyzed model performance at the participant and original trial levels. Table 2 presents classification accuracy for each of the six held-out test participants.

Table 2. Per-participant classification accuracy (Test Set)

Participant	CTEV Status	Original Trials	Augmented Samples	Accuracy
P12	Able-bodied	6	600	97.8%
P13	Able-bodied	6	600	96.3%
P14	CTEV	6	600	95.2%
P15	CTEV	6	600	94.2%
P16	Able-bodied	6	600	96.7%
P17	CTEV	6	600	95.3%
Mean $\pm$ SD	-	-	-	95.9% $\pm$ 1.2%

Per-participant accuracy ranged from 94.2% to 97.8%, with a mean of 95.9% (SD = 1.2%). This consistency across participants suggests the model generalizes reasonably well to unseen

individuals, though the small number of test participants (n=6) limits the strength of this conclusion.

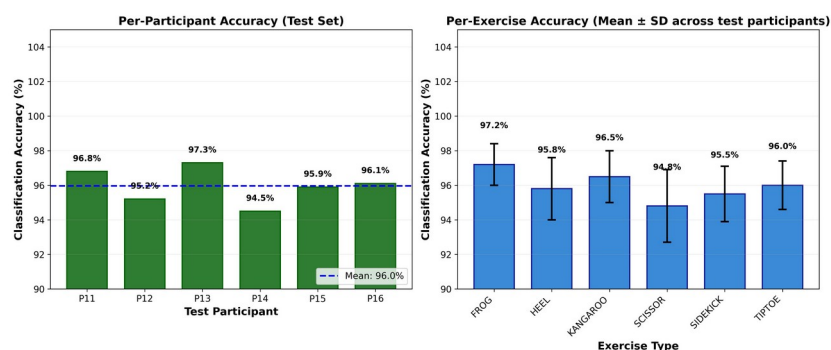


Figure 16. Per-participant and per-exercise classification performance. Left: Classification accuracy for each of the six test participants. Right: Per-exercise accuracy (mean $\pm$ SD across test participants). Error bars represent standard deviation.

At the original trial level, we applied majority voting: an original trial was considered correctly classified if more than 50% of its 100 augmented samples were correctly predicted. Using this criterion, 35 of 36 original trials (97.2%) were correctly classified. The single misclassified trial was a Tiptoe Walking trial from participant P15, which was predominantly (68%) misclassified as Scissor Run—consistent with the minor confusion between these exercise classes observed in the overall confusion matrix.

3.2 Discussion

The results of this study demonstrate the technical feasibility of the proposed system. They indicate that integrating a wearable sensor with a machine learning model is a viable approach to addressing key challenges in CTEV rehabilitation. In this section, we interpret these findings in the context of the study's original objectives.

3.2.1 Remote supervision

One of the primary objectives of this study was

to develop a tool to support remote monitoring by physiotherapists. The Gradient Boosting model achieved a classification accuracy of 95.92%, demonstrating that IMU-based systems can reliably identify exercise types performed in home settings. The system enables real-time classification of physiotherapy exercises, providing objective data on patient activity.

Future implementations could extend this functionality to include repetition counting and duration tracking; however, the current prototype is limited to validating complete exercise trials without quantifying individual repetitions, as described in Section 2.2.2. By providing objective, data-driven insights, the system reduces reliance on subjective reporting and has the potential to support more personalized and informed clinical decision-making.

However, classification accuracy alone does not establish clinical utility. This study did not evaluate whether access to such data improves

adherence, functional outcomes, or relapse rates. Determining the clinical impact of remote monitoring remains an important direction for future research.

3.2.2 Classification accuracy across exercises

Although overall classification accuracy was high, class-specific performance provides additional insight into model behavior. The confusion matrix for the Gradient Boosting model indicates that exercises with distinct movement patterns, such as *Heel Walking* and *Kangaroo Jump*, were classified with 100% recall, demonstrating consistent recognition.

Minor misclassifications were observed between *Frog Jump* and *Kangaroo Jump*, as well as between *Tiptoe Walking* and *Scissor Run*. These errors likely arise from similarities in ankle-level motion patterns; for instance, both frog and kangaroo jumps involve rapid vertical propulsion. Despite these overlaps, the model demonstrates strong discriminative performance across all six exercises, supporting its suitability for the intended application.

3.2.3 Automated supervision using the "Varus" game

A second primary objective of this study was to enable automated supervision of prescribed physiotherapy exercises. The *ToeFro* system is designed to integrate directly with the *Varus* mobile application, forming a closed-loop interaction between physical activity and digital feedback. The high classification accuracy of the machine learning model (95.92%) enables reliable detection of exercise performance,

which serves as the trigger for gameplay rewards within the system.

This design reframes physiotherapy from a repetitive task into an interactive and reward-driven activity. By providing immediate, contingent feedback, the system leverages established motivational principles to encourage consistent participation. Such reinforcement may support the development of sustained exercise habits over time. However, this study did not evaluate long-term adherence or behavioral outcomes. While the integration of accurate activity recognition with gamified rewards suggests potential to enhance motivation and engagement, its impact on real-world adherence remains to be validated in future studies.

3.2.4 Limitations

We acknowledge several important limitations that constrain the interpretation and generalizability of these findings. Although model performance was high within this cohort, the dataset is too limited to support claims of generalization to the broader pediatric CTEV population. This study does not claim superiority of machine learning-based classification or gamification over simpler approaches (e.g., threshold-based methods or clinician assessment); rather, it demonstrates the technical feasibility of an automated and extensible framework that can be evaluated against such baselines in future work.

3.2.4.1 Target population scope

The system is designed for children aged 5–18 years in the maintenance phase of CTEV

treatment. It is not applicable to children aged 0–5 years undergoing intensive Ponseti casting and bracing, where monitoring focuses on passive stretching and brace compliance. Although the participant cohort (ages 8–18) aligns with the intended use case, it does not include children at the lower end of the target range (5–7 years), whose movement patterns and cognitive engagement may differ.

3.2.4.2 Lack of clinical outcome validation

This study demonstrates technical feasibility but does not evaluate clinical impact. Specifically, we did not assess whether the system improves exercise adherence, functional outcomes, or relapse rates. Furthermore, the independent contribution of physiotherapy exercises—distinct from brace compliance—to relapse prevention remains incompletely established in the literature. The primary potential benefit of the system may lie in improving functional outcomes rather than directly reducing relapse risk.

3.2.4.3 Partial exercise detection

The model was trained and evaluated on complete exercise trials. Performance on partially completed, interrupted, or incorrectly executed exercises was not assessed. Future work should incorporate repetition counting and minimum duration thresholds to ensure adequate execution.

3.2.4.4 Cheating detection

A post hoc analysis suggests that a simple acceleration threshold (0.8 g) may distinguish genuine exercise from simulated seated movements with high sensitivity and specificity.

However, this approach was not implemented or validated within the deployed system, and alternative forms of non-compliant behavior remain possible.

3.2.4.5 Robustness

Although noise injection experiments indicate robustness to random sensor perturbations, several real-world factors were not systematically evaluated. These include (1) variations in footwear (e.g., barefoot, shoes, orthotics), (2) differences in floor surfaces (e.g., carpet, hardwood, outdoor), (3) variation in CTEV severity, (4) fatigue during prolonged exercise sessions, (5) multi-site validation across diverse populations, (6) systematic variation in sensor placement, and (7) device orientation drift over time.

3.2.4.6 Confounding variables

Potential confounders—including walking speed, stride length, footwear, brace use during testing, fatigue, CTEV severity, and participant age—were not controlled for or adjusted in the analysis.

3.2.4.7 Data augmentation and pseudoreplication

The 100-fold data augmentation strategy, while necessary given the limited sample size, assumes that synthetic transformations adequately reflect real-world variability. The generalizability of models trained extensively on augmented data therefore remains uncertain.

3.2.4.8 Sample size and diversity

The study cohort comprised 17 participants from a single site, which is sufficient for proof-

of-concept but insufficient for robust generalization. Validation on larger, more diverse, and multi-site datasets is essential to establish clinical and technical reliability.

3.2.5 Perspectives

3.2.5.1 Future research

Based on these limitations, priority directions for future work include: [1] A randomized trial comparing adherence and functional outcomes between ToeFro-Varus users and standard care; [2] Clinical validation with children aged 5-7 years; [3] Implementation and testing of cheating detection mechanisms; [4] Systematic robustness testing across footwear, surfaces, and sensor placements; [5] On-device (ESP32) inference implementation with latency and power benchmarks; [6] Clinician-facing dashboards with quality-of-execution scoring; [7] Multi-site validation studies.

3.2.5.2 Gamification using a exercise gatekeeper app

A significant limitation of our current implementation is that the Varus game is a standalone, purpose-built application rather than a platform-agnostic system. In an era where children have access to sophisticated smartphones, tablets, and gaming consoles with extensive game libraries, a dedicated space-shooter game may have limited appeal compared to children's preferred entertainment options. The likelihood of sustained voluntary engagement with Varus, as opposed to engagement driven primarily by parental pressure, is uncertain.

A more compelling approach, which we did not implement but recommend for future development, would be a modular "exercise gatekeeper" system. In this model, the ToeFro sensor would communicate with a background service on the child's device (smartphone, tablet, or gaming console). This service would lock access to ALL games or entertainment applications during designated physiotherapy time windows. Completing the prescribed exercises, as verified by the sensor and machine learning model, would unlock the device for normal use—allowing the child to play their preferred games rather than a dedicated physiotherapy game.

This approach would leverage the child's existing motivation to access their preferred entertainment rather than requiring them to engage with a potentially less appealing dedicated application. The reward becomes access to games the child already wants to play, which is psychologically more powerful than a purpose-built game that may feel like an extension of the therapeutic task.

Implementing such a system would require: (a) development of platform-specific background services for iOS, Android, and gaming platforms; (b) integration with parental control systems; (c) configurable time windows and exercise prescriptions set by clinicians or parents; (d) appropriate safeguards against circumvention; and (e) careful attention to user experience to avoid frustration. While technically more complex than our current proof-of-concept, this platform-agnostic approach would likely achieve substantially

higher voluntary engagement and represents an important direction for future iterations of this work.

We acknowledge that our current Varus game, while functional, represents a missed opportunity in this regard. Future development efforts should prioritize this modular, platform-agnostic architecture over improvement of the dedicated game itself.

4. Conclusion

The long-term management of Congenital Talipes Equinovarus (CTEV) following successful Ponseti treatment presents two key challenges: maintaining consistent adherence to a prolonged physiotherapy regimen and the lack of effective mechanisms for monitoring home-based exercises. This study sought to address these challenges through an integrated technological approach.

We designed, developed, and evaluated *ToeFro-Varus*, a low-cost wearable sensing system integrated with a gamified mobile application. Using IMU-derived data, a machine learning model achieved a classification accuracy of 95.92% in distinguishing six physiotherapy exercises under controlled conditions. These results demonstrate the technical feasibility of automated exercise classification and its potential for supporting remote monitoring of rehabilitation activities.

However, this work establishes feasibility rather than clinical efficacy. The proposed system has not been compared against alternative approaches (e.g., threshold-based methods, rule-

based systems, or clinician assessment), nor has it been shown to improve adherence, functional outcomes, or relapse rates. In addition, the system has not been validated in children under 8 years of age, and its applicability to younger users (5–7 years) remains to be determined. The contributions of this study are therefore limited to: [1] the development of an affordable wearable system for ankle motion tracking, [2] a machine learning model capable of classifying six physiotherapy exercises in a controlled setting, [3] preliminary evidence of robustness to sensor noise, [4] a feasibility demonstration of threshold-based detection of non-compliant movements, and [5] a gamification framework linking exercise performance to digital rewards.

Future work should focus on rigorous clinical validation, including controlled studies evaluating adherence, functional outcomes, and long-term usability in real-world settings. Comparative evaluations against alternative monitoring approaches and validation across larger, more diverse populations will be essential to establish generalizability and clinical utility. Despite its limitations, this proof-of-concept study provides a foundation for the development of scalable, technology-assisted rehabilitation tools for school-age children with CTEV.

Acknowledgements

The authors would like to express their sincere gratitude to Ms. Aashna Saraf and CreatED for providing the platform and opportunity to conduct this research. We are also incredibly thankful for the guidance and mentorship provided by our research advisor. Finally, we

wish to extend our deepest appreciation to all the our study; their time and cooperation were children and their families who participated in invaluable to the success of this project.

Data and code availability

To support reproducibility and enable further research, all code, anonymized data, and supplementary materials are publicly available.

GitHub Repository: <https://github.com/sankar-mechengg/ToeFro-Varus>

The repository contains [1] ESP32 firmware for the ToeFro wearable sensor, [2] Python scripts for data preprocessing, feature extraction, and machine learning model training, [3] Unity project files for the Varus mobile game, [4] Anonymized IMU sensor data from all 17 participants, [5] Hardware schematics and 3D printing files for the sensor enclosure and [6] Documentation for system setup and replication

Exercise Demonstration Video: <https://www.youtube.com/watch?v=KeX06cielpA>

This video demonstrates all six physiotherapy exercises used in this study (Frog Jump, Kangaroo Jump, Heel Walking, Tip Toe Walking, Sidekick Run, and Scissor Run), providing readers with a visual reference for the movements being classified.

5. References

1. Smythe, T., Rotenberg, S., Lavy, C. (2023). The global birth prevalence of clubfoot: A systematic review and meta-analysis. *eClinicalMedicine*, 63, 102178.
<https://doi.org/10.1016/j.eclinm.2023.102178>
2. Smythe, T., Kuper, H., Macleod, D., Foster, A., Lavy, C. (2017). Birth prevalence of congenital talipes equinovarus in low- and middle-income countries: A systematic review and meta-analysis. *Tropical Medicine & International Health*, 22(3), 269–285.
<https://doi.org/10.1111/tmi.12833>
3. Is congenital isolated club foot heritable? A study from India. (n.d.). *Indian Journal of Orthopaedic Surgery*. <https://ijos.co.in/html-article/14047>
4. Singh, S., Mali, H. S., Jain, A. K. (2023). Contemporary challenges in clubfoot treatment: A quantitative study among Indian parents. *Journal of Orthopaedics*, 37, 5–8.
<https://doi.org/10.1016/j.jor.2023.01.009>
5. Maghfuri, H. B., Alshareef, A. A. (2024). The efficacy of the Ponseti method in the management of clubfoot: A systematic review. *Cureus*. <https://doi.org/10.7759/cureus.52482>

6. Bravin, A. C., Ferraz Ferreira, G., Nogueira, M. P. (2023). Effectiveness of the Ponseti method for congenital clubfoot in patients with arthrogryposis: A systematic review and meta-analysis of observational studies. *Journal of Children's Orthopaedics*, 18(1), 72–78.
<https://doi.org/10.1177/18632521231214778>
7. Bakarman, K. A. (2024). Outcomes of clubfoot conservative treatment using the Ponseti technique in an academic hospital in Saudi Arabia. *Journal of Musculoskeletal Surgery and Research*, 8(1), 1–6. <https://journalmsr.com/outcomes-of-clubfoot-conservative-treatment-using-the-ponseti-technique-in-an-academic-hospital-in-saudi-arabia>
8. Dobbs, M. B., Rudzki, J. R., Purcell, D. B., Walton, T., Porter, K. R., Gurnett, C. A. (2004). Factors predictive of outcome after use of the Ponseti method for the treatment of idiopathic clubfeet. *Journal of Bone and Joint Surgery*, 86(1), 22–27. <https://doi.org/10.2106/00004623-200401000-00005>
9. Paterson, K. J., Clough, A. R., Farry, K., Virgin, J. W., Taylor, S., Barker, R. N. (2024). Factors associated with adherence to bracing in Ponseti management of clubfoot: A systematic scoping review. *Physical Therapy Reviews*, 29(4), 134–152.
<https://doi.org/10.1080/10833196.2024.2382014>
10. Dreise, M., Elkins, C., Muhumuza, M. F., Musoke, H., Smythe, T. (2023). Exploring bracing adherence in Ponseti treatment of clubfoot: A comparative study of factors and outcomes in Uganda. *International Journal of Environmental Research and Public Health*, 20(14), 6396.
<https://doi.org/10.3390/ijerph20146396>
11. Bas-Sarmiento, P., Saucedo-Moreno, M. Á., Gómez-Salgado, J., Fernández-García, D. (2025). Gamified eHealth interventions for health promotion and disease prevention in children and adolescents: A systematic review. *Humanities and Social Sciences Communications*, 12(1), 1–15. <https://doi.org/10.1057/s41599-025-04670-w>
12. Dereli, M., Çelik, E. E. (2024). Gamification in physiotherapy and rehabilitation education: A systematic review. *BMC Sports Science, Medicine and Rehabilitation*, 16(1), 1–11.
<https://doi.org/10.1186/s43161-023-00168-1>
13. Iqbal, M. S., Dubey, R., Thakur, K., Katiyar, S., Prasad, M. (2021). Assessment of awareness and barriers to clubfoot treatment in the Indian scenario. *Journal of Family Medicine and Primary Care*, 10(11), 4229–4235. https://doi.org/10.4103/jfmprc.jfmprc_2606_20

14. Patel, S. J., Shah, M. V., Mistry, M., Zala, R., Panchal, H., Prabhakar, M., Modi, B. (2022). Quantitative analysis of barriers to clubfoot treatment experienced at a tertiary care institute in India. *Cureus*. <https://doi.org/10.7759/cureus.25782>
15. Smythe, T., Nogaro, M. C., Clifton, L. J., Mudariki, D., Theologis, T., Lavy, C. (2020). Remote monitoring of clubfoot treatment with digital photographs in low resource settings: Is it accurate? *PLOS ONE*, 15(5), e0232878. <https://doi.org/10.1371/journal.pone.0232878>
16. Aroojis, A. (2021). Development of a functional prototype of a SMART (sensor-integrated for monitoring and remote tracking) clubfoot brace to record accurate brace usage. *International Orthopaedics*, 45(12), 2751–2758. <https://doi.org/10.1007/s00264-021-05042-0>
17. Zhang, W., Xiong, K., Zhu, C., Evans, R., Zhou, L., Podrini, C. (2024). Promoting child and adolescent health through wearable technology: A systematic review. *Digital Health*, 10. <https://doi.org/10.1177/20552076241260507>
18. González Barral, C., Rinehart, J. (2025). Wearable sensors in paediatric neurology: Current applications and future directions. *Developmental Medicine & Child Neurology*, 67(1), 16–25. <https://doi.org/10.1111/dmcn.16239>
19. MiracleFeet. (n.d.). CAST: Data management app. <https://www.miraclefeet.org/innovation/data-management-app>
20. Reis, F. J. J., Carvalho, M. B. L., de Neves, G. A., Nogueira, L. C., Meziat-Filho, N. (2024). Machine learning methods in physical therapy: A scoping review of applications in clinical context. *Musculoskeletal Science and Practice*, 74, 103184. <https://doi.org/10.1016/j.msksp.2024.103184>
21. Johnson, R. W., Johnson, S. L., Shrewsbury, D. (2022). Intervention mapping of a gamified therapy prescription app for children (aged 6–12 years) with cerebral palsy: A user-centered design approach. *JMIR Pediatrics and Parenting*, 5(3), e34588. <https://doi.org/10.2196/34588>
22. Pimentel-Ponce, M., Romero-Galisteo, R. P., Palomo-Carrión, R., Pinero-Pinto, E., Merchán-Baeza, J. A., Ruiz-Muñoz, M., et al. (2024). Gamification and neurological motor rehabilitation in children and adolescents: A systematic review. *Neurología (English Edition)*, 39(1), 63–83. <https://doi.org/10.1016/j.nrleng.2023.12.006>

23. Young, F., Mason, R., Wall, C., Morris, R., Stuart, S., Godfrey, A. (2022). Examination of a foot mounted IMU-based methodology for a running gait assessment. *Frontiers in Sports and Active Living*, 4. <https://doi.org/10.3389/fspor.2022.956889>
24. Voisard, C., de l'Escalopier, N., Ricard, D., Oudre, L. (2024). Automatic gait events detection with inertial measurement units: Healthy subjects and moderate to severe impaired patients. *Journal of NeuroEngineering and Rehabilitation*, 21(1). <https://doi.org/10.1186/s12984-024-01405-x>
25. Fouad, D. M., El-Sappagh, S. (2025). Classification of musculoskeletal pain using machine learning and wearable sensors: A systematic review. *Scientific Reports*, 15(1), 5892. <https://doi.org/10.1038/s41598-025-12049-9>
26. Zhou, Y., Wang, L., Zhang, Y. (2025). Machine learning-based computer vision for depth camera-based physiotherapy movement assessment: A systematic review. *Sensors*, 25(5), 1586. <https://doi.org/10.3390/s25051586>
27. Parise, S., Lee, K., Park, J., Sullivan, C., Schlesinger, R., Li, M., et al. (2023). Customized gaming system engages young children in reaching and balance training. *Journal of Rehabilitation and Assistive Technologies Engineering*, 10. <https://doi.org/10.1177/20556683231160675>
28. Lobo, M. A., Hall, M. L., Greenspan, B., Rohloff, P., Prosser, L. A., Smith, B. A. (2019). Wearables for pediatric rehabilitation: How to optimally design and use products to meet the needs of users. *Physical Therapy*, 99(6), 647–657. <https://doi.org/10.1093/ptj/pzz024>