Proceedings of the 2025 SICE Festival with Annual Conference
September 9-12, 2025, Chiang Mai, Thailand

Histogram-based Gradient Boosting based Techniques for Predicting the Knee Abduction Moment

Palawich Giraruchataporn[1†,2†], Inês Pinto[1†,3†], Kittipong Ekkachai[4], and Waree Kongprawechnon[1]

[1]School of ICT, Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani, Thailand
(E-mail: palawich.gir@gmail.com[1†], waree@siit.tu.ac.th[1])
[2]School of Information Science, Japan Advance Institute of Science and Technology, Ishikawa, Japan
(E-mail: palawich@jaist.ac.jp)
[3]Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal
(E-mail: inescfpinto@tecnico.ulisboa.pt)
[4]SMR Research Team, National Electronics and Computer Technology Center (NECTEC), Thailand
(E-mail: kittipong.ekkachai@nectec.or.th)

Abstract: The knee abduction moment (KAM) is a critical biomechanical metric linked to medial knee osteoarthritis (OA) and Anterior Cruciate Ligament (ACL) injury risk. Conventional methods for measuring KAM rely on laboratory-based 3D motion capture systems and force plates, which are accurate but impractical for real-world applications due to their bulk and cost. To address this limitation, we propose a portable KAM prediction system using a smart shoe embedded with six force-sensitive resistors (FSRs) and an inertial measurement unit (IMU), combined with a machine learning model. Sensor data from the wearable (shoe), including force and motion signals at 100 Hz, were collected in parallel with ground-truth KAM data from a 3D tracking system. Two datasets from two person subjects were used, totaling 97 gait cycles. Missing values in the sensor data were addressed using both linear interpolation and K-Nearest Neighbor(KNN) imputation. A Histogram-Based Gradient Boosting Regressor (HGBR) was trained to predict KAM, and model performance was evaluated using R^2, RMSE, and percentage error. Results show that individual-specific models with linear interpolation achieved the highest accuracy ($R^2 = 0.986$), while the generalized model offered good performance across subjects. In contrast, K-Nearest Neighbors imputation significantly reduced accuracy. This research confirms the viability of wearable sensors combined with machine learning for reliable KAM monitoring, providing practical and scalable solutions for injury prevention and rehabilitation in non-laboratory environments.

Keywords: Wearable Sensor, Knee Abduction Moment, Sensor Fusion, Portable Health Monitoring, Human Gait Analysis.

1. INTRODUCTION

The Knee Abduction Moment (KAM) quantifies the external torque on the knee joint in the frontal plane during dynamic activities like walking or stair climbing. Elevated KAM is strongly linked to medial knee osteoarthritis (OA) and the risk of anterior cruciate ligament (ACL) injuries [1-3]. Accurate KAM estimation enables clinicians to assess joint health, predict injury risk, and tailor rehabilitation strategies.

KAM is particularly relevant for two high-risk groups: athletes, who experience repetitive high-impact loads, and older adults, who are prone to degenerative joint changes. While conventional measurement relies on 3D motion capture systems and force plates, these lab-based tools are costly, bulky, and unsuitable for continuous or real-world use [6, 7].

Recent developments in wearable sensors offer practical alternatives. For example, instrumented insoles with force sensors paired with neural networks have predicted KAM with high accuracy (correlation $\sim$0.96, error $<$1%) [4]. Similarly, LSTM-based models using only inertial measurement units (IMUs) have yielded clinically useful predictions outside motion labs [5], demonstrating the feasibility of portable gait analysis systems.

Machine learning (ML) methods have emerged as effective tools for KAM estimation using wearable sensor data such as FSRs and IMUs. Neural networks capture complex nonlinear relationships between sensor inputs and joint kinetics. Prior studies have shown convolutional networks trained on IMU data achieving low normalized RMSE (0.05–0.07 Nm/kg) across diverse gait tasks [6]. Others demonstrated that combining IMUs with pressure-sensitive insoles allows real-time 3D KAM estimation during running with strong agreement to lab measurements (ICC up to 0.90) [8]. These approaches eliminate the need for force plates and complex biomechanical models [6, 8, 9].

Such data-driven systems support real-time monitoring, scale well for widespread use, and can generalize across users, movement styles, and environments. As wearable hardware becomes more advanced and accessible, ML-based models pave the way for continuous, personalized assessment of joint loading.

In this work, we propose a machine learning-based approach using a Histogram-Based Gradient Boosting Regressor (HGBR) to estimate KAM from wearable sensor data collected via a smart shoe embedded with six force-sensitive resistors and an IMU. The objective is to provide a portable, low-cost system for real-world applications in injury prevention and rehabilitation. This method bridges

† Palawich Giraruchataporn and Inês Pinto contributed equally to this work.

the gap between biomechanics research and practical deployment in athletic and clinical settings.

2. METHODOLOGY

The conventional method for KAM estimation employs a 3D motion tracking system with force plates in a lab setting. Although accurate, this setup is bulky, costly, and impractical for diverse or real-world use. To overcome these limitations, this study proposes a wearable sensor-embedded shoe paired with a machine learning model for portable and reliable KAM estimation.

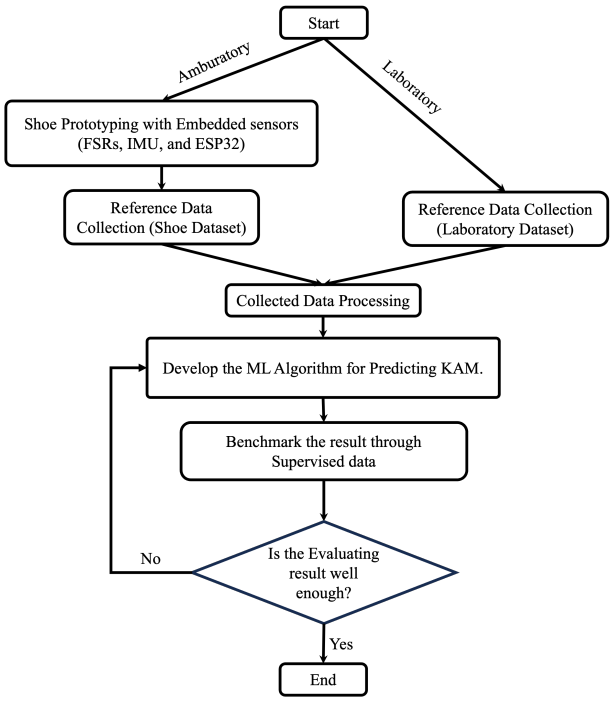

Fig. 1 Workflow diagram.

Figure 1 outlines the methodology. A smart shoe prototype was developed with six Force Sensitive Resistors (FSRs) and an Inertial Measurement Unit (IMU). An ESP32 microcontroller manages the sensors, integrated into the shoe as shown in Figure 2.

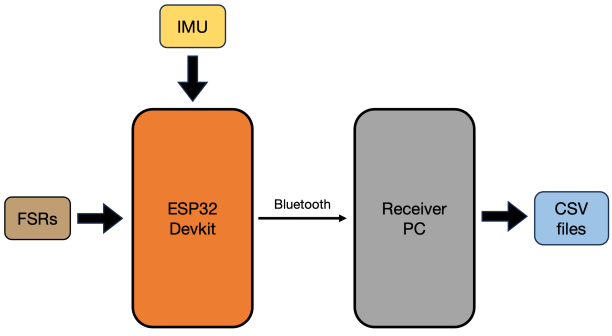

Fig. 2 System diagram.

Data collection involves two parallel streams: lab and shoe sensor data. Lab data (Figure 3) is acquired using a 3D motion capture system and GRF plates at 100 Hz, serving as the ground truth. Simultaneously, the smart shoe (Figure 4) collects FSR data (in Newtons) and IMU angular velocities (roll, pitch, yaw) at the same frequency, aligned via timestamps with the gait cycle.

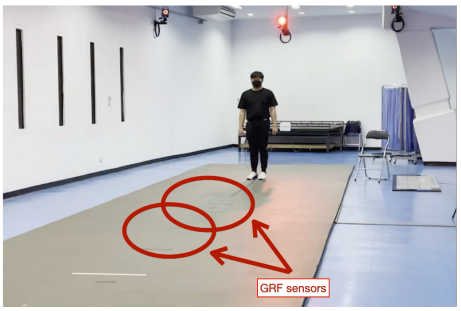

Fig. 3 3D motion capture system [10].

Fig. 4 Hardware diagram [10].

The dataset includes 97 steps from two subjects (height: 176 and 180 cm; weight: 102 and 80 kg). A sample is shown in Figure 5. Data was split into two configurations: (1) individual training per subject (Data1 and Data2), and (2) combined dataset. Each configuration was split 80:20 for training and testing.

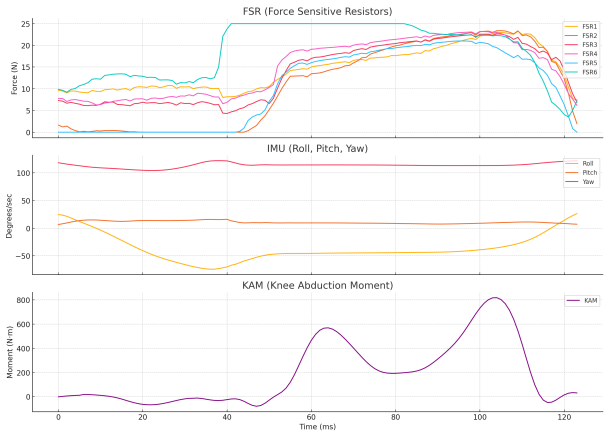

Fig. 5 Collected shoe and lab data.

To handle missing values, we used two imputation methods. First, linear interpolation, estimating values via straight-line approximation:

$$y = y_0 + \frac{(y_1 - y_0) \cdot (x - x_0)}{x_1 - x_0} \quad (1)$$

Second, K-Nearest Neighbors (KNN) imputation, which predicts missing values from nearest feature-space neighbors. These were applied prior to training.

We then trained a supervised model to estimate KAM from sensor data. A Histogram-based Gradient Boosting Regressor (HGBR) was selected for its native handling of missing values [11]. HGBR uses histogram binning to speed training and reduce memory. It learns optimal data split directions, builds tree ensembles correcting prior residuals, and handles categorical inputs natively. We used scikit-learn's default: 100 iterations, learning rate 0.1, max 31 leaf nodes, with seed 42 for reproducibility.

The model input was a 9-dimensional vector: six FSR readings (12-bit analog converted to Newtons) and three IMU angular velocities (deg/sec). The output was KAM from lab ground truth. Table 1 summarizes the configuration.

Table 1 Configuration of the HGBR model

Component	Details
Model Type	HGBR
Input Features	9 total (6 from FSRs, 3 from IMU)
Output Variable	KAM
Training/Test Split	80% training, 20% testing
Learning Rate	0.1
Number of Iterations	100
Maximum Leaf Nodes	31
Minimum Samples per Leaf	20
L2 Regularization	0.0
Early Stopping	Enabled (10% validation set)

To evaluate model performance, we used three regression metrics. Let y_i be true KAM, $\hat{y}_i$ the prediction, and $\bar{y}$ the mean of true values.

R^2 (Equation 2) measures variance explained:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

RMSE (Equation 3) gives average error magnitude:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

Percentage Error (Equation 4) quantifies relative deviation:

$$PercentageError = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{4}$$

These metrics provide a well-rounded assessment of model accuracy and consistency. The next section discusses results under different dataset and imputation strategies.

3. EVALUATION AND RESULT

To evaluate the model's effectiveness in predicting KAM, we conducted experiments in two settings: (1) individual-specific modeling, where models are trained separately per subject, and (2) a generalized modeling approach that combines data from both individuals into a single dataset. We also compared two imputation techniques: KNN and linear interpolation.

Table 2 Model performance under each data handling method and dataset setting.

Data Handling Method	Data Type	R^2	RMSE	% Error
Linear Interpolation	Data1	0.986	9.24	0.779%
	Data2	0.964	16.3	2.59%
	Total datasets	0.929	12.5	1.66%
KNN Imputation	Data1	0.883	103	143%
	Data2	0.771	162	223%
	Total dataset	0.823	130	198%

As shown in Table 2, KNN imputation led to substantially lower accuracy across all scenarios. On "Data1", which had relatively clean sensor signals, the model achieved R^2 = 0.883 and RMSE = 103 (143% error). For "Data2", performance dropped further to R^2 = 0.771, RMSE = 162, and 223% error. When combining both datasets, the generalized model yielded RMSE = 130 and 198% error, indicating high prediction instability.

These results highlight the limitations of distance-based imputation for time-series sensor data. KNN's reliance on nearest neighbors in feature space often disrupts temporal continuity, which is crucial for accurate biomechanical predictions like KAM. The method can introduce discontinuities or amplify noise, degrading model consistency and reliability.

In contrast, linear interpolation consistently outperformed KNN shown in Figure 6. On "Data1", it achieved the best result with R^2 = 0.986, RMSE = 9.24, and 0.779% error. Even for "Data2", it produced strong outcomes (R^2 = 0.964, RMSE = 16.3, error = 2.59%), showing robustness across subjects.

When using the combined dataset, the generalized model maintained high performance (R^2 = 0.929, RMSE = 12.5, error = 1.66%). Although slightly less accurate than the individual-specific model, it demonstrated that a single model can effectively generalize across individuals ideal for scalable, real-world deployment where subject-specific training may not be practical.

In summary, linear interpolation is superior to KNN for handling missing data in gait-based KAM prediction. Its structure-preserving nature aligns well with the temporal dynamics of biomechanical signals. The HGBR model, when paired with linear interpolation, delivered strong and stable performance across both personalized and generalized settings. These findings emphasize the importance of choosing imputation strategies that suit the sequential structure of the data.

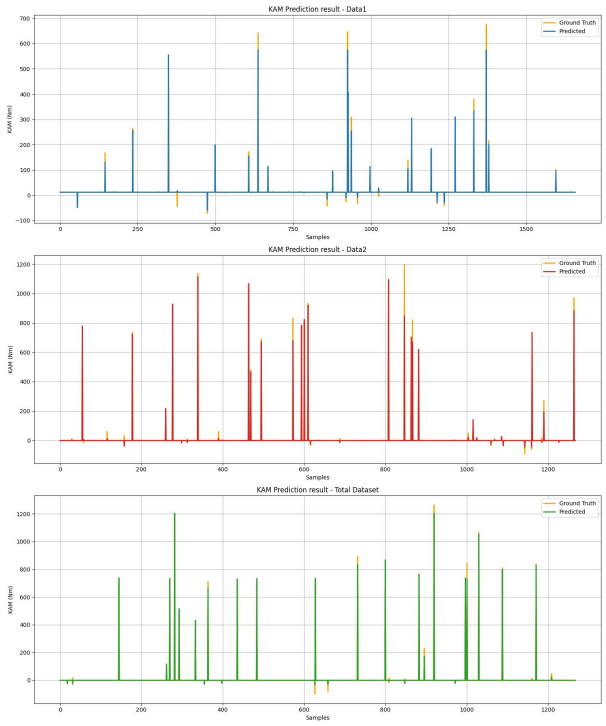

Fig. 6 KAM prediction results using linear interpolation for Data1, Data2, and the combined dataset.

4. CONCLUSION

This study presents a histogram-based gradient boosting regressor approach to predict KAM using data from a wearable smart shoe equipped with FSRs and an IMU. By eliminating the reliance on bulky, expensive laboratory equipment such as 3D motion tracking capture systems and GRF, our system provides a portable and accessible solution for KAM monitoring in natural environments. Two datasets collected from different individuals were used to train and evaluate the model with the HGBR, with performance assessed using R^2, RMSE, and percentage error.

The results demonstrate that our HGBR models using linear interpolation for missing data imputation yielded the highest prediction accuracy, particularly for subjects with more consistent gait patterns. A generalized model trained on combined data achieved slightly lower accuracy but demonstrated better generalizability, making it more suitable for broader applications. In contrast, using KNN for data imputation significantly degraded performance across all metrics, suggesting that simpler methods like linear interpolation may offer greater robustness for this application.

5. ACKNOWLEDGEMENT

We would like to thank Sirindhorn International Institute of Technology, Thammasat University for funding this research. We also thank to the Department of Sport Science and Sports Development, Faculty of Allied Health Science, Thammasat University for providing the Vicon-3D-tracking system and helping us for collecting the supervised data.

REFERENCES

[1] A. Cronström, M. W. Creaby, and E. Ageberg, *Do knee abduction kinematics and kinetics predict future anterior cruciate ligament injury risk? A systematic review and meta-analysis of prospective studies*, BMC musculoskeletal disorders, vol. 21, pp. 1–11, 2020.

[2] E. Wellsandt, T. Kallman, Y. Golightly, et al., *Knee joint unloading and daily physical activity associate with cartilage T2 relaxation times 1 month after ACL injury*, J. Orthop. Res., vol. 40, no. 1, pp. 138–149, 2022.

[3] D. Simon, R. Mascarenhas, B. M. Saltzman, M. Rollins, B. R. Bach Jr, and P. MacDonald, *The Relationship between Anterior Cruciate Ligament Injury and Osteoarthritis of the Knee*, Adv. Orthop., vol. 2015, p. 928301, 2015.

[4] S. J. Snyder, E. Chu, J. Um, Y. J. Heo, R. H. Miller, and J. K. Shim, *Prediction of Knee Adduction Moment Using Innovative Instrumented Insole and Deep Learning Neural Networks in Healthy Female Individuals*, The Knee, vol. 41, pp. 115–123, 2023.

[5] D. Jung, C. Lee, and H. S. Jeon, *Multi-Model Gait-Based KAM Prediction System Using LSTM-RNN and Wearable Devices*, Applied Sciences, vol. 14, no. 22, p. 10721, 2024.

[6] Z. Altai, I. Boukhennoufa, X. Zhai, A. Phillips, J. Moran, and B. X. W. Liew, *Performance of Multiple Neural Networks in Predicting Lower Limb Joint Moments Using Wearable Sensors*, Frontiers in Bioengineering and Biotechnology, vol. 11, 2023.

[7] Kistler Group, *3D Force Plate*, Kistler, Available: `https://www.kistler.com/PT/pt/3d-force-plate/C00000090`, [Accessed: April 12, 2025].

[8] L. Höschler, C. Halmich, C. Schranz, J. Fritz, S. Čigoja, M. Ullrich, A. D. Koelewijn, and H. Schwameder, *Wearable-Based Estimation of Continuous 3D Knee Moments During Running Using a Convolutional Neural Network*, Sports Biomechanics, pp. 1–19, Advance online publication, 2025.

[9] I. Boukhennoufa, Z. Altai, X. Zhai, V. Utti, K. D. McDonald-Maier, and B. X. W. Liew, *Predicting the Internal Knee Abduction Impulse During Walking Using Deep Learning*, Frontiers in Bioengineering and Biotechnology, vol. 10, 2022.

[10] P. Giraruchataporn, K. Ekkachai, P. Peuchpen, S. Kijpaiboonwat, W. Kongprawechnon, and S. Hasegawa, *Smart Shoe for Predicting Knee Abduction Moment*, Proceedings of the 13th Asian Control Conference (ASCC), pp. 162–166, 2022.

[11] A. Perez-Lebel, G. Varoquaux, M. Le Morvan, J. Josse, and J. B. Poline, *Benchmarking missing-values approaches for predictive models on health databases*, GigaScience, vol. 11, giac013, 2022.