Knee Osteoarthritis Classification System Examination on Wearable Daily-Use IMU Layout

Chengshuo Xia
Keio University
Japan
csxia@keio.jp

Tsubasa Maruyama
National Institute of Advanced Industrial Science and Technology
Japan
tbs-maruyama@aist.go.jp

Haruki Toda
National Institute of Advanced Industrial Science and Technology
Japan
haruki-toda@aist.go.jp

Mitsunori Tada
National Institute of Advanced Industrial Science and Technology
Japan
m.tada@aist.go.jp

Koji Fujita
Tokyo Medical and Dental University
Japan
fujiorth@tmd.ac.jp

Yuta Sugiura
Keio Univeristy
Japan
sugiura@keio.jp

ABSTRACT

The diagnostic approach for knee osteoarthritis that draws on kinematic characteristics provides a solution other than imaging medicine. However, the gait-based kinematic analysis still requires a motion capture suit as a prerequisite to ensure a reliable calculation, which limits the daily screening of the end user. To further reduce the cost, in this paper we investigated a wearable inertial measurement unit (IMU)-based knee osteoarthritis classification system based on daily-use wearing IMU layout and machine learning approaches. The acceleration and angular velocity signal output from the IMU were used as the input data; the different features from the time and frequency domains were examined with different handcrafted feature classifiers, as well as the deep learning method. From the results, using three IMUs could reach a 0.82 area under the curve value, with a sensitivity of 86% and a specificity of 78%. The results showed that using daily IMU devices to establish a diagnostic system with an on-body sensor layout is feasible.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

KEYWORDS

Inertial Measurement Unit; Layout; Knee Osteoarthritis; Screening

ACM Reference Format:
Chengshuo Xia, Tsubasa Maruyama, Haruki Toda, Mitsunori Tada, Koji Fujita, and Yuta Sugiura. 2022. Knee Osteoarthritis Classification System Examination on Wearable Daily-Use IMU Layout. In *The 2022 International Symposium on Wearable Computers (ISWC '22), September 11–15, 2022, Cambridge, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3544794.3558459

ISWC '22, September 11–15, 2022, Cambridge, United Kingdom

ACM ISBN 978-1-4503-9424-6/22/09…$15.00
https://doi.org/10.1145/3544794.3558459

1 INTRODUCTION

Knee osteoarthritis (OA) is one of the most common diseases in older adults [7]. The disease is typically accompanied by symptoms of knee stiffness and pain, as well as motion range influence [11]. The clinical diagnosis of knee OA plays an important role in preventing and decreasing its effects. Powerful machine learning tools have been employed to assist professionals and experts in making diagnoses and are effective in identifying knee OA severity, treatment planning, and so on [9].

Medical data, such as X-ray and magnetic resonance imaging (MRI) results, can assist diagnostic systems in determining knee OA patients' Kellgren–Lawrence (KL) grades to understand the disease process. Moreover, as an orthopedic disease, other biomechanical data can also be employed to determine the salient features, such as knee joint angle and ground reaction force (GRF), of normal healthy individuals versus those with knee OA [6]. However, currently, creating a patient-friendly pervasive system for screening and diagnosis still has limitations. The requirement of expert medical equipment impedes disease diagnosis of several potential patients in daily life. So, commonly accessible out-laboratory devices have been extensively employed to establish a pervasive diagnosis method, especially in identifying a cost-effective way to help patients determine their risk of disease.

Inertial measurement units (IMUs) can be used as wearable devices in combination with electronic devices. Thus, the utilization of IMU data in identifying and determining the disease status of patients with knee OA is important for the promotion of mobile medical diagnoses and health systems. The application of IMU in knee OA-related systems has been studied. Most of these studies have focused on using IMU to analyze gait parameters in patients with knee OA. Such systems investigate the interpretation of specific biomechanical parameters that differ between patients with knee OA and normal patients. However, these investigations typically require a specific number of sensors mounted on a specific limb site to obtain the desired gait parameters; for example, five to seven sensors are required for lower limb movement reconstruction. Thus, the gait parameter-based feature of such systems complicates their ubiquitous application.

Thus, IMU data can be used to build a classifier to identify the user's motion data and determine the risk of illness to analyze and discern the user's situation directly. This paper investigates the use of IMU sensors for the disease classification of knee OA. By using walking as the main activity, the effects of different classifiers, different numbers of sensors, and different sensor locations on the performance of the classifier were examined. This paper provides a basis for further screening of daily knee OA using simpler end-user electronics.

2 RELATED WORK

IMU has been widely used in various motion detection systems due to its convenience, such as in orthopedic disease studies, to help in disease detection and patient rehabilitation, and so on [4, 10, 16]. IMUs can be employed to measure temporal and spatial gait metrics, for example, using peak detection to assess the stance and swing time [5], and to estimate the joint angle, as well as to integrate acceleration to calculate the stride length [3, 12]. Using the IMU data in medical diagnosis, Parkinson's disease has been extensively studied to evaluate disease characteristics, diagnostic systems, and post-treatment management. Caramia et al. [2] evaluated several gait features and sensor locations using an IMU device to classify patients with Parkinson's disease and healthy people. Lin et al. [8] also proposed a neural network model to achieve early-stage detection of Parkinson's disease using IMU data. However, given the differences in disease characteristics, directly transferring the relevant findings of the Parkinson's disease system to knee OA is difficult. Thus, this paper discusses the use of IMU output data to construct a knee OA diagnostic system regarding the data source and sensor locations/numbers.

3 MATERIALS

3.1 Participants

Thirty-six individuals (average age: 73.4 years; 9 males and 27 females) who had knee OA of different KL grades (KL > 2) were in the knee OA patient group. The control group consisted of 14 normal participants (average age: 69.4 years; 2 males and 12 females). All participants could walk independently and did not require any external assistance. The study and the experimental protocol was approved by the local institutional review board (M2018-123). All participants had written a consent agreement before the measurements.

3.2 IMU Data Collection and Processing

IMUs (MTw; Xsens Technologies Inc.) were adopted to capture the acceleration, and angular velocity data [13]. Thirteen sensors were deployed on the participants' key body segments (Figure 1). During the experiment, the participants wore personal shoes to ensure a natural and suitable walking condition. The main motion tested was walking. After the entire IMU system was calibrated, the participants were requested to walk on a 12 m-long straight walkway with a self-determined walking speed. Moreover, the effective data were extracted by first conducting up-sampling for captured acceleration and angular velocity to realize a 120 Hz sampling rate. A 55-second data frame was gained from each subject's

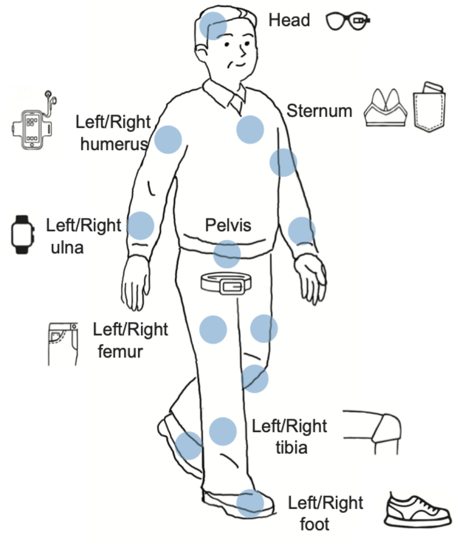

Figure 1: Examined IMU sensors' locations and possible practical usage with daily equipment.

whole walking data. A low-pass Butterworth filter was utilized to help eliminate noise.

3.3 Classification System

The overall task can be recognized as a binary classification to identify patients with knee OA and normal people. Thus, both the handcrafted feature and deep learning model were examined, including the SVM, decision tree, random forest, voting classifier (based on SVM, decision tree, and random forest), and baseline convolutional neural network (CNN) classifier, to test the performance of a system related to the specific sensor layout. System accuracy was calculated by exploiting leave-one-out-cross-validation (LOOCV), as the main application scenario was related to disease screening. The AUC value, sensitivity, and specificity were calculated to assess the classification system. The classification systems are illustrated in Figure 2.

The handcrafted features were extracted from the time domain and frequency domain [1]. The employed features were: *mean*, *variance*, *standard variance*, *75th percentile*, *25th percentile*, *mean and median value of power spectrum*, *mean and median frequency of power spectrum*, and *Entropy*. The data augmentation method from [14] was employed during the LOOCV test to balance the healthy dataset (during the evaluation, the tester's data was not augmented while other healthy people's data used for training were augmented). The signal's magnitude warping, permutation, and time-warping were adopted to retrofit the initial healthy data and thus augment the healthy dataset to balance the training dataset. Thus, the size of the patient dataset was 6 (axis) × 55 (time length) × 120 (Hz) × 36 (people), and the control dataset was 6 (axis) × 55 (time length) × 120 (Hz) × 14 (people) × 2 (augmentation). In addition,

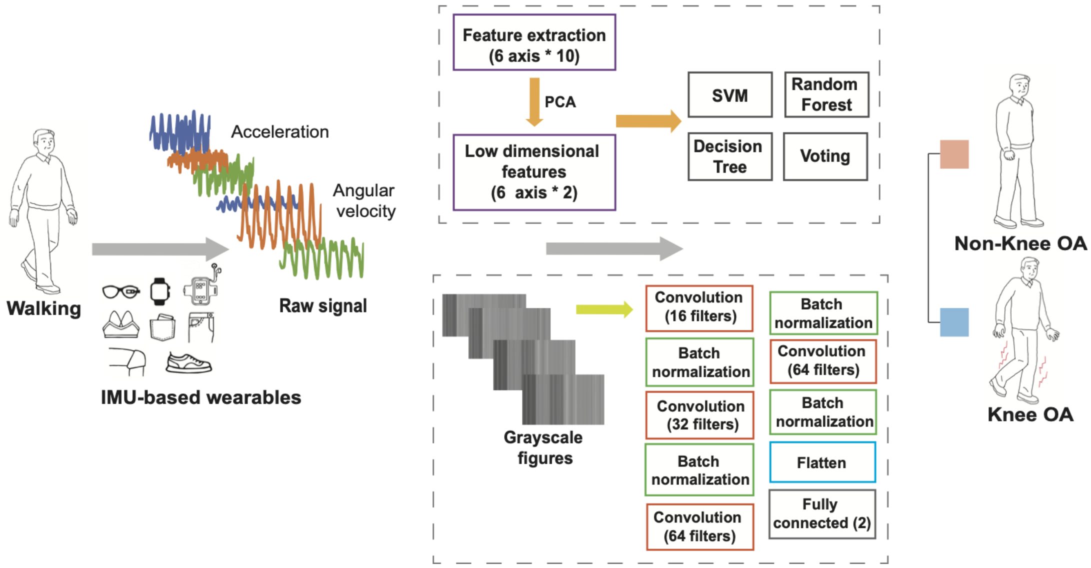

Figure 2: Classification testing systems built with various classifiers.

for the deep learning method, the accelerometer and gyroscope's time-series data were aligned to 6 rows and 55 × 120 columns. The data were then converted to grayscale figures for classification [15].

3.4 Sensor Layout Examination

Considering the practical sensor wearing placement, Figure 1 also shows all possible wearing locations, from the anatomy of the human body structure to the practical possible wear situation (combined with daily accessories and clothes). In evaluating a more practical wearable IMU layout scheme, considering both the sensor number and system performance is necessary. With more sensors attached, the higher dimensional data would benefit from analyzing more motion kinematic and kinetic information. However, the wearables must focus on a more portable and convenient condition; adopting additional sensors also creates shortcomings in the utilization and application for the end user. Thus, the sensor numbers from 1 to 3 should be investigated to evaluate the practical layout scheme.

Thus, the sensor layout studies would consider total of 13 sensor positions as shown in Figure 1, including the *head, sternum, left humerus, right humerus, left ulna, right ulna, pelvis, left femur, right femur, left tibia, right tibia, left foot and right foot.* An iterative approach was used to find the optimal one sensor position, two sensor position combinations and three sensor position combinations among these 13 positions.

4 RESULTS

The experiments were performed on a laptop (Thinkpad X1) with Python 3.6. The deep learning method was developed based on TensorFlow 2.1, and the epoch size was 10. The results of the experiment are presented in Table 1, 2 and 3.

Table 1: Testing result for 1-sensor used with AUC, sensitivity (Sen) and specificity (Spe).

Classifier	Best sensor location	AUC	Sen	Spe
Random Forest	Left foot	0.53	72%	36%
SVM	Left foot	0.61	72%	50%
Decision Tree	**Right lower arm**	**0.65**	**66%**	**64%**
Voting	Right lower leg	0.62	50%	50%
CNN	Pelvis	0.65	72%	57%

In general, the performance of different classifiers was similar, and the handcrafted features showed good performance. Specifically, with only one IMU, the classification between the healthy and patients with knee OA reached 66% sensitivity and 64% specificity. When two IMUs were employed, the SVM, decision tree, and voting classifier could present an acceptable and similar result. The AUC value was 0.71–0.73, and the sensitivity was 71%–75%. The performance improved when an additional sensor (three IMUs). The decision tree and voting classifier showed good results, with an AUC value of over 0.81. For the voting classifier, the sensitivity was 86%, and the specificity was 78%.

Based on the preliminary results, the end limbs could introduce a good result for the wearing sensor layout, such as the lower arms and feet. The end limb's acceleration and angular velocity data could normally maintain more considerable variations during the walking motion. During the experiment, the participants were requested to walk naturally. The lower arm's and foot's motion may present a different pattern, as the walking status would differ between the healthy and patient groups. Among other knee OA screening or diagnosis systems, an AUC value of over 0.7 was generally reported

Table 2: Testing result for 2-sensor used.

Classifier	Best sensor location	AUC	Sen	Spe
Random Forest	Pelvis Right lower arm	0.69	75%	64%
SVM	**Right upper arm Left foot**	**0.73**	**75%**	**71%**
Decision Tree	Right upper arm Left lower arm	0.71	71%	71%
Voting	Pelvis Neck	0.71	71%	72%
CNN	Right foot Neck	0.68	85%	50%

Table 3: Testing result for 3-sensor used.

Classifier	Best sensor location	AUC	Sen	Spe
Random Forest	Pelvis Right upper leg Left foot	0.67	78%	57%
SVM	Pelvis Left upper arm Right upper leg	0.74	78%	71%
Decision Tree	Right lower arm Left lower arm Right foot	0.81	83%	78%
Voting	**Right lower arm Left lower arm Right foot**	**0.82**	**86%**	**78%**
CNN	Pelvis Left lower leg Left foot	0.71	93%	50%

[6]. The preliminary results showed that using one sensor is not enough to ensure a good classifier (Table1), the two and three sensors deployed were mainly analyzed and discussed.

5 DISCUSSION

5.1 Sensor Layout

For two sensors used, the combination of *right upper arm* + *left foot* presented an AUC value of 0.73, a sensitivity of 75%, and a specificity of 71%. Therefore, the combination based on the two sensors showed lower performance in classifying patients with knee OA. Increasing the sensor number resulted in improving the performance of the classification system. As additional dimensional data distribution is examined, more effective features can be found and used in classifier training. From the results, the combination of several sensors presented good results. For example, the *right forearm* + *left forearm* + *right foot* showed an AUC value of 0.82, with a sensitivity of 86% and a specificity of 78%. For other possible combinations, the *sternum* + *right upper leg* + *left lower leg* showed an AUC value of 0.78, with a sensitivity of 78% and a specificity of 86%. The *right upper leg* + *left upper leg* + *left lower leg* showed an AUC value of 0.75, with a sensitivity of 84% and a specificity of 64%. Thus, using the combinations distributed between the sternum and lower limbs, as well as concentrated on the lower limbs, may be feasible in building a good classification system.

5.2 Classifiers and Handcrafted Features

Better results could be obtained when handcrafted feature-based classifiers, such as SVM, decision trees, and voting classifiers, were used. In contrast to a vision-based system, low-dimensional data typically cannot contribute to more efficient features, so the deep learning model is hard to show a better performance. Moreover, no type of classifier shows dominance. In addition, the main features extracted from the raw acceleration and angular velocity were related to their time and frequency characteristics. From an explainable aspect, such metrics are all related to the walking pattern and its kinematic characteristics, such as walking pace and steps. Setting the data length processed by classification to 55 seconds provided more information on the natural walking status of the different groups. Principal component analysis was also significant in decreasing the feature dimensions to balance the effective information and redundant features.

5.3 More Sensors Used

So far, only 1-3 sensor layouts were assessed from 13 candidate on-body sensor positions. And the preliminary test showed that with more sensors used, it is possible to obtain a good-performance classification system. However, the sensor used here is considered from both the practical situation and richness of data distribution. From the data perspective, as more sensors are used, more data distribution information is likely to use and figure out more salient features. However, the homogeneity of the data may also lead to the overfitting of the classifier. While on the other hand, using more sensors will also increase the burden on the user side, making it challenging to use in practice. Finding a good balance between the practical situation and classification performance still needs more attention, and it is important to building a reliable daily screening system.

6 CONCLUSION

This study mainly explored the feasibility of utilizing raw IMU data to build a classification system for knee OA screening. Through its design and validation, the handcrafted feature-based classifier could perform the classification with only two or three IMUs used. The results showed that combining the upper-end limbs (like the forearm) and lower-end limbs (e.g., the foot) may lead to good performance for classification. It presents the opportunity to fuse the commercial off-the-shelf device to realize a daily screening at the user end. However, the current training and testing dataset are still at a small level, and collecting additional datasets is still necessary to improve the screening system. Thus, for the next step, expanding the dataset and exploring more efficient feature expressions will be the focus.

ACKNOWLEDGMENTS

This work was supported by JST PRESTO Grant Number JP-MJPR2134.

REFERENCES

[1] Ahmed Ayman, Omneya Attalah, and Heba Shaban. 2019. An efficient human activity recognition framework based on wearable imu wrist sensors. In *2019 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 1–5.

[2] Carlotta Caramia, Diego Torricelli, Maurizio Schmid, Adriana Munoz-Gonzalez, Jose Gonzalez-Vargas, Francisco Grandas, and Jose L Pons. 2018. IMU-based classification of Parkinson's disease from gait: A sensitivity analysis on sensor location and feature selection. *IEEE journal of biomedical and health informatics* 22, 6 (2018), 1765–1774.

[3] Ye Chen, Weijian Hu, Yu Yang, Jijian Hou, and Zhelong Wang. 2014. A method to calibrate installation orientation errors of inertial sensors for gait analysis. In *2014 IEEE International Conference on Information and Automation (ICIA)*. IEEE, 598–603.

[4] Richard AW Felius, Marieke Geerars, Sjoerd M Bruijn, Jaap H van Dieën, Natasja C Wouda, and Michiel Punt. 2022. Reliability of IMU-based gait assessment in clinical stroke rehabilitation. *Sensors* 22, 3 (2022), 908.

[5] Alberto Ferrari, Pieter Ginis, Michael Hardegger, Filippo Casamassima, Laura Rocchi, and Lorenzo Chiari. 2015. A mobile Kalman-filter based solution for the real-time estimation of spatio-temporal gait parameters. *IEEE transactions on neural systems and rehabilitation engineering* 24, 7 (2015), 764–773.

[6] C Kokkotis, S Moustakidis, E Papageorgiou, G Giakas, and DE Tsaopoulos. 2020. Machine learning in knee osteoarthritis: A review. *Osteoarthritis and Cartilage Open* 2, 3 (2020), 100069.

[7] Michelle J Lespasio, Nicolas S Piuzzi, M Elaine Husni, George F Muschler, AJ Guarino, and Michael A Mont. 2017. Knee osteoarthritis: a primer. *The Permanente Journal* 21 (2017).

[8] Chin-Hsien Lin, Fu-Cheng Wang, Tien-Yun Kuo, Po-Wei Huang, Szu-Fu Chen, and Li-Chen Fu. 2022. Early Detection of Parkinson's Disease by Neural Network Models. *IEEE Access* 10 (2022), 19033–19044.

[9] Erika N Ringdahl and Sandesh Pandit. 2011. Treatment of knee osteoarthritis. *American family physician* 83, 11 (2011), 1287–1292.

[10] Robbin Romijnders, Elke Warmerdam, Clint Hansen, Julius Welzel, Gerhard Schmidt, and Walter Maetzler. 2021. Validation of IMU-based gait event detection during curved walking and turning in older adults and Parkinson's Disease patients. *Journal of neuroengineering and rehabilitation* 18, 1 (2021), 1–10.

[11] Leena Sharma. 2021. Osteoarthritis of the knee. *New England Journal of Medicine* 384, 1 (2021), 51–59.

[12] Ryo Takeda, Giulia Lisco, Tadashi Fujisawa, Laura Gastaldi, Harukazu Tohyama, and Shigeru Tadano. 2014. Drift removal for improving the accuracy of gait parameters using wearable sensor systems. *Sensors* 14, 12 (2014), 23230–23247.

[13] Haruki Toda, Tsubasa Maruyama, Koji Fujita, Yuki Yamauchi, and Mitsunori Tada. 2021. Self-reported walking difficulty associated with stiff-knee gait in Japanese patients with knee osteoarthritis: a preliminary cross-sectional study. In *Healthcare*, Vol. 9. MDPI, 1308.

[14] Terry T Um, Franz MJ Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 2017. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th ACM international conference on multimodal interaction*. 216–220.

[15] Chengshuo Xia, Ayane Saito, and Yuta Sugiura. 2022. Using the virtual data-driven measurement to support the prototyping of hand gesture recognition interface with distance sensor. *Sensors and Actuators A: Physical* 338 (2022), 113463.

[16] Shaghayegh Zihajehzadeh, Omar Aziz, Chul-Gyu Tae, and Edward J Park. 2018. Combined regression and classification models for accurate estimation of walking speed using a wrist-worn IMU. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 3272–3275.