

RESEARCH

Open Access



Multi-sensor fusion based on multiple classifier systems for human activity identification

Henry Friday Nweke^{1,2*}, Ying Wah Teh^{1*}, Ghulam Mujtaba^{1,3}, Uzoma Rita Alo⁴ and Mohammed Ali Al-garadi⁵

*Correspondence:

henrynweke@siswa.um.edu.my; tehyw@um.edu.my

¹ Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

Full list of author information is available at the end of the article

Abstract

Multimodal sensors in healthcare applications have been increasingly researched because it facilitates automatic and comprehensive monitoring of human behaviors, high-intensity sports management, energy expenditure estimation, and postural detection. Recent studies have shown the importance of multi-sensor fusion to achieve robustness, high-performance generalization, provide diversity and tackle challenging issue that maybe difficult with single sensor values. The aim of this study is to propose an innovative multi-sensor fusion framework to improve human activity detection performances and reduce misrecognition rate. The study proposes a multi-view ensemble algorithm to integrate predicted values of different motion sensors. To this end, computationally efficient classification algorithms such as decision tree, logistic regression and k-Nearest Neighbors were used to implement diverse, flexible and dynamic human activity detection systems. To provide compact feature vector representation, we studied hybrid bio-inspired evolutionary search algorithm and correlation-based feature selection method and evaluate their impact on extracted feature vectors from individual sensor modality. Furthermore, we utilized Synthetic Over-sampling minority Techniques (SMOTE) algorithm to reduce the impact of class imbalance and improve performance results. With the above methods, this paper provides unified framework to resolve major challenges in human activity identification. The performance results obtained using two publicly available datasets showed significant improvement over baseline methods in the detection of specific activity details and reduced error rate. The performance results of our evaluation showed 3% to 24% improvement in accuracy, recall, precision, F-measure and detection ability (AUC) compared to single sensors and feature-level fusion. The benefit of the proposed multi-sensor fusion is the ability to utilize distinct feature characteristics of individual sensor and multiple classifier systems to improve recognition accuracy. In addition, the study suggests a promising potential of hybrid feature selection approach, diversity-based multiple classifier systems to improve mobile and wearable sensor-based human activity detection and health monitoring system.

Keywords: Wearable sensors, Multiple classifier systems, Activity identification, Multiple sensor fusion, Activity detection, Multi-view stacking ensemble, Feature-level fusion

Introduction

In recent times, sensor technologies for health monitoring have advanced greatly due to the decrease in the cost and availabilities of sensor-embedded devices. The implementations and analysis of sensor data generated by these devices are vital in wide areas of applications such as smart homes, cyber-physical applications, assisted living, security, elderly care, lifelogging, and sports activities. In health-based applications, sensor data are analyzed to identify various simple and complex activities such as walking, running and doing basic household activities or operating industrial machinery [1]. In addition, sensor data analytics provide a mechanism to detect fall and inaccurate posture in the elderly population that may present a high risk of fall and Identification of what constitutes actual fall would aid prevention with their negative health cost tendencies [2].

Generally, human activity identification has been explored in various sensors types. These include wearable, video, ambient and smartphone-based methods [1, 3]. However, video-based methods are affected by lightening variability, inability to differentiate between target and non-target information during data collection, and issue bothering on user privacy. Besides, ambient sensors devices deployed for collections of data such as sound, pressure, temperature and vital signs are mostly installed in particular locations and may not be effective for ubiquitous health monitoring [1]. Lately, the use of wearable and smartphone embedded sensors for human activity identification have also attracted high interest among researchers. Wearable and smartphones are ubiquitous devices with varieties of built-in sensors such as accelerometers, GPS, gyroscopes, magnetometer, microphones, etc. for consistent monitoring of physiological signals, comprehensive health check, indoor localizations and pedestrian navigations [3, 4]. Applications of these devices for identification of various activity details are as results of their pervasiveness, continuous tracking of human activity details and provision of continuous monitoring through cyber-physical systems. Therefore, wearable and smartphone devices provide a better alternative for ubiquitous and continuous monitoring of activity details.

Although, there are many studies in human activity detection and health monitoring [5–9], they mainly focus on the use of several classification algorithms over extracted feature from single sensor modality. Moreover, this approach is built on the assumption that various sensor modalities provide the same statistical properties. In contrast, different sensor modalities embedded in mobile and wearable devices provide various statistical properties that ensure accurate detection of activity details. For instance, a motion-based sensor such as accelerometer measure acceleration forces that dynamically sense movement and vibration and ensure dynamic detection of movement patterns. On the other hand, gyroscope sensor measure angular velocity and orientation that provide complementary information for the detection of activities of similar patterns and strong displacement activities [10]. Furthermore, the magnetometer sensor helps to eliminate the effects of gravity, ensure independent device orientation, and differentiate between sporadic and static activities [11, 12]. Other sensor modalities embedded in mobile and wearable devices such as pulse rate, location-based sensor (GPS), altimeter, barometer, pressure, and heart rate are inclined to health applications such as energy expenditure estimation, strength training during vigorous exercise, mental load identification, health status monitoring, and disease management in elderly. Therefore, the use of a single

machine-learning algorithm over concatenated feature vectors for human activity detection might limit their performances. Also, it difficult to understand the contribution of each sensor modalities and find optimal features for activity classification. Finally, the approach may result in increased misclassification rate and inability to handle high dimensional data for comprehensive physical activity detection [3]. According to recent comprehensive evaluations of several classification algorithms [13, 14], no single classification model is sufficient for a particular human activity detection task.

To provide enhanced performance and diversity in recognition of human activity details, multi-sensor fusion strategies have been proposed in recent studies [15, 16], in which various sensor modalities outlined earlier are integrated using raw sensor signal, extracted features or decisions predicted by individual classification algorithms. The central ideas of fusion protocols are to incorporate diverse sensor modalities to increase reliabilities, robustness and enhanced generalization of human activity detection frameworks. In addition, multi-sensor fusion methods help to minimize uncertainty and effects of indirect captures which is quite challenging to eliminate with only single sensor modality [17].

Among the various methods for multi-sensor fusion, multiple classifier systems [3] that fuse several diverse machine learning algorithms to arrive at superior decisions than single classifier provide the best alternative. Multiple classifier systems methods are highly recommended to resolve issues on complexity, high dimensionality, and disparity in sensor data. This would result in improved accuracy, robustness and generalization of activity classification framework. Typical multiple classifier system methods that have played a vital role in human activity detection and health monitoring include bagging, boosting, sensor feature manipulation, model initialization and stacking ensemble [3, 18]. These methods take a random sampling of the training data or different weak classification algorithms to create diversities of opinions integrated through voting, fuzzy decision rule, Dempster–Shafer theory or Random committee.

However, few studies [16, 19, 20] have address the issue of multiple classifier system methods for human activity identification by utilizing diverse multimodal sensor data and classification algorithms. Specifically, these studies only developed protocols to integrate multiple accelerometer sensors attached at different body locations, which limit their implementation for robust activity recognition. Furthermore, accelerometer signals are sensitive to sensor location, drift and are ineffective for identification of dynamic or orientation based activities [21]. In contrast, this study utilizes motion sensors such as accelerometer, gyroscope, and magnetometer commonly found in wearable and smartphones devices to develop high performance, robust and efficient human activity recognition framework using multiple classifier system methods. Based on extensive comparative analysis, the paper proposes a robust multi-view stacking ensemble algorithm to detect common and complex daily activities. Stacking ensemble [22] is a multiple classifier method that exploits the predictive values of base classification algorithms to improve the generalization ability of human activity recognition framework. Therefore, multi-view stacking ensemble integrates data from multiple heterogeneous sources to build robust and efficient systems [23]. In this case, each sensor modality (accelerometer, gyroscope, and magnetometer) depicted as heterogeneous data represents different entities and feature space. The motivations for using the proposed multi-view stacking

ensemble algorithm is, to utilize complementary but distinct feature vectors from each sensor modalities and diversity base classifiers to build robust, flexible and efficient activity identification system.

Although recent study [18] have shown that the use of multi-view stacking ensemble methods can greatly improve human activity identification for mobile and wearable sensor data. However, there are still issues to tackle in order to implement a comprehensive and robust activity detection framework. First, current multi-view stacking ensemble algorithm utilizes the same classification algorithm as base classifiers to train each view and combined the predictive values with the same classification algorithm. In a recent study in other domain [24], multi-view stacking ensemble produces robust and efficient results with the implementation of diverse classification algorithms. Moreover, this paper utilized computationally efficient classification models such as decision tree, k-Nearest Neighbor, and logistic regression to implement the proposed activity detection framework. Second, many datasets for human activity recognition show high level of class imbalance [6]. Class imbalance problem is a difficult issue in human activity identification as it may lead to performance results that produce bias predictive values towards the majority activity classes and low performance towards the minority activity classes. And this issue frequently occurs in human activity identification and health monitoring after feature selection to reduce the size feature vectors [6]. The dataset used in this paper show some level of class imbalance, in which activity classes such as jumping, ascending and descending stairs have less number of instances compared to walking activity frequently performed in real life. To solve the problem, we apply Synthetic Minority Over-Sampling Technique (SMOTE) to the sensor data to increase the minority class and this approach improves the performance of our activity identification system [25]. Finally, the use of irrelevant feature vectors in human activity classification task would lead to overfitting, low performances and increased computation time. The paper proposes to implement a bio-inspired meta-heuristic evolutionary search algorithm integrated with correlation-based feature selection to produce compact feature vectors.

To overcome the weakness of existing multi-view stacking ensemble method and influenced by work done by [18], this study proposes diversity and multi-modal based human activity detection by incorporating enhanced methods to improve performance generalizations. As such, the goal of the paper is to evaluate the impact of the proposed multi-view stacking ensemble algorithms to improve the performance of human activity detection systems. In addition, we provide comprehensive comparison of the proposed methods against single sensor modality, feature-level fusion and three baselines to show significance. The experimental results and comparison provide practical applications for robust activity detection and monitoring, and serves as references for further implementation of multi-view based human recognition system. In addition, the extensive comparison in the paper will act as start-of-art methods to evaluate and compare future implementation of human activity detection framework, multi-sensor fusion and multiple classifier systems.

Contributions

The major contributions of this paper are presented below:

1. To propose a robust and efficient multi-view stacking ensemble algorithm for human activity identification and health monitoring. The developed algorithms are in three phases of implementations. *First*, k-Nearest Neighbors and decision tree were used as base classifiers to train each view in the dataset and k-Nearest Neighbors as meta-classifier (k-NN-DT-k-NN). *Second*, logistic regression, k-Nearest Neighbors and decision tree were used as base classifiers and logistic regression was used as meta-classifier (LR-k-NN-DT-LR). *Finally*, logistic regression, k-Nearest Neighbors and decision tree were deployed as base classifiers and average performance results of k-Nearest Neighbors and logistic regression as meta-classifiers (k-NN-DT-LR-(k-NN-LR)); we built the models after an extensive evaluation of the four single classification algorithms used in this paper. We compare both the proposed multi-view stacking methods, feature-level fusion, and single classifier performances;
2. To evaluate the impact of bio-inspired metaheuristic evolutionary search algorithm integrated with correlation based features selection algorithm to produce compact feature vectors for implementation of computationally efficient human activity identification framework.
3. To demonstrate the impact of Synthetic Minority Over-Sampling Technique (SMOTE) to balance the minority activity classes and reduce bias towards majority activity classes.
4. To provide analysis of the recent approaches for multi-sensor based human activity recognition.
5. Extensive experiments to explore the effectiveness of the proposed methods using two publicly available datasets and compare the significance of the multi-view stacking ensemble with *weighted majority voting*, *Bagging* and *Random Subspace* ensemble [16, 26, 27] based multiple classifier system methods.

Outline

The rest of this paper is structured as follows. “[Review of related works](#)” section presents the background and related works. “[Problem formulation](#)” section describes problem formulation for human activity identification using multi-view stacking ensemble methods. “[Proposed methodology](#)” section presents the multi-view stacking ensemble algorithm architecture that includes signal processing, feature extraction and normalization, feature selection and proposed algorithms. “[Experiments](#)” section discusses the experimental setups; the results obtained at each evaluation and compare the multiple classifier system methods with existing methods. “[Conclusion and future works](#)” section concludes the study.

Review of related works

Human activity recognition, detection, identification and monitoring are terms used interchangeably by various studies that implement approaches to assess the level of physical activities undertaken by individual using motion sensors [3, 28, 29]. In addition,

the process encompasses procedures for implementation of mobile and wearable sensors based activity assessment using various sensor modalities. These procedures include data collection; signal processing, feature extraction, feature selection and activity classification. For instance, Biagetti et al. [30] proposed wireless architecture for data acquisition and monitoring of sport activities using surface electromyography (sEMG) and accelerometer sensors. The authors achieved 83.7% accuracy using k-Nearest Neighbor classifier. In a recent study, Bhattacharjee et al. [31] evaluated various machine learning algorithms for daily activity monitoring. Machine learning algorithms evaluated using motion sensors collected with smartphone device include support vector machine, perceptron neural networks, backpropagation neural networks and recurrent neural networks. Further information on human activity monitoring using various sensor modality are reported in recent studies [3, 32]. In this paper, our review is focused on studies that developed protocols to integrated data, features and multiple classification algorithms for the purpose of activity monitoring and assessment.

A number of fusion methods have been proposed for the implementation of comprehensive human activity identification and monitoring using multiple sensors in recent years. These approaches are grouped into data-level, feature-level, and multiple classifier (decision fusion) frameworks. These methods integrate different sensors, feature vectors and classification algorithms for the purpose of human activity detection, assessment, prediction and monitoring. The data and feature-level fusion discussed in this section are those related to human activity recognition and monitoring using motion sensors of various modalities. In addition, the multiple classifier system methods are those related to activity prediction and classification. Moreover, we discuss the sensor used, the number of subjects for data collections, number of activities, sensor types used, strength, weakness of each method. Table 1 summarizes the various metrics of each paper implemented recently for human activity detection in the area of data fusion, feature-level fusion, and multiple classifier systems.

Data-level fusion

The use of a single source of information for human activity identification and health monitoring is challenging for effective recognition of complex activity details, comprehensive health monitoring and follow up recommendation [3]. Therefore, data or sensor level fusion is required to ensure effective activity classification. Consequently, data level fusion methods integrate raw sensor data obtained from various sensor modalities to improve performance efficiency and reliability. Recently, various methods have been implemented to fuse multiple sensors. Some of these methods include Dempster–Shafer theory (DST), Bayesian networks (BS), Kalman filtering, particle filter and graph-based theory [3, 33]. For instance, Tolstikov et al. [34] proposed Bayesian and Dempster–Shafer theory to combine various binary sensor data for daily activity detection of elderly citizens. The authors processed the sensor separately with dynamic Bayesian network and Dempster–Shafer theory giving rise to different operational efficiency and accuracy. Similarly, Amoretti et al. [35] evaluated the use of Bayesian Network for the fusion of different sensor modalities in ambient assisted livings environments. Tunca et al. [10] proposed the fusion of motion sensors (accelerometer and gyroscope) data using Kalman filtering for pathological gait analysis. Other modified versions of Kalman filtering have

Table 1 Summary of recent studies on human activity recognition through data fusion

Authors	Activities	Sensors/position	Algorithms/evaluation	Fusion type	Strength	Weakness
Tolstikov [34], Amoretti et al. [35], Sebak et al. [38], Tunca et al. [10], Qui et al. [4]	Static: sitting, standing, lying, sleeping, idle Mobility based: walking, climbing stairs, leaving home Household: prepare breakfast, prepare dinner, drink Daily hygiene: shower, use toilet	Sensors: Camera, accelerometer, gyroscope, magnetometer, binary sensor Position: ankle, toilet flush, cupboard Sampling rate: ~100 Hz	Algorithm: dynamic Bayesian network, Kalman filtering, Dempster-Shafer Evaluation metrics: error rates, accuracy, computation time, precision, recall, F-measure	Data fusion	Provide a simple and real-time, computationally efficient and independent implementation of human activity recognition	Inability to handle a long sequence of activities. Moreover, the approach is sensitive to sensor position, noise and sometimes impractical to implement
Nishida et al. [47], Spinante et al. [5], Shoaib et al. [9], Xu et al. [50], Chen and Wang [8], Berenguar et al. (2017), Zdravetski et al. [74], Fong et al. [45], Dobbins et al. [6], Köping et al. [46], San-Segundo et al. [48], Li et al. [49], Pires et al. [29]	Static: sitting, lying down, standing, reading, making calls, kneeling Mobility based: walking, jogging, ascending descending stairs, biking, stretching, object-lifting, bending, falling forward, falling left, falling backward, talking, cycling, nordic walking, jumping, car, running Household: eating, watching TV, open door, close door, open fridge, close fridge, open dishwasher, close dishwasher, open drawer, close drawer, close table, drink form cup, vacuuming, ironing, shopping, cooking Daily hygiene: brush teeth, bathing, washing clothes, drying clothes, house cleaning Transition activities: stand up from sitting, stand up from laying, going up, going down, laying down from standing Office: Typing, Writing, walking at computer Harmful habit: smoking	Sensor: accelerometer, gyroscope, magnetometer, linear acceleration, gravity, heart rate, location sensor, air pressure, video Position: front pocket, chest, ankle, thigh, forearm, wrist, waist, back, feet, Right shoulder Sampling rate: 20-200 Hz	Algorithms: k-NN, ANN, DT, SVM, NB, LR, RF, Hoefding tree, HMM, RNN, CNN, LSTM, LDC, QDC, POLYC, PARZENC, Gaussian Mixture Model, MLP, FNN, DNN Evaluation metric: AUC, accuracy, precision, recall, F-measure, specificity, computation time, error rate, Kappa, FP, FN, Confusion Matrix	Feature fusion	Use to fuse sensor of diverse modalities and less sensitive to noise	Feature incompatibility, instability to sensor failure and signal variation reduce performance

Table 1 (continued)

Authors	Activities	Sensors/position	Algorithms/evaluation	Fusion type	Strength	Weakness
Catal et al. [52], Gjoreski et al. [27], Peng et al. [54], Chowdhury et al. [19], Garcia-Ceja et al. [18], Saha et al. [16], Peng et al. [59]	Static: lying, sitting, standing, kneeling, sleeping, recreational activities Mobility based activities: running, walking, cycling, ascending stairs, descending stairs, jogging, travel, sports Household: washing dishes, mop floor, sweep the floor, eat chips, watch TV, shopping, brush teeth Hygiene based: cleaning, wash hand Office activities: working on the computer, meeting	Sensor: accelerometer, calorimeter, biosensor, body-media, accelerometer, sound, accelerometer, location data, vital signs Positions: chest, wrist, mouth, pocket, table (sound), shirt pocket, belt, bag sampling rate: 20–200 Hz	Base classifiers: MLR, SVR, GPR, M5P, MLP, SVM, RF, BDT, DNN, Adaboost, LR, J48, KNN Fusion method: posterior probability, majority voting, single classifier based multi-view stacking Evaluation metrics: F-measure, recall, precision, accuracy, Confusion Matrix	Multiple classifier systems	Can handle complex activity details, high dimensional sensor data and uncertainty by systematic classifier fusion. Combine heterogeneous and homogeneous classifier to reduce variance and ambiguity that are likely to occur in the single classifier	External knowledge dependencies and may be computationally complex based on the base classifier

also been implemented such as extended Kalman filtering, Quaternion based extended Kalman filtering and Rao-Blackwellization unscented Kalman filtering to deal with various challenges such as sensor orientation, postural instabilities and sensor placements in human activity detection [36, 37].

Furthermore, Qiu et al. [4] proposed an extended Kalman filtering approach to integrate motion sensors for pedestrian navigation application and noted that such method provides a robust algorithm for human activity detection. Sebbak et al. [38] proposed a Dempster–Shafer theory method to fuse varieties of sensor modalities for human activity identification and comprehensive health monitoring. The use of Dempster–Shafer theory helps to reduce uncertainty and imprecision in sensor representation and increases reliability. While Phan et al. [39] developed a graph-based theory to integrate social sensor and physical sensor data for context-aware activity recognition in order to reduce computation complexity in mobile-based implementation.

The main advantages of data-level fusion are the ability to provide simple, real-time, computationally efficient and problem independent implementation of human activity recognition [33]. However, some of the data level fusion methods are challenging to handle long sequence activity at real-time (such as Dempster–Shafer and Bayesian theory). In addition, data-level fusion is sensitive to sensor positions, noise and sometimes impractical to implement in real-time [40].

Feature-level fusion

Integration of feature vectors extracted from various sensor modalities is the most implemented fusion methods for human activity recognition. Feature-level fusion methods combine features extracted from mobile and wearable sensors such as ECG, GPS, accelerometer, gyroscope, magnetometer, visual sensors, etc. using various machine learning algorithms. The main attractions of feature-level fusion are the ability to fuse sensor from diverse devices and less sensitive to noise. In the last few decades, various studies have been published for human activity classification using feature level fusion. Here, this paper only discusses recent implementation while further discussion can be found in a recent review in the area [3, 7].

Sensor fusion using feature concatenation methods are simple to implement with less computation complexity and various studies in human activity recognition have proposed several techniques in these regards for inertial sensor and multimodal sensor fusion. For instance, Spinsante et al. [5] investigated frameworks for monitoring of physical activities in the workplace to minimize sedentary lifestyle by fusion of motion sensors using a decision tree classification algorithm. The techniques categorized the activities into an active or non-active, developed a mechanism for feedback update, and achieved high-performance accuracy. Also, to recognize concurrent activities, Chen and Wang [8] proposed a hierarchical algorithm for the fusion of accelerometer and gyroscope. Concurrent activities are performed simultaneously and include walking while brushing teeth, making a phone call while preparing a meal or watching TV and such activities require sensor of multi-modalities to recognize. In [41], the authors proposed aggregation of features extracted from inertial motion sensor for real-time posture detection and how determine the correlation between posture and action. Furthermore, the method was deployed to correct the effect of activity drift in pre-impact fall

detection, recognition of transition activities, human motion tracking, real-time context-aware navigation, and pedestrian location navigations. Zdravevski et al. [42] developed enhanced and real-time multimodal sensor-based activity detection and monitoring using logistic regression with a fusion of inertial sensors and physiological signals. Feature concatenation methods that involve the fusion of vision based sensors and inertial sensors have also been proposed using machine learning for human activity detection and health monitoring. The fusion methods enable identification of mobility changes, complex and concurrent activity details and behaviour tracking [43]. However, each sensor modalities provide different statistical properties for recognition of particular activity details and maybe not be optimal to aggregate these features before applying learning algorithms [44]. Moreover, the fusion of vision-based sensors with other sensor modalities is still challenging due to issues bothering on privacy and lack of scene semantics.

Recently, Fong et al. [45] proposed shadow features for efficient activity classification and health monitoring. The proposed feature vectors were computed from the dynamic nature of the human body motion and machine-learning algorithms were applied to infer dynamic body movement and underlying momentum of activity details. The main improvements of the proposed shadow features over previous studies are the incremental nature, simplicity and low computation time of shadow features for mobile and wearable device implementation. Shoaib et al. [9] evaluated the fusion of motion sensor for complex human activity detection using machine-learning algorithms. The authors extracted computationally efficient feature vectors from accelerometer, gyroscope, and magnetometer and combine these features using Naïve Bayes classifiers. On the other hand, Köping et al. [46] proposed comprehensive frameworks to integrate varieties of sensor modality by utilizing a codebook feature learning approach. The proposed framework integrates features extracted from smartphones, smartwatches, and smart glasses. The use of a codebook approach to extract underlying sensor data helps to summarize the local characteristics of the data and thereby improve activity detection accuracy. In a related method, Nishida et al. [47] evaluated a Gaussian mixture model-based fusion of accelerometer and acoustic sound for human activity recognition. Feature vectors were extracted from the sensors separately and then trained with a Gaussian mixture model, while combination was done using a weighted likelihood estimate to recognize indoor and outdoor activities. Also, San-Segundo et al. [48] evaluated motion sensor data fusion using features computed from time and frequency domain transformation of each sensor data. The authors modelled time variation in activity details that makes the algorithm robust against degradation. In addition, they propose long short-term memory to model long-term dependencies in activity variation.

In addition, a hybrid approach that combine conventional feature and automatic feature representation was proposed by Li et al. [49], in which the authors comprehensively evaluated both handcrafted features extraction methods and deep learning based features for human activity recognition. The authors concluded that the fusion of two deep learning algorithm (CNN and LSTM) provide better performance results. Furthermore, Dobbins et al. [6] propose a fusion of features extracted from multiple accelerometers attached at different body position for comprehensive health monitoring and activity recognition using 10 classification algorithms. The proposed method was enhanced by integrating visualization protocol to enable real-time activity recognition

using smartwatches. Finally, Xu et al. [50] proposed three-phase multi-level complementary feature learning approach to integrate low-level, mid-level and high-level features extracted from the motion sensor using kernel-based support vector machine. Moreover, evaluation of the proposed approach on three publicly available datasets shows enhanced performance improvements against existing low-level methods.

However, issues such as feature incompatibility, robustness to sensor failure and vulnerability to uncertain noise or interference due to variation sensitivity greatly reduce the performance of feature-level fusion methods. Furthermore, finding optimal features and feature extraction methods require extensive domain knowledge which is time-consuming [19]. Besides, there may be an issue related to high computation time from the extraction of semantic-based features and dictionary that results to inclusion of irrelevant features. The above challenges and limitations have made the use of feature-level fusion and single machine learning impractical for robust and efficient implementation of human activity detection system.

Multiple classifier systems

Recently, multiple classifier system methods that integrate decisions obtain from different machine learning algorithms to improve activity identification and comprehensive health monitoring have received great deal to research efforts [3]. The use of decision fusion approaches for human activity identification is necessitated by the need to improve the performance accuracy, robustness, efficiency, and generalizability of the single classification algorithm. Hence, multiple classifier systems fusion is appropriate to handle complex activity details, high dimension sensor data, and uncertainty by deploying systematic integration of individual classifier to produce consensus opinions. In addition, multiple classifier system methods combine heterogeneous and homogeneous classifiers to reduce the ambiguity that is unlikely when such classifier is used alone [15]. Moreover, multiple classifier system methods for human activity classification provide a mechanism to resolve issues related to diagnostic errors using classifier diversity, bias and variances, reduce computation complexity and better algorithm representation [3]. Therefore, integration of multiple decisions from individual classifier minimize issues related to overfitting, increase the probability of finding optimal solutions and enable efficient implementation of learning algorithms. Implementation of multiple classifier system would help to resolve issues such as pattern variations, signal degradation, sensor failures, and spatial variability of data, environmental fluctuation and insufficient computation resources can be minimized by the use decision fusion approaches for human activity classification [33, 51].

Here, we review some of the studies that recently developed multiple classifiers for human activity detection to set the stage and need for multiple classifier systems when developing human activity detection and health monitoring system. Gjoreski et al. [27] proposed multiple context decision ensemble for energy expenditure estimation from physical activity details. The authors trained multiple regression-based algorithms on different contexts (features) extracted from multiple sensors and combined the individual approach using majority voting. The proposed multiple context ensemble approach outperformed other ensemble algorithms such as Random space and bagging ensemble methods. Chowdhury et al. [19] proposed posterior adapted class label fusion method to

integrate multiple accelerometer sensor data attached at different placement positions of the body. The proposed method calculates class weights for each model and then fine-tuned these weights based on score functions using the posterior probability of the predicted class labels. Then, the class label with the highest score was selected as the final prediction. In their recent implementation, the authors further [26] evaluated different methods to combine decisions predicted by classification algorithms for human activity recognition. These include weighted majority voting, Naïve Bayes combiner and Behavior Knowledge Space (BKS) using multiple accelerometer sensor data. They noted that the majority voting outperformed ensemble learning approach such as random forest and bagged decision tree.

In a related study, Catal et al. [52] evaluated ensemble methods to integrate decisions generated with classification algorithms such as decision tree, multi-layer perceptron and logistic regression for human activity identification using accelerometer sensor data collected from 36 subjects. The evaluation showed the impact of decision-level fusion for human activity classification as the authors achieved 98% accuracy with an average of probability fusion method. Tripathi et al. [53] investigated the fuzzy decision rule algorithm that uses simple combination rule for adaptive based human activity identification. The authors formulated new classifier as a batch of new activity details. In addition, Peng et al. [54] proposed hierarchical complex human activity recognition frameworks by fusion of acceleration and physiological signals. The paper utilized diverse feature vectors computed from various sensor modalities such as acceleration and physiological signals by exploiting their differing modalities. Moreover, clustering algorithms were utilized to generate a component of complex activities and topic model to generate latent semantic of complex activity details. The output of the final classification is combined at the classifier level. They noted that the approach helps to reduce information loss and burden of data annotation. Guan and Plötz [55] implemented epoch bagging method for human activity identification by utilizing probabilities selection of the subset of the original data for mini-batch based training of Long Short Term Memory and stochastic based gradient descent learning. The main advantage of the approach is the ability to generate robust decision from each epoch values to improve the performances of the human activity detection framework.

Recently, a hierarchical algorithm that integrates sensors trained separately using each machine learning algorithm was developed by [56, 57]. The hierarchical fusion method train acceleration sensor attached at different body positions and combine the feature vectors with an asymmetrically weighted decision provided by each sensor with recall and precision as metrics for inclusion and rejection. The proposed method helps to resolve the problem of sensor anomalies and failures. Nonetheless, the methods were only applied on accelerometer sensor attached on different body locations. Besides, accelerometer is inefficient in recognition of the activity of similar patterns such as descending or ascending stairs or concurrent activities that include reading while watching TV, cooking while making calls [58]. Also, Peng et al. [59] proposed hierarchical complex activity recognition by fusion of accelerometer, location, and vital sign data. The data were processed and learned separately and combined at the classifier level in order to achieve generalizability and independence of different activity contexts. While Saha et al. [16] propose a two-phase ensemble algorithm for human activity recognition

by exploiting position specific condition to improve performance results. Therefore, the training and testing data were drawn from different placement positions.

One of the early studies for multi-view stacking ensemble for human activity recognition was proposed by [18] to independently combine feature vectors extracted from the accelerometer and acoustic sound sensor. In this case, each of the accelerometer and sound sensor was trained with Random forest as the base classifier and the predicted labels of each sensor were combined using Random Forest Meta-classifier. They noted that the stacked generalization fusion approach helps to preserve the statistical characteristics of each sensor thereby enhancing the performance accuracy and reliabilities of the activity recognition system. However, acoustic sensor data are ineffective in recognition of different activity details such as ambulatory activities, health monitoring and assisted living for the elderly and fall detection. Others include fitness tracking for effective living, postural identification and mobility changes [4, 9, 60]. Identification of these activities provides comprehensive health condition and well-being in people with special need and their health status for caregivers. The under-listed activities are accurately and effectively performed using motion sensors that involve whole body motion or local interactions with the sensor attached to the objects. Therefore, the acoustic sound sensor is only deployed for differentiating between indoor and outdoor activities [47].

Second, this present paper differs distinctly from previous studies by proposing an innovative and unified framework for the evaluation of feature-level and multiple classifier systems for human activity recognition. The paper comprehensively evaluates the impact of class imbalance issues and meta-heuristic feature selection approach for selection of relevant and compact feature vectors to enhance human activity recognition framework and reduce computation time. Third, the paper implement diversity based multi-view stacking ensemble algorithm to improve human activity detection by the integration of different classification models at both base classifiers and meta-classifier level. The use of different classification models for multi-view stacking ensemble algorithm provides flexibility, enhanced performance generalization, robustness, reduce uncertainty and ambiguity by classifier-level fusion of outputs generated by various classification model. Finally, the paper provides an extensive evaluation of the proposed methods using challenging motions sensor data and the performance results indicated improvement over baseline methods. The summary of recent studies on human activity recognition using data fusion is presented in Table 1.

Problem formulation

This paper aims to investigate how to improve the performance of human activity detection algorithm through a multi-view ensemble approach. To achieve that, the paper extensively investigated two fusion methods for human activity recognition. The methods investigated are feature-level fusion and multiple classifier systems as discussed in “Review of related works” section. In feature level fusion, features extracted from acceleration (acc_i), gyroscope (gyr_i) and magnetometer (mag_i) sensors were concatenated and trained with a single classification algorithm. The input to the activity detection framework are feature vectors extracted from each sensor modality. The concatenated features F_i and activity classes b_i are represented in the expression in Eq. (1).

$$F_i = (acc_i \oplus gyr_i \oplus mag_i, b_i) \quad (1)$$

Then, the classification algorithm is used to map the training feature vectors to the activity classes and this process is shown in Eq. 2,

$$M : F_i \rightarrow b_i, \quad (2)$$

where M is the classification algorithm and b_i is activity classes.

However, the use of a single machine-learning algorithm on concatenated features would fail provide an efficient and robust activity detection framework. Therefore, the paper further proposed a second evaluation method termed multi-view stacking for human activity recognition. This multiple classifier system (ensemble algorithms) methods train the feature vectors extracted from the individual sensor (*accelerometer, gyroscope, and magnetometer*) using different classification algorithms and then fused the intermediate output at the classifier level. Given the motion sensor data, $X_s = (F, b)$ that represent the training feature vectors generated from the motion sensor at each time windows. Where F is the feature vectors and b is the activity classes.

Here, the motion sensor represents the accelerometer, gyroscope, and magnetometer sensor data described earlier. The aim is to develop an innovative evaluation procedure to build adaptive decision fusion for human activity detection. Following the same approach as depicted in [59], we detail the problem formulation as follows.

$$F = (acc_i, gyr_i, mag_i, b_i), \quad (3)$$

represent the accelerometer, gyroscope, and magnetometer while b_i representing the activity details (Eq. 3). The training data (features from each sensor), X_s are trained with M number of base classifiers, where $M = \{m_1, \dots, m_n\}$ represents individual classifiers. We have two or three base classifier combination in three implementation procedures. The predicted class label after training the first-level classifier is shown in Eq. (4)

$$b \leftarrow \arg \max M_n(x), \quad (4)$$

where $M_n(x)$ is the prediction probabilities returned by each base classifiers M_n when input X_s is trained with an activity label b .

Then, the output prediction probabilities generated by each base classifiers are then combined with the meta-classifier as shown in Eq. (5).

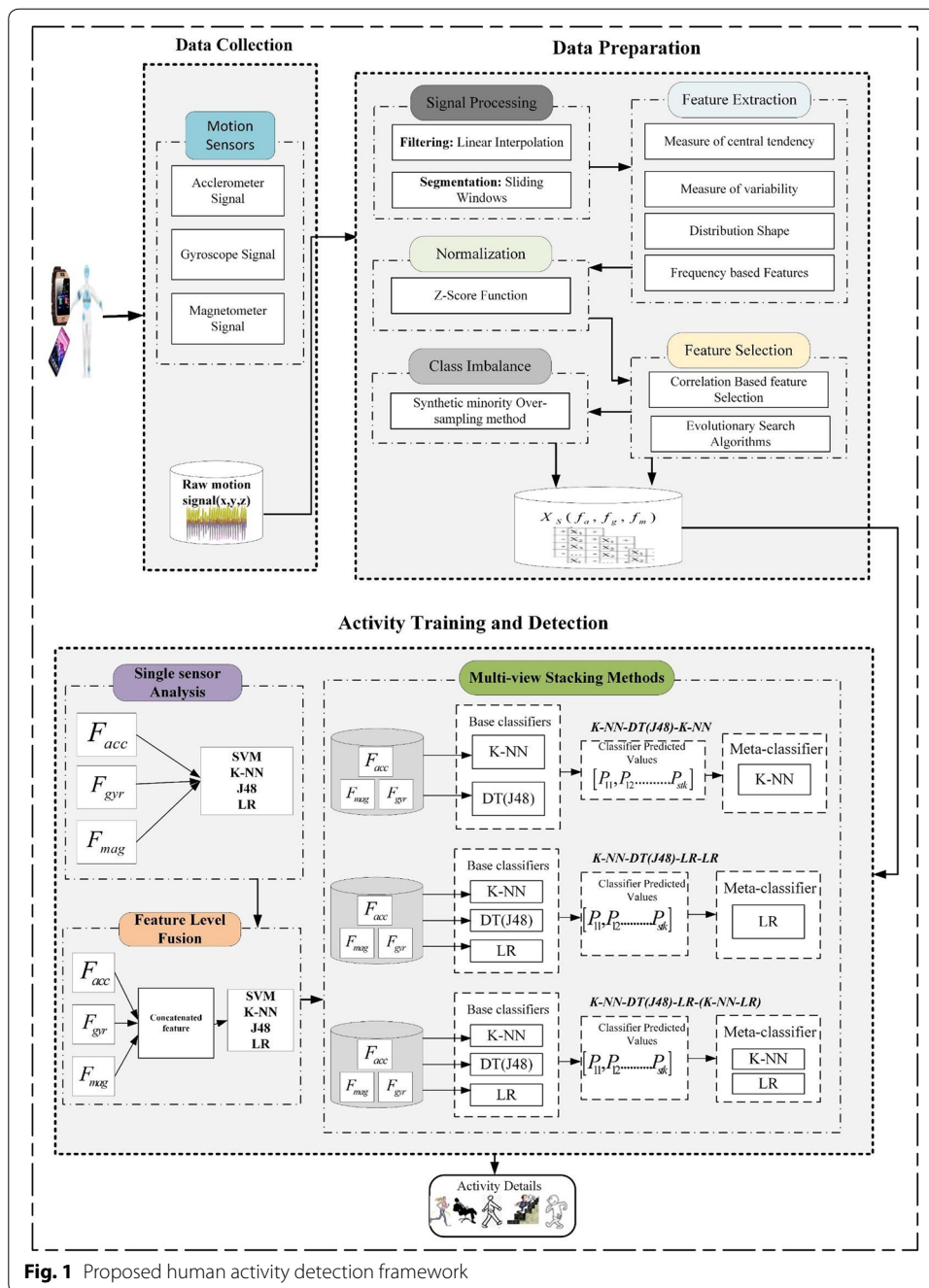
$$MSC : P \rightarrow b, \quad (5)$$

where $P = \{M_1(acc), M_2(gyr), M_3(mag)\}$ and MSC is the multiple classifier systems. "Decision fusion using multi-view stacking ensemble method" section presents further explanation of procedures for training the multi-view stacking method for human activity detection.

Proposed methodology

The proposed multiple sensor modalities fusion for human activity identification consist of six steps as depicted in Fig. 1.

These steps include data collection, signal processing, feature extraction and normalization, feature selection and classification of physical activity details. The experimental evaluation steps consist of (1) single sensor analysis, (2) sensor fusion using feature concatenation and (3) multi-view stacking that combine the predictive probabilities of different sensor modalities before fusion. In addition, the study evaluates the impact of



class imbalance by using Synthetic minority over-sampling techniques (SMOTE) to balance the activity classes with less number of instances.

The first step in human activity detection is data collection (Fig. 1). Here, two datasets collected with wearable devices named *Dataset 1* and *Dataset 2* were used in this study. The datasets contain different data of various modalities and include accelerometer, gyroscope, and magnetometer collected at a frequency of 204.8 Hz and 50 Hz respectively. Sensor data collected with mobile and wearable sensor are corrupted with sensor impurities due to signal degradation, therefore the signal processing help to remove

noise before feature extraction as shown in the step. The third step of the process is feature extraction.

Feature extraction is the most important aspect of human activity recognition as the process helps to transform the raw signal into descriptive feature vectors. Here, different features broadly categorized into time and frequency domain were extracted from the raw sensor data. These features were then normalized to limits the features to certain ranges and that have shown to enhance classifiers' performances. "[Feature extraction and normalization](#)" section provide further explanation on the feature extraction process.

Furthermore, most of the features extracted may not contribute positively for activity identification; the paper proposes a combination of bio-inspired metaheuristic feature selection aided by correlation based features selection to select the most discriminant features before activity classification. "[Feature selection](#)" section provide full description of feature selection processes adopted. The final steps depicted in the proposed method are single sensor analysis, feature-level fusion and multi-view stacking ensemble approach. In single sensor analysis, feature vectors extracted from each sensor modality and placement is fed to classification algorithm to build model for activity detection. On the other hand, feature-level fusion stage integrate feature vectors extracted from the sensor modalities before activity detection using machine-learning algorithms. Finally, multi-view stacking approach integrate decision from different classification algorithms and sensor modality. "[Decision fusion using multi-view stacking ensemble method](#)" section describes the multi-view stacking ensemble algorithms. We applied these steps simultaneously to all the positions utilized in experimental settings. The positions considered in this study include the ankle, chest, and wrist.

In Fig. 1, these steps are depicted in details.

Signal processing

Wearable and mobile inertial sensors (accelerometer, gyroscope, and magnetometer) based human activity classification requires sequence of procedures to process the sensor data before actual activity classification using different machine learning algorithms. Raw sensor data are corrupted by signal artifacts such as noise, missing values due to signal degradation or loss of battery life. We converted the raw sensor data, made of three axes (x , y , z) to time series and then filtered to remove noise. Filtering is important in human activity classification framework in order to remove low-frequency data, the geometric bias of sensor dimension, improve correlation and linear relationship between each data point [16]. In this study, linear interpolation was used to input missing values and the values at the end of the activity sequence were replaced with the previous activity data. This approach is very effective data transformation and signal processing method in human activity recognition system [11, 19]. To reduce computation time and accurately recognize activity details, data segmentation approach was applied on the raw sensor data to divide the data into a series of segments. The sliding window approach is considered in this study for its effectiveness in human activity detection [61].

The most important consideration in the sliding window approach is how to set the window size and this has proven to be important in recognition of certain activity details [62]. Here, the paper empirically set the window size and utilize previously tested

window sizes to enable accurate comparison with other published works in human activity recognition. For Dataset 1, we adopted the procedure recommended in the original research [63] to segments the data using 5 s (1024 samples with 204.8 Hz sampling rate) with 50% overlapping at each window size. While Dataset 2, a window size of 2-s (100 samples with 50 Hz sampling rate) was used as recommended in [19] without overlapping to accurately capture the activity details. We choose the small window sizes because most of the activities involved in the study are ambulatory activities that require small window sizes to recognize. These include activities such as walking, descending stairs, running or jogging [42, 64]. Each of the sensor modality (accelerometer, gyroscope, and magnetometer) used in this study were separately processed by applying the linear interpolation function $L()$ and data segmentation function $Sg()$ developed and then saved for feature extraction.

Feature extraction and normalization

Feature extraction process reduces the signal into feature vectors that are discriminative enough to describe the activity details, minimize activity classification error and reduce computation time. In human activity classification, several features have been proposed for human activity detection which can be broadly classified into time and frequency domain features [65]. Time domain features involve the extraction of signal or statistical metrics from raw signals and show how signal changes with time. The main advantages of time domain features are their ability to provide low computational time and are applicable for online and real-time activity detection. In contrast, frequency domain features help to show the distribution of signal energy and are efficient for recognition of repetitive activities [6, 9]. Therefore, given window sizes of d_t seconds ($N = f_s \times d_t$ samples), we extracted different feature vectors to characterize the original signal and present a compact representation of the activities performed at each window samples. Here, f_s represent the sampling frequency of each inertial signal used in this study. For each 3-axis of the motion signal (accelerometer, gyroscope, and magnetometer), we extracted 18 feature vectors of both time and frequency domain listed in Table 2. Fifty-four features were extracted from each motion sensor. From Dataset 1, these features

Table 2 List of extracted features from each sensor modality

Feature	Formula	Feature	Formula
Mean (μ)	$\bar{s} = \frac{1}{N} \sum_{i=1}^N s_i$	Root mean square (R_{ms})	$rms = \sqrt{\frac{1}{n} \sum_{i=1}^N (s_i)^2}$
Median (M_e)	$median_i(s_i)$	Peak amplitude (P_a)	$\max(s_i) - \min(s_i)$
Maximum (M_a)	$\max_i(s_i)$	Pitch angle (P_k)	$\arctan\left(\frac{x_i}{\sqrt{y^2+x_i^2}}\right)$
Minimum (M_i)	$\min_i(s_i)$	Signal power (S_p)	$\sum_{i=1}^N s_i^2$
Harmonic mean (H_m)	$\frac{1}{N} \sum_{i=1}^N \frac{1}{s_i}$	Kurtosis (K_r)	$E[(s_i - \bar{s})^4] / E[(s_i - \bar{s})^2]^2$
Standard deviation (σ)	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})^2}$	Skewness (S_k)	$E\left[\left(\frac{s_i - \bar{s}}{\sigma}\right)^3\right]$
Variance (σ^2)	$\sigma^2 = \frac{\sum (s_i - \bar{s})^2}{N}$	Energy (E)	$\frac{\sum_{i=1}^N s_i ^2}{length(s_i)}$
Coefficient of variation (C_v)	$\frac{\sigma_{s_i}}{\mu_{s_i}}$	Entropy (H)	$-\frac{\sum_{i=1}^N S_i \log S_i }{length(S_i)}$
Interquartile range (I_r)	$Q_3(s_i) - Q_1(s_i)$	Mean frequency (μF)	$\sum_{i=1}^N (i s_i(F)) / \sum_{j=1}^N s_j(F)$

were extracted from accelerometer and gyroscope placed on the ankle, chest, and wrist. Therefore, we computed 324 feature vectors from *Dataset 1*. We extracted the same set of features from *Dataset 2* (accelerometer, gyroscope, and magnetometer) placed at the ankle and wrist. In total, 324 feature vectors were extracted from *Dataset 2*. The feature sets extracted have been subdivided into different categories [66] as explained below.

- a. Time domain features: Time domain features extract statistical or mathematical quantities from raw signals in order to depict signal characteristics. The time domain feature vectors extracted from raw sensor data are a measure of central tendencies, degree of variation and distribution of signal shape.
 - i. *Measure of central tendency* These features describe how the signal is close to the central values and depict the location of the central points. Features extracted as measure of central tendency include mean, maximum, minimum, median and harmonic mean of each 3-axis of the raw signal. These features require less computation time to process with minimal computational requirement. Furthermore, measure of central tendency based features have shown significant performance improvement for posture recognition, differentiation of static and dynamic activities and energy expenditure estimation [3, 65].
 - ii. *Measure of variability* These feature set represent degree at which the motion sensor signals are distributed over a distances between the central points. The higher the degree of variation, the worse the distribution of the raw signals. The feature extracted as measure of variabilities are standard deviation, variance, coefficient of variation, interquartile range, root mean square, signal magnitude area, magnitude of area under the curve, pitch angle, signal power and peak amplitude. Measure of variability based features have low computation cost and important to determine the stability and probability distribution of raw signals [65, 66].
 - iii. *Distribution of shape* Signal based feature helps to understand the shape and distribution of the raw sensor signal. The distribution of shape-based features computed from the raw sensor signal includes skewness and kurtosis. Skewness measures the asymmetric probability distribution of the signals while kurtosis determines the flat or spike of the sensor data distribution. Distribution of shape-based features have extensively shown impressive results in human activity detection, health monitoring and related applications [3, 19, 66].
- b. Frequency-based features: Frequency based features are important for the analysis of repetitive activities. In addition, they are required in human activity recognition. Here, the raw signal data were transformed into the frequency domain using the Fast Fourier Transform (FFT) function. From the frequency domain data, we extracted different feature vectors. Feature vectors extracted from the transformed data include energy, entropy and weighted mean frequency of the transformed signal data. Entropy provides a means to differentiate between signals with the same energy but corresponds to different activities. These features were extracted following the procedure explained in [19]. Moreover, the features have been previously used for

the detection of activities such as cycling, jogging, running and so on [65]. All the feature vectors extracted from both *Dataset 1* and *Dataset 2* are listed in Table 2 with their corresponding formula used to calculate each feature.

After the feature extraction process, the computed features were then normalized to zero mean and unit variance in order to reduce the features to certain ranges. This process has proven to be effective for improved classification accuracy in the human activity detection system especially for features with dynamic range. The Z-score normalization procedure [26, 67] was utilized in this study. The mean value of the feature vectors was subtracted from the individual signal data point and divided by the standard deviation. The Z-score normalization is shown in Eq. (6) below, where \bar{x} and σ represent the mean and standard deviation of each data point respectively.

$$x' = \frac{x_i - \bar{x}}{\sigma} \quad (6)$$

Feature selection

However, not all the extracted feature vectors from each sensor modality used in this study (accelerometer, gyroscope, and magnetometer) shown in Table 2 may be useful for developing effective and efficient activity detection framework. The use of unnecessary feature would lead to overfitting, low performance and high computation time [3, 26]. Therefore, we deployed feature selection methods to reduce the features and increase performances. In this case, optimal features vectors were selected by utilizing different feature selection methods [68]. However, choosing the best feature selection methods to reduce the dimensionality of the data is still challenging as different feature selection approach works differently in different training data. For this reason, we created an intersection of two feature selection methods. Here, the study proposes bio-inspired meta-heuristic (evolutionary search algorithm) integrated with correlation-based feature selection to select the most appropriate feature vectors and reduce redundancy and computation time. The evolutionary search algorithm is a meta-heuristic and wrapper-based feature selection method while the correlation-based feature is a filter based method. The combination of these two feature selection methods would help to select the most discriminative feature vectors. First, we applied correlation based features selection on the training data to estimate the correlation between each class and feature vectors. The features were ranked based on the correlation and features with correlation threshold of >0.15 were selected to be used for further dimensionality reduction using bio-inspired metaheuristic feature selection approach. Correlation-based feature selection approach is fast and efficient to select discriminant feature vectors and to effectively rank the computed features. To further improve the features and classification results, correlation-based features are integrated with Bio-inspired feature selection (evolutionary search algorithm) to further reduce the dimensionality of the data. Using correlation based features with selection criteria, we selected some feature vectors, but these feature vectors may not be representative enough to ensure accurate activity classification. On the other hand, an evolutionary search algorithm is a form of the meta-heuristic search algorithm that simultaneously explores several points in search space and navigates the search space stochastically to avoid being trapped in local minima. Therefore, the

evolutionary algorithm maintains an individual population over the training instances and time as depicted in Eq. (7)

$$p(t) = \{x_i^t, \dots, x_n^t\}, \quad (7)$$

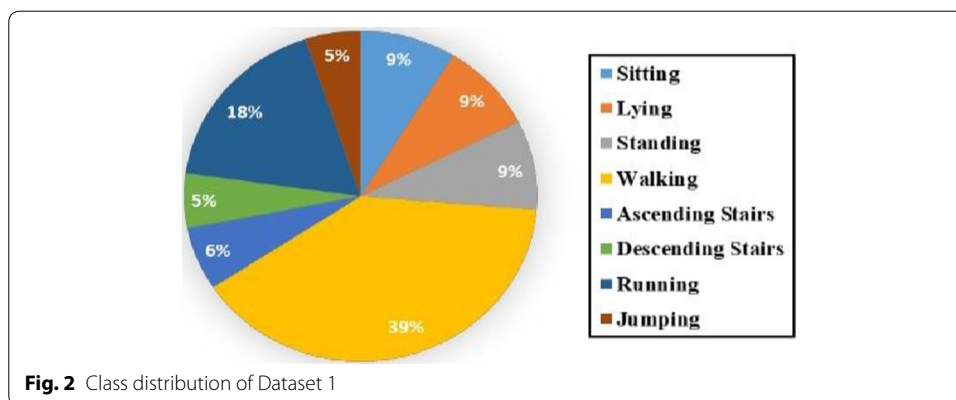
where x represent the potential to the problem at hand. Evolutionary search algorithm exploits biologically inspired mechanism inform of recombination, mutation, fitness, and selection to iteratively select the best training instances for activity classification and health monitoring [69, 70]. The processes for developing an evolutionary algorithm for feature selection are described below:

1. For initial set value $t = 0$, construct new population values $p(t) = \{x_i^t, \dots, x_n^t\}$ that represent a set of starting point to explore evaluation instances;
2. Evaluate each selected features to give their abilities to predict the intended target which is some measure of its fitness;
3. Select a new population over training instances $p(t + 1)$ by stochastically selecting the individual from the initial population $p(t)$;
4. Some of the new population are further transformed by means of genetic operators to form new solutions.
5. Step 1 to 4 are recursively repeated until certain criteria are reached. These criteria include reaching the number of iteration, given fitness score is achieved or the evolutionary algorithm converges to near-optimal solutions. The termination criteria depend on the one that is achieved first.

The parameters used in choosing the best feature sets for human activity classification using an evolutionary algorithm are default parameters in WEKA machine learning toolkits to ensure reproducibility. The list of the selected features using evolutionary algorithm (EA) and correlation-based feature selection methods are shown in [Appendix](#). The selected feature vectors were used to develop the proposed feature-level and decision-level fusion through multi-view stacking.

Activity class imbalanced distribution

In various real-world applications such as medical diagnosis, fraud detection, activity detection, and health status monitoring, data are expected to be imbalanced. In this case, a certain activity that is frequently performed in nature has majority classes while less performed activities have less number of classes. Such imbalanced class distribution tends to overwhelm the minority classes and produce accuracy that is skewed towards the majority classes [6]. Class imbalanced occur when there is a high variance between the majority activity classes and minority activity classes. Therefore, balancing the training data would improve the classification algorithm performances. Two methods commonly used for solving the class imbalanced problem are the Random under-sampling (RUS) and Synthetic Minority Oversampling Techniques (SMOTE) [71]. Random under-sampling methods reduce the majority classes to equal the number of minority classes. However, it may lead to loss of important information when Random under-sampling is applied to balance the dataset. On the other hand, SMOTE generates new training data from the nearest neighbor of the line joining the minority classes. The approach augments the training dataset of the minority classes by randomly generating new instances



from the original training data and avoid over-fitting. Here, we utilize the SMOTE [25] to increase the minority activity classes following a recent study in human activity recognition [6]. We oversampled the minority classes such as Descending stairs, Jumping and descending stairs to solve the problem related to imbalanced dataset [71]. Figure 2 shows the activity class imbalanced distribution in *Dataset 1*.

To access the impact of class imbalance on the dataset and proposed method, area under the curve (AUC) that performance was introduced. Area under the curve is more robust than other performance metrics such as accuracy, recall and F-measure in class imbalance scenario. Moreover, AUC is independent of data skewness and class activity distributions. In addition, area under the curve has been used to assess the impact of class distribution in human activity identification and related applications [6, 71].

Classification algorithms

In human activity classification, it is difficult to build accurate and effective decision fusion using the single classification model. Therefore, this paper first evaluates the individual classifier on the sensor modalities used in our study and then, the same classifiers were used to build the feature-level fusion and multi-view stacking ensemble methods. Moreover, it is challenging to identify all the activity details using a single classification algorithm. To this end, classification algorithms that have been used for human activity detection to ensure accurate evaluation and comparison were selected. The classification algorithms used in this study include the decision tree (J48), support vector machine, k-Nearest Neighbors (k-NN) and logistic regression [3, 16]. These classification algorithms were chosen because they have produced improved performance results in human activity detection and similar tasks [3, 16, 24]. Moreover, each classification algorithm has different philosophy implementation for the learning process and selecting appropriate classification algorithm is important for building robust and efficient human activity detection system. In subsequent sections, these classification algorithms are discussed and their parameter tuning values are presented in Table 5, “Experimental setup” section.

Decision tree (J48)

Decision tree provide learning and classification algorithm that recursively divide training data or features from sensor modalities into node segments made up of root node, internal splits, and leaves [72]. A decision tree is a non-parametric algorithm that does

not require assumption on the training feature distribution and can model the non-linear relationship between the training feature vectors and activity classes. Various classes of the decision tree have been extensively proposed for human activity detection in recent studies due to its algorithm efficiency and easy to understand process. In addition, the decision tree allows the easy and rule-based hierarchical representation of activity details [73, 74]. In this study, we implemented the J48 decision tree.

Support vector machine (SVM)

Support vector machine (SVM), developed by [75] provide powerful classification algorithms based on statistical learning theory and employ the use of hyperplane that separates the training data using maximal margin [76]. Given a training instance of feature vectors extracted from different sensor modalities $X_S\{x_1, x_2, \dots, x_S\}$, support vector machine optimally separate the classes into different activities. However, the support vector machine requires large training examples to avoid overfitting and extensive parameter tuning in order to obtain high-performance generalizations [42].

K-Nearest Neighbors (k-NN)

k-Nearest Neighbors is a non-parametric and lazy learning algorithm that use instance-learning methods to store instances and classifiers new training data using similarity index measure such as Euclidean distances. k-NN is one of the simplest and most effective algorithms for human activity recognition and has provided competitive performance and pattern recognition problems [3, 77]. Furthermore, k-NN is effective in handling large training features that are too large to fit into memory and use simple Euclidean distance to measure the similarities between training and testing feature vectors in human activity recognition [9, 42]. K-Nearest Neighbors learning activity patterns from training features vectors by comparing the similarity between specific test data sample with a set of training instances based on the closest neighbor k-values that shows the number of neighbors utilized to determine the classes. With the above process, k-Nearest Neighbors provide a faster and more accurate recommendation with desirable quality for activity recognition. In our implementation, the value of k was set to 10, although we tested other values of k , but $k=10$ provided the best result using our datasets.

Logistic regression (LR)

Logistic regression is fast, simple and compact classification model that has been extensively applied for human activity detection and health monitoring [16, 74]. Logistic regression provides an easy interpretation of the model and importance of feature vectors. Moreover, the model is easily parallelizable. In logistic regression model, the relationship between training data and activity detail are modeled in order to accurately detect activity classes. The input values are linearly combined using weights or coefficient values in order to make predictions on the training data. In addition, logistic regression model has provided an efficient algorithm for human activity classification in recent years [78].

Decision fusion using multi-view stacking ensemble method

Recently, data fusion and multiple classifier systems frameworks have been acknowledged as the most effective mechanism to enhance the reliabilities, robustness, and generalizability

of human activity identification systems. With a data fusion approach, issues bothering on data uncertainty, and the effect of direct captures that are challenging to eliminate with single sensor modalities are minimized [3]. As earlier pointed out, multi-sensor development of human activity identification can be achieved at three levels, which includes data fusion, feature level fusion, and multiple classifiers approaches. These fusion mechanisms provide means of integrating multiple sensor modalities for comprehensive human activity monitoring and other related applications. However, existing studies in multiple sensor fusion majorly explore feature level fusion that combines features extracted from multiple sensors and fed to a single machine learning algorithm for human activity detection. The main drawback of such method is the inability to learn sensor specific activities and statistical properties for effective activity monitoring. In addition, there is an issue bothering on feature incompatibility that highly decreases performance results [7, 79]. Therefore, it is challenging to understand the performances of each sensor modalities in human activities detection.

Furthermore, these methods are unable to handle a long sequence of complex activity details. In addition, feature incompatibility and signal variation greatly affect algorithm performance in the feature level fusion approach using single classification algorithms. In order to improve the performance results of human activity detection, this paper proposes decision fusion using multi-view stacking method. In decision fusion method, decisions produced by multiple classifiers are integrated to handle complex systems, high dimensional data and reduce uncertainty especially in heterogeneous sensor scenario [3]. Specifically, the paper proposes multi-sensor fusion through stacking ensemble of heterogeneous classifier fusion to improve human activity detection algorithms. Stacking ensemble (stacking generalization) first proposed in [22] is efficient multiple classifier system for decision fusion of data of different modalities to achieve diversity and reduce misclassification rate. The underlying concept of stacking is how to combine the decision obtain from heterogeneous or homogeneous classifiers with meta-classifiers to improve performance [80]. Therefore, the proposed multi-sensor fusion involves multiple stages that involve choosing the base classifiers, training the base classifiers on the training data and then Meta classifier to integrate the results of the base classifiers [24]. The base classifiers are set of classification algorithms, M_1, \dots, M_i and meta-classifier(s) M_{i+1} are trained on the prediction of the base classifiers. The rationale is to improve classification results by training the meta learner on the misclassification of the base classifiers.

Therefore, given S number of different sensor modalities in which feature vectors are extracted and represented as X_S , in this case from the accelerometer, gyroscope and magnetometer sensors. For each sensor modality in the training data, the training data is trained with K -fold cross-validation (in our case tenfold cross-validation was used) to generate the input training data for Meta classifiers. The base classifiers used in our case are the decision tree (J48), k -Nearest Neighbors (k -NN) and logistic regression (LR). Given the features vectors from each sensor where $x_a b_a$ which represent the attribute instances and class labels (activities) of each sensor modality. Then, by applying K -fold cross-validation, each sensor data is divided into nearly equal blocks and trained using one of the base classifiers M_1, \dots, M_i and the output concatenated to train the Meta-Learner M_{I+1} . Furthermore, randomly dividing the training data into K equal part gives rise to $X_{S1}, X_{S2}, \dots, X_{SK}$ and for each K -fold cross-validation, the base classifiers (k -NN, decision tree and logistic regression) were trained on $X_S^{(-k)} = X_S - X_{Sk}$ and tested on X_{Sk} respectively [81].

The output predicted values of each K-fold P_{Stk} comprised of the output predicted probabilities of each class and predicted class label of each base classifier is then generated and trained with Meta classifier(s) M_{i+1} . With training and testing for K-fold cross-validation, the prediction outputs for all sensor modalities are pooled together into $[P_{Stk} = P_{A11}, \dots, P_{Stk}]_b$ where t represent the classifier index, k is the test part of the K-fold cross-validation and b is the total number of classes (in our case, eight activity classes). The pooled output predicted probabilities along with the predicted labels and true class labels for all the sensor modalities are represented as $[P_{S_{ik}} = P_{A11}, \dots, P_{S_{ik}}, X_s, y_a]_b$.

During the testing stage, given the feature vector from each sensor modalities, the same procedures were followed to generate the output predicted probabilities using the base classifiers. Then, the output predicted probabilities are then fed to the Meta classifier to produce the final activity detection results. The multi-view stacking process is shown in Algorithm 1. The major issues in developing adaptive stacking ensemble method are the choice of base classifiers and Meta classifier that would provide the best empirical results [82]. Wolpert [22] noted that there is no specific base classifier and Meta classifier for building score based stacking method [24]. Other studies have also empirically evaluated the choice of base classifiers and Meta classifier in multi-view stacking implementations. In a recent study, Ahmed et al. [24] observed that score based stacking ensemble performed excellently well with a combination of the heterogeneous classifier at both base classifier and meta-classifier level. To ensure effective implementation, we propose a heterogeneous classifier based multi-view stacking method for human activity identification. Here, three stacking methods were constructed to under-study their impact on human activity detection system.

Algorithm 1: Multi-sensor fusion through Multi-view stacking ensemble algorithms

Input: Set of features extracted from different sensor modalities (accelerometer, gyroscope, magnetometer) represented as X_S

Output: Set of activities details

Given a set of feature vectors extracted from different sensor modalities.

1. The feature vectors from each sensor $X_S = \{(x_a, y_a), a=1, \dots, N\}$ were randomly divided into the nearly equal block and trained with base classifiers (in this case, Decision tree (J48), k-Nearest Neighbors, Logistic Regression) to generate the combined feature vectors to train the Meta-classifier M_{I+1} .
2. Randomly dividing the training data into K equal part given rise to $X_{S1}, X_{S2}, \dots, X_{SK}$ and for each K-fold cross-validation, Decision tree (J48), k-Nearest Neighbors and Logistic Regression were trained on $X_S^{(-k)} = X_S - X_{Sk}$ and tested on X_{Sk} respectively;
3. The output predicted the probability of each class and predicted the label of each base classifier is then generated and trained on Meta-classifier M_{i+1} . Here, the meta-classifiers include the k-NN, Logistic Regression and Majority Voting
4. The output prediction of all the sensor modalities is pooled together into $[P_{Stk} = P_{A11}, \dots, P_{Stk}]_b$ where t , k and b represent the classifier index, k-fold cross-validation and a total number of classes respectively.
5. All the pooled output predicted probabilities along with their predicted class label and true class labels for the sensor modalities is represented as $[P_{S_{ik}} = P_{A11}, \dots, P_{S_{ik}}, X_s, y_a]_b$.
6. The final multi-sensor fusion technique is built by retraining the entire data of the feature vectors from all the sensor with k-NN, Logistic Regression, and average (k-NN and Logistic Regression).
7. For testing stages, given features vectors generated from the sensor modalities, then the output predicted probabilities of testing data are generated with k-NN, Decision Tree and Logistic Regression as the feature vectors for Meta-level classifier.

The output predicted probabilities features is then classified by k-NN, Logistic Regression and average results of k-NN and LR to produce the final activity detection results.

These include:

- Multi-view stacking with k-NN, J48, and k-NN (MST-k-NN-J48-k-NN): k-NN and decision tree were used as based base classifiers and k-NN as the Meta classifier;
- Multi-view stacking with LR, k-NN, and J48 (MST-LR-k-NN-J48-LR): LR, k-NN, J48 were used as base classifiers and LR as the Meta classifier.
- Multi-view stacking with LR, k-NN, J48 and (k-NN-LR): (MST-LR-k-NN-J48 (k-NN-LR)): LR, k-NN, and J48 as base classifiers and LR and k-NN as the Meta classifier. The final result is the average of the prediction of the two classification algorithms.

In each experimental configuration, tenfold cross-validation was used on the training data to generate the meta-level classifier data. This approach is important to avoid overfitting the training data [18]. Algorithm 1 shows the generic process for training the stacking generalization approach adopted in our experiments.

Experiments

This section presents the experimental implementation of feature-level fusion and multi-sensor fusion using multi-view stacking methods for human activity recognition. Furthermore, the impact of increasing the minority class by apply Synthetic Minority Over-sampling (SMOTE) is also evaluated. Moreover, the section presents the datasets, model validation, experimental setup, and performance evaluation.

Datasets description

Dataset 1: This dataset was first collected and analyzed in [63] for hierarchical multi-sensor based classification of activity of daily living using Shimmer sensor nodes placed on the right ankle, chest, right hip and right wrist. All the sensors used in the experiment were a motion sensor that records 3D accelerometer and 3D gyroscopes while 19 subjects perform 13 activities of daily living. These activities include sitting, lying, standing, washing dishes, vacuuming, sweeping, walking, ascending stairs, descending stairs, treadmill running, bicycling on the ergometer (50 w), bicycling on the ergometer (100 w) and rope jumping recorded at a sampling rate of 204.8 Hz. The range of the accelerometer used in the experiment was ± 6 g. The range of gyroscope was ± 500 °/s for sensor nodes placed on the chest, hip and wrist and ± 2000 °/s for sensor nodes on the hip. Our study utilizes the sensor placed at the ankle, chest and wrist and subsets of the activities for evaluation of our proposed multi-sensor fusion based on multi-view stacking. The subset of activities used includes sitting, lying, standing, walking, ascending stairs, descending stairs and jumping using a segmentation window size of 5 s (1024 samples for each sensor data) with 50% overlapping at adjacent windows. This was the original window size segmentation used in the evaluation of the data in hierarchical multi-sensor based classification of activity of daily living [63] with a mean classification rate of 89.6%. Details of the two datasets used in our experimental evaluation are presented in Table 3.

Dataset 2: Dataset 2 was collected in [83, 84] as a benchmarked dataset for implementation of an open framework for agile development of mobile health applications. The dataset provides a framework to build tools that support multidisciplinary mobile health applications. Moreover, the dataset provides multimodal human activity data comprises

Table 3 Summary of datasets used and data processing methods

	Dataset 1	Dataset 2
Sensors	Shimmer sensor devices containing 3D accelerometer, 3D gyroscope	Shimmer2 sensor device IMU containing 3D accelerometer, 3D gyroscope, 3D magnetometer and 2-lead electrocardiography (ECG)
Placement	Right ankle, chest and right wrist	Ankle and wrist
Physical activities performed	sitting, lying, standing, washing dishes, vacuuming, sweeping, walking, ascending stairs, descending stairs, running, bicycling on the ergometer (50 w), bicycling on the ergometer (100 w), rope jumping.	Standing still, sitting and relaxing, lying down, walking, climbing stairs, waist bend forward, the frontal elevation of arms, knee bending(crouching), cycling, jogging, running, jumping
Number of activities	13	12
Number of participants	19	10
Sampling rate	204.8 Hz	50 Hz
Filtering method	Linear interpolation	Linear interpolation
Window type and size	5 s with 50% overlap	2 s
Feature selection methods	Evolutionary search method, CorrelationAttributeEval, Ranker	Evolutionary search method, CorrelationAttributeEval, Ranker
Evaluation method	Tenfold cross validation	Tenfold cross validation

of accelerometer, gyroscope and vital sign data collected from ten subjects while performing 12 physical activities while Shimmer2 sensor device was used to record the data from each subject in the experiment. During data collection, the motion sensors were placed on the subject's left ankle, right wrist, and chest. The sensor placed on the chest also provided 2-lead electrocardiography (ECG) measurement for health monitoring to ascertain the effect of exercise on the participants. The sampling rate used for sensing activities was 50 Hz in which the sensor monitor 3D acceleration (± 6 g) placed on the ankle, chest, and wrist, 2-lead electrocardiography (ECG), 3D gyroscope placed on the ankle, chest, and wrist, and 3D Magnetometer on the ankle and wrist. The subjects wore the sensor while performing activities that included standing still, sitting and relaxing, lying down, walking, climbing stairs, and waist bends forward, frontal elevation of arms, knees bending (crouching), cycling, jogging, running and jumping front and back. These activities were performed for 1 min range and some were performed $20\times$. In this paper, we analyze accelerometer, gyroscope and magnetometer sensors placed at the ankle and wrist for multimodal data fusion approach. However, 2-lead ECG measurement is not analyzed in our study in line with the original experiments. In addition, subsets of the data sets were considered in our experiments, these include Standing still, sitting and relaxing, lying down, walking, climbing stairs, cycling, jogging and running.

Experimental setup

The signal preprocessing, data segmentation, features extraction and normalization discussed previously were implemented in MATLAB 2016 (reference <https://in.mathworks.com/>) for each activity sequence and combined into master feature vectors (MFVs). The extracted master feature vectors were stored as .csv files and converted to attribute relation file format (.arff). Then, the dimension of the master feature vectors was reduced to select the most discriminant features by applying the correlation-based features and followed by Evolutionary search algorithm to further reduce the feature vectors. The

Table 4 Classification algorithms and parameter values

Classification algorithm	Parameters
Support vector machine	batchSize=100;buildCalibrationModels=False;c=1.0;calibrator=Logistic-R 1.0E-8-M-1-num-decimal-places 4;checksTurnedOff=False; debug=False;doNotCheckCapabilities=False;epsilon=1.0E-12;filterType=Normalize;kernel=PolyKernel-E 1.0-C 250007; numDecimalPlaces=2;numFolds=-1;randomSeed=1;toleranceParameter=0.001
k-Nearest Neighbors	KNN=10;batchSize=100;crossValidate=False;debug=False;distanceWeighing=No distance weighing;doNotCheckCapabilities=False;meanSquared=false;nearestNeighbourSearchAlgorithm=LinearNNSearch;numDecimalPlaces=2;windowSize=0 Standard
Decision tree (J48)	batchsize=100;binarysplits=false;collapseTree=True;confidenceFactor=0.25; debug=false;doNotCheckCapabilities=false;doNotMakeSplitPointActualValue=false; minNumObj=2;numDecimalPlaces=2;numFolds=3;reduceErrorPruning=false;saveInstanceData=false; seed=1;subtreeRaising=True; unpruned=false;useLaplace=false;useMDLcorrection=true
Logistic regression	batchSize=100;debug=false;doNotCheckCapabilities=false;maxit=-1;numDecimalPlaces=4; ridge=1.0E-8;useConjugateGradientDescent=false

proposed method utilized the WEKA (Waikato Environment for Knowledge Analysis) machine learning toolkits [85] as an implementation platform for feature selection and classification. The classification algorithms selected for both implementation of feature-level fusion and multi-view stacking method include logistic regression (LR), Sequential minimal optimization (SMO), k-Nearest Neighbors and decision tree (J48). The classification algorithm was selected based on their performance in similar human activity classification evaluations [86].

Furthermore, the parameter values used in all our experiments are shown in Table 4. These parameter settings are default values and were chosen based on recent empirical evaluations of various classification algorithms in pattern recognition [87] and reported improved performance results. In addition, these values are default values to ensure reproducibility. We use the same parameters throughout the experiments to evaluate each activity detection method, both for a single sensor, feature fusion, and decision fusion methods. The implementation of the signal processing, feature extraction, and classification algorithms was conducted on system computer running on Windows 10 Operating system. The system is using an Intel Core™ I7-6700 CPU @ 3.400 GHz with installed Random Access Memory (RAM) capacity of 16 GB.

Model validation

In general, approaches such as hold-out, leave-one-out and k-fold cross-validation methods are used to validate model performances for human activity detection and health monitoring [68]. Each validation methods depends on the task and the size of the training dataset. In holdout cross validation partitions, the datasets are divided into training and testing data, and then the training set is used to train the algorithms and evaluated on the test set. In contrast, leave-one-subject-out cross-validation segments the training data into the number of subjects used for the experiment, then, use one subject for training and the resting for testing. This approach is repeated for each subject in the experiment, thereby using the entire subject for training and testing. Likewise, stratified K -fold cross-validation method divided the whole training data into K equal part, use $K - 1$ parts of the data for training classification model and K part for testing. This

procedure is repeated K times and the final result is computed as the average of all the tests performance of all the folds [6, 9].

The main advantage of K-fold cross-validation is that all instances of the training dataset are trained and tested with the model thereby providing lower variance within the estimator. The method ensures accurate prediction with less bias of the true rate estimator and important for model selection [62]. In this study, we applied tenfold stratified cross-validation for building the multi-view stacking ensemble fusion method and the feature-level fusion used in the experiments. As earlier outlined, the use of tenfold cross validation provides a means to avoid overfitting [18]. To measure the statistical significance of the proposed decision fusion method, confidence interval approach [67, 88] was adopted. The performance metrics are considered statistically significant if the differences between decision fusion, feature fusion, and single sensor evaluation are higher than the confidence interval. This is considered statistically significant with a 95% probability. The equation for computing the confidence interval is shown in Eq. (8), where P is the performance metrics and N is the number of instances.

$$\delta = \pm 1.96 \times \sqrt{\frac{P(100 - P)}{N}} \quad (8)$$

Performance evaluation

The performance of individual classification algorithm and multi-view stacking fusion methods were evaluated using different performance metrics. These performance metrics include accuracy, recall, precision, error rate and area under the curve (AUC). For each activity class, the prediction was measured with the ground truth labels and the number of true-positive (TP), true negative (TN), false-positive (FP) and false-negative (FN) were calculated using the Confusion matrix of each prediction. These performance measures are shown in Table 5 with the corresponding measurement equations. In addition, these performance metrics have been extensively applied in the evaluation of human activity detection systems and related applications [68]. All the performance metrics were computed based on individual class represented as N . To compute the Area under the curve (AUC) to measure the ranking of each activity detection algorithms and impact of class imbalance, we adopted the proposed approach in [89] for multi-class activity detection and classification problem.

The area under the curve is deployed to evaluate the impact of class imbalance on the performance of human activity detection systems and was recently implemented in diabetes mellitus prediction [71].

Experimental results and discussions

We conduct different experiments to investigate the impact of different fusion of motion sensor generated by mobile and wearable sensor data for human activity detection. These experiments can be categorized into: *first*, baseline evaluation that present experiments of using features extracted from acceleration, gyroscope, and magnetometer without feature selection or over-sampling method. *Second*, present the results of applying feature selection method and fusion of both feature level fusion and multi-view stacking methods. *Third*, analysed the impact of the over-sampling method (SMOTE) on both feature-level fusion and use of multiple classifier systems approaches. *Finally*, compare

Table 5 Performance evaluation measures

	Evaluation measures	Equation	Description
Higher values indicate better performances	Accuracy	$\frac{1}{N} \sum_{i=1}^N \frac{(TP+TN)_i}{(TP+FP+TN+FN)_i}$	Calculate the rate of correctly classified activities classes out of the total number of activity instances
	Recall	$\frac{1}{N} \sum_{i=1}^N \frac{(TP)_i}{(TP+FN)_i}$	Measure the number of correctly predicted instances as positive instances
	Precision	$\frac{1}{N} \sum_{i=1}^N \frac{(TP)_i}{(TP+FP)_i}$	Measure the ability of the proposed algorithm to accurately classify actual activity details
	F-measure	$\frac{1}{N} \sum_{i=1}^N 2 \cdot \frac{(precision \cdot recall)_i}{(precision+recall)_i}$	Calculate the weighted harmonic mean of precision and recall
	Area under the curve (AUC)	$\frac{1}{N} \sum_{i=1}^N 0.5 * \left[\frac{TP_i}{(TP+FN)_i} + \frac{TN_i}{(TN+FP)_i} \right]$	Measure the rate of performance of the algorithms across all activity details. AUC is the plot between recall and specificity drawn from the different threshold
Lower values indicate higher performances	Error	$\frac{1}{N} \sum_{i=1}^N \frac{(FP+FN)_i}{(TP+FP+TN+FN)_i}$	Calculate the rate of activities incorrectly classified out of all the activity details

the multi-view stacking methods with other multiple classifier systems recently implemented for human activity classification. These experiments were analysed for *Dataset 1* and *Dataset 2* as described earlier.

Baseline method: analysis of individual motion sensor

The performances of the individual motion sensor using all the features extracted from *Dataset 1* and *Dataset 2* are presented in Figs. 3, 4. There are different areas in which the results are analysed. These include classification algorithms, sensor modality based and position-wise. The performance results are presented in terms of accuracy, recall, precision, F-measure and AUC for each analysis. The AUC is used to determine the impact of class imbalance on the performance of human activity detection systems [71]. In addition, these performance metrics were used to ensure a comprehensive evaluation of the proposed human activity detection framework.

The essence of this evaluation is to establish if there are irrelevant feature vectors in the extracted features, impact on each sensor modality on activity detection and as a baseline for evaluating the performances of feature-level and the multi-view stacking ensemble methods. *Dataset 1* contains two sensor modalities, accelerometer, and gyroscopes. As shown in Fig. 3 the acceleration sensor placed at the ankle present best performance when compared to the gyroscope sensors. The performance is closely followed by chest position while wrist presents the list performance. In terms of machine learning algorithms utilized in our experiments, J48 decision tree classifier demonstrated the best activity detection accuracy of 95.32%, followed by k-NN (94.78%) for acceleration sensor placed at the ankle. We obtained similar results with gyroscope sensor with accuracy for J48 decision tree (87.56%), k-NN (85.28%) and LR (84.20%). The least performance results were observed in support vector machine. Moreover, k-NN marginally

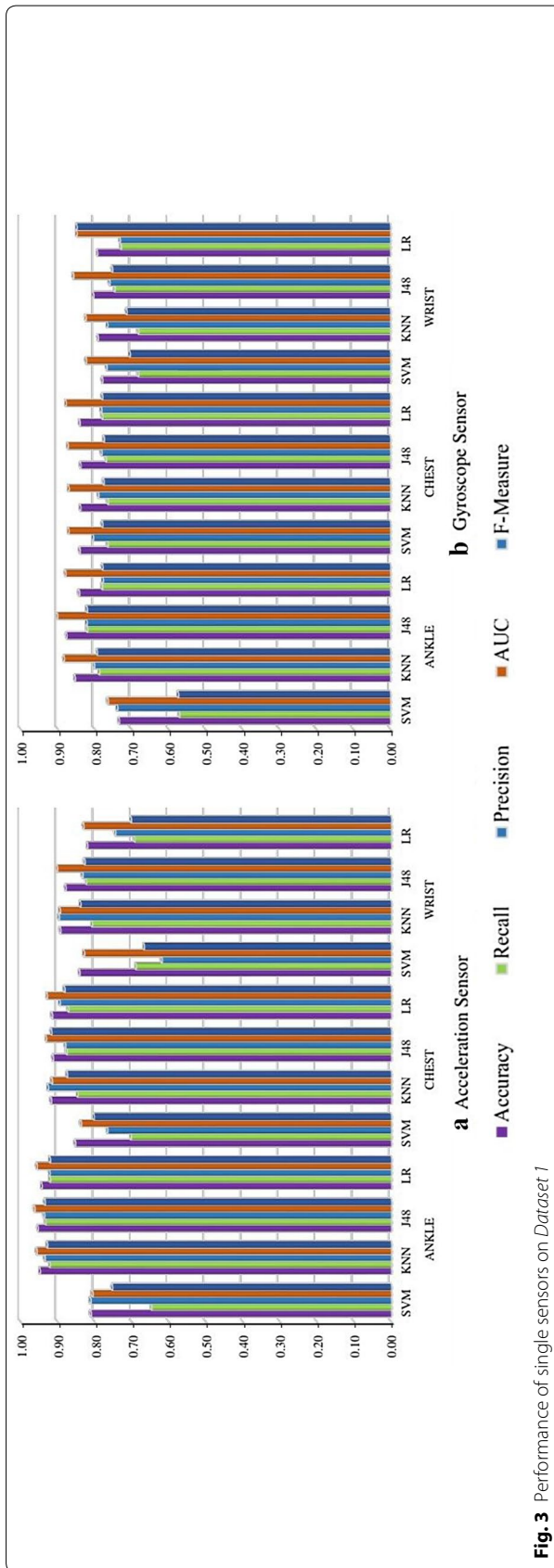


Fig. 3 Performance of single sensors on Dataset 1

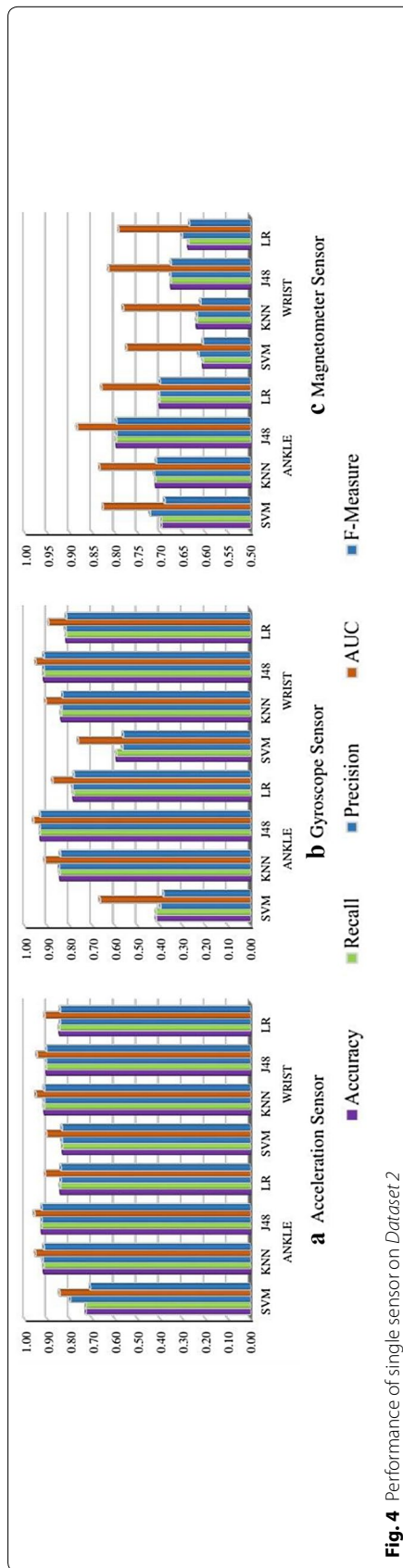


Fig. 4 Performance of single sensor on Dataset 2

outperformed decision tree and logistics regression for acceleration sensor placed on the chest while support vector machine outperformed k-NN and decision tree for gyroscope sensor placed in the same position.

Regarding *Dataset 2* (Fig. 4) that contain three sensor modalities: accelerometer, gyroscope, and magnetometer. We analysed the sensors placed on the ankle and chest in our experiments as these sensors contain all the sensor modalities. Here (Fig. 4), the acceleration sensor clearly outperformed gyroscope and magnetometer for all the sensor positions. The performance results showed that most of the machine learning algorithms used to evaluate the sensors performed very well.

The results demonstrated that the highest accuracy was achieved with J48 decision tree with 93.17%, 92.48% and 79.76% for gyroscope, acceleration and magnetometer sensor placed on the ankle respectively. These performance results were closely followed by k-NN with accuracies of 91.63%, 84.47% and 71.06% for acceleration, gyroscope, and magnetometer respectively. Similar to *Dataset 1*, the lowest results were observed in the support vector machine classification algorithm. As can be seen in the performance results in both datasets, sensors placed at the ankle took preminent performance results in terms of positions for the majority of the classification algorithms while acceleration sensors showed better performances when compared with gyroscope and magnetometer.

Even though some of the classification algorithms showed acceptable results, there is still room for improving the performances of the activity detection framework. The experimental evaluation showed that a single sensor failed to show better performance all the time. In the next section, the paper presents the results of using different fusion methods to improve the activity detection framework.

Sensor fusion for improved human activity detection system

This section presents the performance results on the multi-view stacking ensemble method for the human activity detection model. In the experiments, evolutionary search algorithms and correlation-based feature selection were utilized to reduce the feature vectors in both *Dataset 1* and *Dataset 2*. Then, the feature vectors from accelerometer, gyroscope, and magnetometer sensors were combined at feature-level and decision-level to access the impact of two data fusion methods for human activity detection. For the feature-level fusion, the reduced feature vectors were column concatenated and support vector machine, k-Nearest Neighbors, J48 decision tree and logistic regression were applied to detect activity details. On the other hand, the multi-view stacking method was used to fuse the decisions generated by individual sensors.

As discussed in “[Decision fusion using multi-view stacking ensemble method](#)” section, adaptive multi-view stacking methods were presented in three experiments. *First*, we use J48 and k-NN as base classifiers and k-NN as meta-classifier. *Second*, LR, k-NN, and J48 were used as base classifiers and LR as Meta-classifier. Finally, LR, k-NN, and J48 were used as base classifiers and combination of LR and k-NN were used as meta-classifiers, which the average result of the classification algorithms was presented as the final classification result. The performance results obtained using Dataset 1 and Dataset 2 are shown in Tables 6, 7. The tables present the accuracy, recall, precision, F-measure, error rate and AUC of each classification model used in our analysis for ankle, chest and

Table 6 The performance results of Feature selection, feature-level fusion and multi-view stacking on Dataset 1 (the best results obtained at each sensor position are italicized)

Positions	Methods	Accuracy (%)	Recall	Precision	F-measure	Errors	AUC
Ankle	SVM	93.08	0.9029	0.9312	0.8987	0.0692	0.9465
	KNN	94.67	0.9243	0.9303	0.9272	0.0533	0.9272
	J48	95.13	0.9283	0.9380	0.9330	0.0487	0.9603
	LR	95.60	0.9430	0.9382	0.9404	0.0440	0.9683
	Stacking-KNN-J48-KNN	96.79	0.9592	0.9583	0.9587	0.0321	0.9772
	Stacking-LR-KNN-J48-LR	97.49	0.9648	0.9673	0.9660	0.0251	0.9804
	Stacking-LR-KNN-J48-MV-LR-KNN	97.57	<i>0.9672</i>	<i>0.9669</i>	<i>0.9670</i>	<i>0.0243</i>	<i>0.9818</i>
Chest	SVM	94.09	0.9107	0.9355	0.9210	0.0591	0.9505
	KNN	94.45	0.9148	0.9428	0.9275	0.0545	0.9528
	J48	93.35	0.9107	0.9185	0.9145	0.0665	0.9500
	LR	95.32	0.9379	0.9373	0.9375	0.0468	0.9654
	Stacking-KNN-J48-KNN	95.05	0.9323	0.9405	0.9362	0.0495	0.9621
	Stacking-LR-KNN-J48-LR	95.67	0.9398	0.9477	0.9435	0.0433	0.9665
	Stacking-LR-KNN-J48-MV-LR-KNN	96.02	<i>0.9425</i>	<i>0.9528</i>	<i>0.9474</i>	<i>0.0398</i>	<i>0.9681</i>
Wrist	SVM	90.96	0.8432	0.9293	0.8758	0.0904	0.9134
	KNN	93.32	0.8828	0.9514	0.9111	0.0668	0.9311
	J48	91.89	0.8885	0.9020	0.8947	0.0811	0.9376
	LR	91.38	0.8780	0.8884	0.8828	0.0862	0.9323
	Stacking-KNN-J48-KNN	94.74	0.9209	0.9448	0.9319	0.0526	0.9560
	Stacking-LR-KNN-J48-LR	95.32	0.9324	0.9472	0.9395	0.0468	0.9623
	Stacking-LR-KNN-J48-MV-LR-KNN	<i>95.48</i>	<i>0.9330</i>	<i>0.9501</i>	<i>0.9412</i>	<i>0.0452</i>	<i>0.9628</i>

Italic values show multiple classifiers combinations with the highest values and produce superior results compared to single classifications and feature-level fusion

Table 7 Performance results of feature selection, feature-level fusion and multi-view stacking on Dataset 2 (the best results obtained at each sensor position are italicized)

Positions	Methods	Accuracy (%)	Recall	Precision	F-measure	Errors	AUC
Ankle	SVM	90.24	0.9021	0.9029	0.9019	0.0976	0.9441
	KNN	98.17	0.9617	0.9818	0.9817	0.0183	0.9896
	J48	96.42	0.9642	0.9643	0.9642	0.0358	0.9795
	LR	96.18	0.9617	0.9619	0.9616	0.0382	0.9781
	Stacking-KNN-J48-KNN	<i>98.50</i>	<i>0.9849</i>	<i>0.9850</i>	<i>0.9849</i>	<i>0.0150</i>	<i>0.9914</i>
	Stacking-LR-KNN-J48-LR	98.46	0.9845	0.9845	0.9845	0.0154	0.9912
	Stacking-LR-KNN-J48-MV-LR-KNN	98.37	0.9837	0.9837	0.9837	0.0163	0.9907
Wrist	SVM	94.31	0.9430	0.9454	0.9431	0.0569	0.9674
	KNN	97.44	0.9743	0.9748	0.9744	0.0256	0.9653
	J48	95.33	0.9534	0.9535	0.9534	0.0467	0.9734
	LR	94.15	0.9415	0.9414	0.9414	0.0585	0.9666
	Stacking-KNN-J48-KNN	97.93	0.9793	0.9794	0.9793	0.0207	0.9882
	Stacking-LR-KNN-J48-LR	97.76	0.9777	0.9777	0.9777	0.0224	0.9872
	Stacking-LR-KNN-J48-MV-LR-KNN	<i>98.05</i>	<i>0.9805</i>	<i>0.9806</i>	<i>0.9805</i>	<i>0.0195</i>	<i>0.9889</i>

Italic values show multiple classifiers combinations with the highest values and produce superior results compared to single classifications and feature-level fusion

wrist placements. Regarding Dataset 1 performance results shown in Table 6, multi-view stacking methods clearly outperformed the feature-level fusion using SVM, J48, k-NN, and LR. Moreover, there is an improved result with all classification compared with using the single sensor modality. The lowest performance results were observed with feature-level fusion and SVM. All the experiments using multi-view stacking outperformed feature-level fusion methods. In the multi-view stacking, the highest performance results were achieved when LR-kNN-J48-(LR-k-NN) (97.57%) followed by LR-k-NN-J48-LR (97.49%) and k-NN-J48-k-NN (96.79%). Moreover, for feature-level fusion using the single classification algorithm, the best accuracy was achieved with logistics regression (95.60%) followed by J48 (95.13%) and k-NN (94.67%).

In Table 6, we observed improvement on performance results using the fusion methods than using the single sensor and classification algorithms by 3% to 17% for acceleration sensor and 10% to 24% with gyroscope sensors using multi-view stacking based fusion method. Similar performance results improvement on accuracies were also observed in feature-level fusion. The highest improvement accuracy obtained was on using the support vector machine, which showed a performance increase by 12% and 20% for acceleration and gyroscope sensor respectively. However, there is a marginal decrease for J48 decision tree when acceleration and gyroscope sensors were fused at the feature-level. This support the theory that feature-level fusion methods for multi-sensor analysis is not efficient for human activity detection and fails to guarantee better performance sometimes [78]. In addition, it provides credence for the proposed robust multi-view stacking method that showed superior performances over both feature level fusion and a single sensor for human activity recognition. In other sensor placement such as chest and wrist, we obtained comparable performances in all the performance metrics used in our evaluations. The multi-view achieved 96.02% and 95.48% accuracy for chest and wrist placement respectively.

Regarding *Dataset 2* shown in Table 7, multi-view stacking outperformed feature-level methods. As shown in Table 7, the highest results were achieved with multi-view stacking using J48-k-NN-k-NN (98.50%) followed by LR-k-NN-J48-LR (98.48%) and LR-k-NN-J48-(LR-k-NN) (98.37%). Similarly, for feature-level fusion, k-NN demonstrated competitive performance results with 98.17% accuracy, 96.17% recall, 96.42% F-measure and 98.96% AUC. K-Nearest Neighbor algorithm showed higher performance compared to other classification algorithms used in our experiments. Furthermore, similar performance results were demonstrated by sensors attached at the wrist for *Dataset 2* shown in Table 7. Specifically, LR-k-NN-J48-(LR-k-NN) multi-view stacking methods achieved 98.05% accuracy, 98.06% Precision and 98.89% AUC while feature-level fusion using k-NN obtains 97.44% accuracy, 97.48% precision, 97.44% F-measure and 96.53% AUC. The performance results obtained with the chest and wrist placements are marginally lower compared to performance results demonstrated by ankle placement. When compared with single sensors (Fig. 4), there is an improvement on classification accuracy by 6% to 26% for acceleration sensor, 5% to 57% for gyroscope and 5% to 19% using the multi-view stacking ensemble algorithm. Similarly, they are noticeable improvements for feature-level fusion for SVM (20.14%), LR (11.91%), k-NN (6.54%) for acceleration sensor, SVM (48.41%), LR (17.72%), k-NN (13.70%) for gyroscope sensors and SVM (20.52%), LR (25.98%) and k-NN (27.11%) for magnetometer respectively. J48 decision tree demonstrated the lowest

performance results improvements with accuracies of 3.94%, 3.25% and 16.66% for acceleration, gyroscope and magnetometer sensors placed at the ankle respectively.

To assess the statistical significance of the proposed adaptive multi-view stacking ensemble algorithms against single accelerometer, gyroscope, magnetometer, and feature-level fusion, we used confidence interval probability [88] discussed in “[Model validation](#)” section. The confidences interval is within the range of 0.55% to 0.68% for all the performance metrics. In all the evaluations, the differences between the single sensors, feature-level fusion, and multi-view stacking ensemble are higher than the confidences interval. The results obtained are considered statistically significant with a confidence interval with 95% interval. Specifically, the proposed multi-view stacking methods provided statistically significant improvements except for feature-level fusion using k-Nearest Neighbors algorithms in *Dataset 2* as shown in Table 7. In all the experimental evaluations, implementation of multiple classifier systems provided better performance results by combining multiple weak base classifiers to create robust activity detection algorithm. The performance results obtained have also been justified by recent studies on using multiple classifier systems for human activity detection [27, 52].

Impact of over-sampling on the performance results

This section presents the impact of over-sampling the minority class activities on proposed multi-view stacking methods. Dataset generated for classification task may have an unequal number of activity classes. These are very common in natural settings such as fraud detection, medical diagnosis, and activity detection and this will result in class imbalance. Class imbalance greatly affects classification algorithm performances, as majority classes tend to overwhelm the minority classes. In the dataset used in our analysis, activities such as ascending stairs, descending stairs and jumping have a lower frequency of occurrence compare to other activities performed in our dataset. As earlier discussed in “[Activity class imbalanced distribution](#)” section, we over-sampled the training data to improve the data distribution of these activities. The number of nearest neighbors $K=5$ and percentage values set to 100, we resample the dataset once. The performance results obtain with resampling the dataset for *Dataset 1* and *Dataset 2* are presented in Tables 8, 9.

Regarding *Dataset 1* presented in Table 8, we observed a marginal increase in both feature-level fusion and multi-view stacking by applying the SMOTE algorithm on the training data. Specifically, for multi-view stacking ensemble, the use of SMOTE algorithms improves the performance results to 97.89% accuracy, 97.18% recall, 97.17% F-measure and 98.44% AUC for sensor attached to the ankle. Moreover, we observed similar improvement in other sensor placements and performance metrics in our experiments. This confirms the positive impact of SMOTE algorithms to improve the activity detection system [6]. In *Dataset 2*, there are also improvements in the performances of both feature-level fusion and multi-view stacking ensemble when SMOTE algorithm was applied to increase data distribution among activity classes. The results are presented in Table 9. In multi-view stacking methods, the highest accuracies of 99.18%, 99.13%, and 99.05% were demonstrated by using the three multi-view stacking experiment using the ankle placement. Moreover, similar performance results were obtained using the wrist placement. For feature-level fusion, the highest accuracies of 98.45%, 96.87%, and 96.08% were obtained through k-NN, LR and J48 respectively for ankle placement. Furthermore,

Table 8 Performance using feature selection and SMOTE algorithm on Dataset 1 (the best results obtained at each sensor position are italicized)

Positions	Methods	Accuracy (%)	Recall	Precision	F-measure	Errors	AUC
Ankle	SVM	93.98	0.9060	0.9299	0.8991	0.0602	0.9488
	KNN	95.33	0.9295	0.9314	0.9303	0.0467	0.9614
	J48	96.45	0.9537	0.9560	0.9547	0.0355	0.9742
	LR	96.48	0.9504	0.9505	0.9502	0.0352	0.9727
	Stacking-KNN-J48-KNN	97.25	0.9629	0.9621	0.9624	0.0275	0.9794
	Stacking-LR-KNN-J48-LR	97.92	0.9709	0.9713	0.9711	0.0208	0.9838
	Stacking-LR-KNN-J48-MV-LR-KNN	<i>97.89</i>	<i>0.9718</i>	<i>0.9717</i>	<i>0.9717</i>	<i>0.0211</i>	<i>0.9844</i>
Chest	SVM	94.43	0.9170	0.9386	0.9254	0.0557	0.9541
	KNN	95.20	0.9307	0.9434	0.9364	0.0480	0.9616
	J48	93.79	0.9197	0.9240	0.9218	0.0621	0.9551
	LR	95.42	0.9419	0.9443	0.9430	0.0458	0.9675
	Stacking-KNN-J48-KNN	95.87	0.9466	0.9465	0.9464	0.0413	0.9702
	Stacking-LR-KNN-J48-LR	96.54	0.9565	0.9586	0.9575	0.0346	0.9756
	Stacking-LR-KNN-J48-MV-LR-KNN	<i>96.73</i>	<i>0.9586</i>	<i>0.9603</i>	<i>0.9594</i>	<i>0.0327</i>	<i>0.9768</i>
Wrist	SVM	91.26	0.8956	0.9254	0.9085	0.0874	0.9407
	KNN	95.30	0.9381	0.9534	0.9451	0.0470	0.9654
	J48	92.39	0.9222	0.9243	0.9232	0.0761	0.9552
	LR	91.72	0.9040	0.9067	0.9052	0.0828	0.9457
	Stacking-KNN-J48-KNN	95.91	0.9504	0.9522	0.9513	0.0409	0.9721
	Stacking-LR-KNN-J48-LR	96.17	0.9524	0.9545	0.9477	0.0383	0.9733
	Stacking-LR-KNN-J48-MV-LR-KNN	<i>96.63</i>	<i>0.9572</i>	<i>0.9616</i>	<i>0.9593</i>	<i>0.0337</i>	<i>0.9760</i>

Italic values show multiple classifiers combinations with the highest values and produce superior results compared to single classifications and feature-level fusion

Table 9 Performance using feature selection and SMOTE algorithms on Dataset 2 (the best results obtained at each sensor position are italicized)

Positions	Methods	Accuracy (%)	Recall	Precision	F-Measure	Errors	AUC
Ankle	SVM	93.67	0.9079	0.9573	0.9241	0.0633	0.9489
	KNN	98.45	0.9780	0.9868	0.9820	0.0155	0.9878
	J48	96.08	0.9537	0.9561	0.9549	0.0392	0.9740
	LR	96.87	0.9642	0.9603	0.9622	0.0313	0.9799
	Stacking-KNN-J48-KNN	<i>99.18</i>	<i>0.9892</i>	<i>0.9924</i>	<i>0.9908</i>	<i>0.0082</i>	<i>0.9940</i>
	Stacking-LR-KNN-J48-LR	99.05	0.9880	0.9898	0.9889	0.0095	0.9933
	Stacking-LR-KNN-J48-MV-LR-KNN	99.13	0.9897	0.9905	0.9901	0.0087	0.9942
Wrist	SVM	94.92	0.9392	0.9427	0.9407	0.0508	0.9658
	KNN	98.18	0.9740	0.9840	0.9787	0.0182	0.9856
	J48	96.74	0.9577	0.9605	0.9590	0.0326	0.9765
	LR	94.59	0.9396	0.9383	0.9389	0.0541	0.9658
	Stacking-KNN-J48-KNN	98.99	0.9864	0.9869	0.9866	0.0101	0.9925
	Stacking-LR-KNN-J48-LR	98.98	0.9846	0.9869	0.9842	0.0120	0.9925
	Stacking-LR-KNN-J48-MV-LR-KNN	<i>99.02</i>	<i>0.9868</i>	<i>0.9871</i>	<i>0.9869</i>	<i>0.0098</i>	<i>0.9927</i>

Italic values show multiple classifiers combinations with the highest values and produce superior results compared to single classifications and feature-level fusion

the lowest accuracy was demonstrated by SVM using feature-level fusion. The performance results achieved by applying the SMOTE algorithms are presented in Table 9.

In the nutshell, in terms of sensor placement, we observed that ankle positions outperformed chest and wrist in all our evaluations. The likely reasons for the high performance of the sensor placed at the ankle are due to the activities considered in our experiments. Most of the activities are motion based and ambulatory activities such as walking, cycling, etc. These activities are strong displacement activities which the sensors placed at the ankle provides [74]. In summary, the best performance results were achieved by the proposed multi-view stacking ensemble algorithms especially with the use of two classification models as meta-classifiers followed by other multi-view stacking methods experiments. In addition, feature-level fusion demonstrated impressive results in our analysis using k-NN, LR and J48 classification algorithms. Compared to other classification algorithms used in our experiments, SVM provided the least performance results. It can be concluded that there is a high correlation between the classification algorithms and proposed multi-view stacking methods in our experiments. The base classification algorithms that achieved impressive results on single accelerometer, gyroscope and magnetometer sensors proved to be a good base classifier combination. This can be seen with the use of logistic regression (LR) and k-Nearest Neighbors (k-NN) as both base classifiers and meta-classifiers.

Furthermore, there are performance differences between the two datasets used in our experiments. *Dataset 2* provides better classification results in both feature-level fusion and multi-view stacking compared to *Dataset 1*. This result is in agreement with the recent implementation of human activity recognition using the same dataset [74]. However, fewer improvements were observed between all the experimental evaluation between applying feature selection alone and using both feature selection methods and SMOTE algorithm. Moreover, applying SMOTE algorithms on the data marginally outperformed data fusion with feature selection only. We postulated that the approach can be improved by increasing the values of nearest neighbors and percentage values in SMOTE algorithms especially for *Dataset 1* and this approach were recently evaluated for predicting diabetes mellitus [71]. Nonetheless, to ensure a fair comparison between the two datasets used in the evaluation, we only used values stated as we have already achieved impressive performance results with *Dataset 2*. In future work, increasing the values of the nearest neighbors and percentages for SMOTE algorithms would be considered.

To ensure the statistical significance of the proposed multi-view stacking ensemble, confidence interval discussed in “[Model validation](#)” section was also used to compare the overall performance results of proposed methods with feature level fusion and other baseline evaluations. Statistical significance was observed between multi-view stacking methods, Feature-level fusion and single sensor analysis with 95% confidence interval.

Significance of the proposed methods for human activity detection

To investigate the significance of the proposed multi-view stacking ensemble methods for human activity detection, we compared the proposed methods with other multiple classifier system methods recently implemented for human activity detection. Three recent studies on multiple classifier systems were chosen, these include weighted majority voting, Bagging, and Random Subspace ensemble [16, 26, 27]. In Saha et al. [16], weighted majority voting approach was considered to combine features extracted from accelerometer sensors

using logistic regression. Here, feature vectors such as mean, standard deviation, variance and standard deviation of the magnitude of accelerometer were extracted over 2 s window size and 50% overlap. Chowdhury et al. [26] evaluated different ensemble algorithm using feature vectors extracted from accelerometer signals. The feature vectors selected for implementation in their proposed method are minimum, maximum, mean, standard deviation, variance, percentile, zero crossing rate energy and dominant frequency of the raw accelerometer signal over 10 s window size with 50% overlapping. Then, Binary decision tree (BDT), Support vector machine (SVM), k-Nearest Neighbors (k-NN) and artificial neural network (ANN) were fused using weighted majority voting. To evaluate the method, maximum epoch and learning rate of ANN were set to 250 and 0.001 respectively.

In addition, support vector machine with a linear kernel was considered while the value of k for k-NN was set to 7 as specified in their study. The specified method was implemented using tenfold cross-validation to ensure accurate comparison with the proposed multi-view stacking ensemble methods. Furthermore, Ghogeski et al. [27] evaluated ensemble algorithms for energy expenditure estimation using Bagging and Random space ensemble. The proposed multi-view stacking methods were compared with these two ensemble methods. Bagging based multiple classifier system method, which has been extensively applied for human activity detection [3] randomly divide the training data into subsets, and train random subset using classification algorithms without replacement. The ensemble method uses an approach called bootstrapping replicate of the original data. Then, the decision produced by each subset of the data is combined with majority voting as the final prediction results. Bagging method is very popular for human activity identification and comprehensive health monitoring and we implement the method to compare our proposed adaptive multi-view stacking ensemble method. In our implementation, the batch size and learning cycle were set to 100 respectively.

On the other hand, Random Subspace is a multiple classifier system that randomly selects a predefined number of features space from the whole training feature vectors to create a different training feature sets. This procedure is repeated several times and at each step, the classification algorithm is trained on the features. Then, the final decision is built by fusion of each model prediction outputs using majority voting [27]. In addition, the batch size and number of iteration for random space ensemble implementation were set to 100 and 10 respectively. Logistic regression was used as base classification algorithm in the implementation of both Bagging and Random Subspace ensemble methods. The performance results achieved using different multiple classifier systems are presented in Table 10. The performance results demonstrated by multi-view stacking are highlighted in the table. Multi-view stacking method outperformed other multiple classifier systems for human activity recognition. The high performance of the proposed multi-view stacking ensemble methods is a result of the use of efficient, diverse classification algorithms and enhanced feature selection methods.

Furthermore, the performance results obtained demonstrate the use of multiple classifier systems for human activity recognition provide robustness and generalization that depicts enhance classification algorithm decision making. In addition, the results obtained are consistent with recent studies [52] on the use of multiple classification algorithms for human activity identification. With the performance of the proposed multi-view stacking approach, it can be concluded that the method is promising for developing comprehensive human activity detection framework.

Table 10 Comparison with other multiple classifier system methods

	Methods	Accuracy (%)	Recall	Precision	F-measure	AUC
Dataset 1	Saha et al. [16]	91.85	0.8806	0.8963	0.8872	0.9340
	Chowdhury et al. [26]	95.32	0.9315	0.9414	0.9342	0.9622
	Ghojeski et al. [27] Random subspace	95.05	0.9272	0.9365	0.9303	0.9599
	Ghojeski et al. [27] Bagging	95.71	0.9381	0.9383	0.9378	0.9660
	Proposed method	97.89	0.9718	0.9717	0.9717	0.9844
Dataset 2	Saha et al. [16]	77.52	0.7749	0.7769	0.7718	0.8714
	Chowdhury et al. [26]	92.28	0.9227	0.9239	0.9224	0.9559
	Ghojeski et al. [27] Random subspace	87.52	0.8750	0.8770	0.8746	0.9286
	Ghojeski et al. [27] Bagging	86.26	0.8623	0.8643	0.8620	0.9213
	Proposed method	99.18	0.9892	0.9924	0.9908	0.9940

Conclusion and future works

This paper presents and investigated experimental use of multi-view stacking ensemble to combine different sensors by exploiting their predictive probabilities for human activity detection and monitoring. We evaluated the proposed approach in three ways. *First*, we conducted an experimental evaluation of four classification algorithms on the original feature vectors extracted from accelerometer, gyroscope, and magnetometer respectively, to assess the impact of each sensor modality for human activity detection and health monitoring. *Second*, evolutionary search algorithms and correlation-based feature selection methods were utilized to reduce features vectors and evaluate the impact of feature-level fusion and multi-view stacking methods. Finally, the impact of increasing the minority activity classes using SMOTE algorithms was equally evaluated for both feature-level fusion and multi-view stacking ensemble algorithms.

In all the experiments, we observed superior performances of the proposed multi-view stacking ensemble algorithms for human activity detection using two publicly available datasets. The performance results obtained showed that the overall detection accuracy can be improved, from approximately 71% to 95% while using all the feature vectors extracted from each sensor modality and classification algorithms. In addition, the performance results can be further improved up to approximately 99% with multi-view stacking ensemble, feature selection methods and SMOTE algorithm. Moreover, our proposed method outperformed baseline methods such as Bagging, majority voting and Random Subspace ensemble using publicly available datasets and improve on the performances of the baseline techniques. From the experimental results, we observed that ankle placement consistently outperformed other placement positions in both feature-level and multi-view stacking fusion methods. Our results clearly demonstrate the validity, impact of multi-view stacking fusion, evolutionary search algorithm based feature selection and SMOTE algorithms, and the capacity of these methods to enhance human activity details for mobile and wearable sensor-based human activity detection and monitoring. The contribution of this paper is the implementation of comprehensive evaluation of human activity identification approach using individual sensors, classification algorithms and multiple classifier system methods. The proposed evaluation methods ensure improved classification accuracy, robustness and reduce performance biases.

Despite the promising results obtained with the proposed approach, there are still areas that require further researches to improve the techniques. Consequently,

limitations of the proposed methods that would be solved in future are outlined here. First, the proposed method utilized publicly available dataset that contain only three sensor modalities (accelerometer, gyroscope and magnetometer). To ensure comprehensive activity identification, health monitoring and status recommendation require other sensor modalities such as pulse rate, video, ECG, EMG, location sensors, radar sensors etc. Additional work is required to collect large and comprehensive datasets with diverse subjects, activity details and environments. Furthermore, future works would focus on using deep learning methods to automatically extract discriminative feature vectors from sensor data. In the case of the current study, handcrafted feature vectors were carefully extracted, but such methods are time-consuming and application dependents. The use of deep learning methods would lead to generalized and improved performance.

Moreover, evaluation of cross-locations fusion of sensors can also be tested by fusion of sensor modalities from different locations through adaptive multi-view stacking approach. Another important area of research is the development of mobile cloud-based and cyber-physical system to support seamless community based human activity recognition and integration of a wide range of multimodal sensors.

Abbreviations

ANN: artificial neural networks; AUC: area under the curve; BKS: behaviour knowledge space; BS: Bayesian network; CNN: convolutional neural networks; DNN: deep neural networks; DST: Dempster–Shafer theory; DT: decision tree; EA: evolutionary search; ECG: electrocardiography; EMG: electromyography; FFT: Fast Fourier Transform; FN: false negative; FP: false positive; GPR: Gaussian process for regression; GPS: Global positioning system; IMU: inertial measurement unit; k-NN: k-Nearest Neighbors; LDC: linear discriminant classifier; LR: logistic regression; LSTM: long short term memory; MFV: master feature vectors; MLR: multiple linear regression; MSC: multiple classifier system; NB: Naïve Bayes; QDC: quadratic discriminant classifier; RF: random forest; RUS: random under sampling; sEMG: surface electromyography; SMOTE: synthetic over-sampling technique; SVR: support vector regression; TN: true negative; TP: true positive; WEKA: Waikato Environment for Knowledge Analysis.

Acknowledgements

The authors would like to thank University of Malaya for sponsoring the paper through the BKP Special grants and researchers that collected the datasets that were used to support this research.

Authors' contributions

HFN conceptualize the Multi-sensor fusion methods. HFN and GM implemented the proposed methods and conduct the experiments. HFN and URA wrote the manuscript draft. YWT supervised the experiments and review the final manuscript. MAA provided formal analysis, guidance on the framework and validation of the experiments. All the authors review the final manuscript for submission. All authors read and approved the final manuscript.

Funding

This research is supported by University of Malaya BKP Special Grant no vote BKS006-2018.

Availability of data and materials

The datasets supporting the conclusion of this article are available in (<https://www.mad.tf.fau.de/research/activitynet/daliac-daily-life-activities/>) for Dataset 1 and (<http://archive.ics.uci.edu/ml/datasets/mhealth+dataset>) for Dataset 2. The software used for experiments are MATLAB 2017b and Java-based WEKA toolkit. MATLAB is a trademark of MathWorks, USA (<https://in.mathworks.com/>) and WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>) is an Open-source Java-based Toolkits for multi-platform issued under the GNU General Public License.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia. ² Computer Science Department, Ebonyi State University, P.M.B 053, Abakaliki, Ebonyi State, Nigeria. ³ Department of Computer Science, Sukkur IBA University, Sukkur 65200, Pakistan. ⁴ Computer Science Department, Federal University, Ndufu-Alike, Abakaliki, Ebonyi State, Nigeria. ⁵ Department of Computer Science and Engineering Department, Qatar University, 2713, Doha, Qatar.

Appendix 1

See Table 11.

Table 11 List of selected features using an evolutionary search algorithm and correlation-based feature selection

Placement	Sensors	Features selected
Feature selection with evolutionary search algorithms-Dataset 1		
Ankle	Accelerometer	Median(X), Min(Y), Energy(X), STD(X), Peak-amplitude(Y), Percentile(X), Mean(X), Energy(X), Peak-amplitude(Z), CRV(Z), Min(Z), VAR(Y), STD(Y), CRV(Z), Harmonic-Mean(Y), Signal-power(X), Signal-power(Y)
	Gyroscope	Signal-power(Z), Max(Z), STD(Z), Harmonic-mean(z), CRV(Z), Peak-amplitude(X),,, peak-amplitude(Y),,, peak-amplitude(Z), STD(Y), signal-power(Y), Harmonic-mean(Y), median(X), VAR(Z), CRV(Y), CRV(X)
Chest	Accelerometer	Signal-Power(Y), Min(Y), Peak-Amplitude(Y), Skewness(Y), CRV(Y), Energy(Y), Min(Z), Peak-Amplitude(X), CRV(Z), Max(X), Percentile(Y), CRV(X), VAR(Y), Energy(Z), Median(Z)
	Gyroscope	CRV(Y), Signal-power(Y), STD(Y), Skewness(Y), Max(Y), Median(Y), Min(Y), ZCR(Z), Signal-power(Z), CRV(Z), IQR(Z), median(X), Peak-amplitude(X), Signal-power(X), IQR(Y), Peak-amplitude(Y)
Wrist	Accelerometer	Min(X), Max(Y), Entropy(Y), CRV(Y), Signal-power(Y), Peak-amplitude(X), Min(Y), Signal-power(X), Harmonic-mean(Z), Median(Y), Energy(X), Energy(Z), CRV(X), VAR(X), CRV(Z), Median(X)
	Gyroscope	CRV(Z), Signal-power(Y), CRV(Y), STD(Z), IQR(X), Signal-power(Z), IQR(Y), Peak-amplitude(X), Percentile(X), Mean(X), IQR(Z), Peak-amplitude(Y), Signal-power(X), Min(Z), Min(X), VAR(Z), Mean(Y)
Feature selection with evolutionary search algorithm-Dataset 2		
Ankle	Accelerometer	Median(Z), Harmonic-Mean(Y), min(Y), CRV(Y), Energy(Y) Signal-power(Y), CRV(Z), Signal-power(Z), max(Y), max(Y), median(Y), VAR(Y), percentile(Y), Max(X), CRV(X), Harmonic-Mean(Z), mean(X)
	Gyroscope	Energy(Z), CRV(Z), Min(Y), Energy(X), Median(Y), VAR(Y), Max(X), Median(Z), CRV(X), Energy(Y), Min(Z), percentile(Z), Signal-power(Z), Max(Y), Median(X), Skewness(Z)
	Magnetometer	Entropy(X), max(X), CRV(X), Median(X), Min(X), Signal-power(X), Skewness(X), Energy(Z), Median(Y), STD(Y), STD(Z), Energy(Y), Percentile(Y), min(Y), Signal-power(Z), Peak-amplitude(X)
Wrist	Accelerometer	VAR(Y), Signal-power(Y), Min(Y), Median(Y), Median(Z), Min(X), VAR(X), Mean(X), Min(Z), Energy(Z), max(Z), CRV(Z), Energy(X), Max(X),
	Gyroscope	Median(Y), VAR(Y), Signal-power(Y), Min(X), Percentile(X), Median(X), Median-F(Z), Max(Z), Percentile(Z), CRV(Z), Energy(Y), Energy(X)
	Magnetometer	Energy(X), STD(X), Max(X), Min(X), Median(X), Peak-amplitude(Z), CRV(X), Max(Y), Peak-amplitude(X), STD(Z), Energy(Z), Peak-amplitude(Y), Median(Z), CRV(Z), Median(Y), Signal-power(Y), Mean(Y), Percentile(Y), VAR(Y), Signal-power(Z), Mean(X)

Received: 10 October 2018 Accepted: 26 August 2019

Published online: 09 September 2019

References

1. Cao L, Wang Y, Zhang B, Jin Q, Vasilakos AV (2017) GCHAR: an efficient Group-based Context-aware human activity recognition on smartphone. *J Parallel Distrib Comput* 118:67–80
2. Ponti M, Bet P, Oliveira CL, Castro PC (2017) Better than counting seconds: identifying fallers among healthy elderly using fusion of accelerometer features and dual-task Timed Up and Go. *PLoS ONE* 12:e0175559
3. Nweke HF, Teh YW, Mujtaba G, Al-garadi MA (2019) Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Inf Fusion* 46:147–170
4. Qiu S, Wang Z, Zhao H, Qin K, Li Z, Hu H (2018) Inertial/magnetic sensors based pedestrian dead reckoning by means of multi-sensor fusion. *Inf Fusion* 39:108–119
5. Spinsante S, Angelici A, Lundstrom J, Espinilla M, Cleland I, Nugent C (2016) A mobile application for easy design and testing of algorithms to monitor physical activity in the workplace. *Mobile Inf Syst*. <https://doi.org/10.1155/2016/5126816>

6. Dobbins C, Rawassizadeh R, Momeni E (2017) Detecting physical activity within lifelogs towards preventing obesity and aiding ambient assisted living. *Neurocomputing* 230:110–132
7. Cornacchia M, Ozcan K, Zheng Y, Vellipalsalar S (2017) A survey on activity detection and classification using wearable sensors. *IEEE Sens J* 17:386–403
8. Chen Y, Wang ZL (2017) A hierarchical method for human concurrent activity recognition using miniature inertial sensors. *Sens Rev* 37:101–109
9. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJ (2016) Complex human activity recognition using smartphone and wrist-worn motion sensors. *Sensors* 16:426
10. Tunca C, Pehlivan N, Ak N, Arnrich B, Salur G, Ersoy C (2017) Inertial sensor-based robust gait analysis in non-hospital settings for neurological disorders. *Sensors* 17:825
11. Ordonez FJ, Roggen D (2016) Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16:115
12. Shoaib M, Bosch S, Scholten H, Havinga PJ, Incel OD (2015) Towards detection of bad habits by fusing smartphone and smartwatch sensors. In: *Pervasive computing and communication workshops (PerCom Workshops)*, 2015 IEEE international conference on, pp 591–596
13. Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R (2009) Activity identification using body-mounted sensors—a review of classification techniques. *Physiol Meas* 30:R1
14. Janidarman M, Fekr AR, Radecka K, Zilic Z (2017) A comprehensive analysis on wearable acceleration sensors in human activity recognition. *Sensors* 17:26
15. Jurek A, Nugent C, Bi Y, Wu S (2014) Clustering-based ensemble learning for activity recognition in smart homes. *Sensors* 14:12285–12304
16. Saha J, Chowdhury C, Biswas S (2018) Two phase ensemble classifier for smartphone based human activity recognition independent of hardware configuration and usage behaviour. *Microsyst Technol* 24:2737–2752
17. Banos O, Damas M, Pomares H, Rojas I (2012) On the use of sensor fusion to reduce the impact of rotational and additive noise in human activity recognition. *Sensors* 12:8039–8054
18. Garcia-Ceja E, Galván-Tejada CE, Brena R (2018) Multi-view stacking for activity recognition with sound and accelerometer data. *Inf Fusion* 40:45–56
19. Chowdhury A, Tjondronegoro D, Chandran V, Trost S (2017) Physical activity recognition using posterior-adapted class-based fusion of multi-accelerometers data. *IEEE J Biomed Health Inform* 99:1
20. Baños O, Damas M, Pomares H, Rojas I (2013) Activity recognition based on a multi-sensor meta-classifier. In: *International work-conference on artificial neural networks*, pp 208–215
21. Chen C, Jafari R, Kehtarnavaz N (2017) A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools Appl* 76:4405–4425
22. Wolpert DH (1992) Stacked generalization. *Neural Netw* 5:241–259
23. Cano A (2017) An ensemble approach to multi-view multi-instance learning. *Knowl Based Syst* 136:46–57
24. Ahmed M, Rasool AG, Afzal H, Siddiqi I (2017) Improving handwriting based gender classification using ensemble classifiers. *Expert Syst Appl* 85:158–168
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–357
26. Chowdhury AK, Tjondronegoro D, Chandran V, Trost SG (2017) Ensemble methods for classification of physical activities from wrist accelerometry. *Med Sci Sports Exerc* 49:1965
27. Gjoreski H, Kaluza B, Gams M, Milic R, Lustrek M (2015) Context-based ensemble method for human energy expenditure estimation. *Appl Soft Comput* 37:960–970
28. Gravina R, Ma CC, Pace P, Aloï G, Russo W, Li WF et al (2017) Cloud-based Activity-as-a-Service cyber-physical framework for human activity monitoring in mobility. *Future Gener Comput Syst* 75:158–171
29. Pires I, Garcia N, Pombo N, Flórez-Revuelta F, Spinsante S (2018) Approach for the development of a framework for the identification of activities of daily living using sensors in mobile devices. *Sensors* 18:640
30. Biagetti G, Crippa P, Falaschetti L, Orcioni S, Turchetti C (2017) A portable wireless sEMG and inertial acquisition system for human activity monitoring. In: *Rojas I, Ortuno F (eds) Bioinformatics and biomedical engineering, Iwbbio 2017, Pt II, vol 10209*. Springer International Publishing Ag, Cham, pp 608–620
31. Bhattacharjee S, Kishore S, Swetapadma A, IEEE (2018) A comparative study of supervised learning techniques for human activity monitoring using smart sensors. *IEEE*, New York
32. Cvetković B, Szecklicki R, Janko V, Lutowski P, Luštrek M (2017) Real-time activity monitoring with a wristband and a smartphone. *Inf Fusion* 43:77–93
33. Saeedi S, Moussa A, El-Sheimy N (2014) Context-aware personal navigation using embedded sensor fusion in smartphones. *Sensors* 14:5742–5767
34. Tolstikov A, Hong X, Biswas J, Nugent C, Chen L, Parente G (2011) Comparison of fusion methods based on dst and dbn in human activity recognition. *J Control Theory Appl* 9:18–27
35. Amoretti M, Copelli S, Wientapper F, Furfari F, Lenzi S, Chessa S (2013) Sensor data fusion for activity monitoring in the PERSONA ambient assisted living project. *J Ambient Intell Humaniz Comput* 4:67–84
36. Al-Jawad A, Barlit A, Romanovas M, Traechtler M, Manoli Y (2013) The use of an orientation Kalman filter for the static postural sway analysis. In: *3rd international conference on biomedical engineering and technology—Icbet 2013, vol 7*, pp 93–102
37. Chen JIZ (2011) An algorithm of mobile sensors data fusion tracking for wireless sensor networks. *Wireless Pers Commun* 58:197–214
38. Sebbak F, Benhammedi F, Chibani A, Amirat Y, Mokhtari A (2014) Dempster-Shafer theory-based human activity recognition in smart home environments. *Ann Telecommun* 69:171–184
39. Phan T, Kalasapur S, Kunjithapatham A (2014) Sensor fusion of physical and social data using Web SocialSense on smartphone mobile browsers. In: *Consumer communications and networking conference (CCNC)*, 2014 IEEE 11th, pp 98–104

40. Luo RC, Chang CC, Lai CC (2011) Multisensor fusion and integration: theories, applications, and its perspectives. *IEEE Sens J* 11:3122–3138
41. Berenguer M, Bouzid M-J, Makni A, Lefebvre G, Noury N (2017) Evolution of activities of daily living using inertia measurements: the lunch and dinner activities. *J Int Soc Telemed eHealth* 5:10-1
42. Zdravevski E, Stojkoska BR, Standl M, Schulz H (2017) Automatic machine-learning based identification of jogging periods from accelerometer measurements of adolescents under field conditions. *PLoS ONE* 12:e0184216
43. Banos O, Villalonga C, Bang J, Hur T, Kang D, Park S et al (2016) Human behavior analysis by means of multimodal context mining. *Sensors* 16:1264
44. Ghasemzadeh H, Amini N, Saeedi R, Sarrafzadeh M (2015) Power-aware computing in wearable sensor networks: an optimal feature selection. *IEEE Trans Mob Comput* 14:800–812
45. Fong S, Song W, Cho K, Wong R, Wong KKL (2017) Training classifiers with shadow features for sensor-based human activity recognition. *Sensors* 17:476
46. Köping L, Shirahama K, Grzegorzec M (2018) A general framework for sensor-based human activity recognition. *Comput Biol Med* 95:248–260. <https://doi.org/10.1016/j.combiomed.2017.12.025>
47. Nishida M, Kitaoka N, Takeda K (2014) Development and preliminary analysis of sensor signal database of continuous daily living activity over the long term. In: Paper presented at the Asia-Pacific signal and information processing association, 2014 annual summit and conference (APSIPA), pp 1–6
48. San-Segundo R, Blunck H, Moreno-Pimentel J, Stisen A, Gil-Martin M (2018) Robust human activity recognition using smartwatches and smartphones. *Eng Appl Artif Intell* 72:190–202
49. Li F, Shirahama K, Nisar M, Köping L, Grzegorzec M (2018) Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors* 18:679
50. Xu Y, Shen Z, Zhang X, Gao Y, Deng S, Wang Y et al (2017) Learning multi-level features for sensor-based human action recognition. *Pervasive Mob Comput* 40:324–338
51. Fatima I, Fahim M, Lee Y-K, Lee S (2013) A genetic algorithm-based classifier ensemble optimization for activity recognition in smart homes. *TIIS* 7:2853–2873
52. Catal C, Tufekci S, Pirmitt E, Kocabag G (2015) On the use of ensemble of classifiers for accelerometer-based activity recognition. *Appl Soft Comput* 37:1018–1022
53. Tripathi AM, Baruah D, Baruah RD (2015) Acoustic sensor based activity recognition using ensemble of one-class classifiers. In: 2015 IEEE international conference on evolving and adaptive intelligent systems (Eais), p 7
54. Peng L, Chen L, Wu X, Guo H, Chen G (2016) Hierarchical complex activity representation and recognition using topic model and classifier level fusion. *IEEE Trans Biomed Eng* 64:1369–1379
55. Guan Y, Plötz T (2017) Ensembles of deep lstm learners for activity recognition using wearables. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 1:11
56. Banos O, Damas M, Guillen A, Herrera LJ, Pomares H, Rojas I et al (2015) Multi-sensor fusion based on asymmetric decision weighting for robust activity recognition. *Neural Process Lett* 42:5–26
57. Banos O, Damas M, Pomares H, Rojas I (2013) Activity recognition based on a multi-sensor meta-classifier. In: Rojas I, Joya G, Cabestany J (eds) *Advances in computational intelligence*, Pt II, vol 7903. Springer, Berlin, pp 208–215
58. Banos O, Toth MA, Damas M, Pomares H, Rojas I (2014) Dealing with the effects of sensor displacement in wearable activity recognition. *Sensors* 14:9995–10023
59. Peng L, Chen L, Wu M, Chen G (2018) Complex activity recognition using acceleration, vital sign, and location data. *IEEE Trans Mob Comput* 18:1488–1498
60. Khan SS, Taati B (2017) Detecting unseen falls from wearable devices using channel-wise ensemble of autoencoders. *Expert Syst Appl* 87:280–290
61. Bulling A, Blanke U, Schiele B (2014) A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput Surv* 46:1–33
62. Banos O, Galvez J-M, Damas M, Pomares H, Rojas I (2014) Window size impact in human activity recognition. *Sensors* 14:6474
63. Leutheuser H, Schuldhuis D, Eskofier BM (2013) Hierarchical, multi-sensor based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset. *PLoS ONE* 8:e75196
64. Banos O, Galvez JM, Damas M, Guillen A, Herrera LJ, Pomares H, et al (2014) Evaluating the effects of signal segmentation on activity recognition. In: *Proceedings Iwbbio 2014: international work-conference on bioinformatics and biomedical engineering*, Vols 1 and 2, pp 759–765
65. Figo D, Diniz PC, Ferreira DR, Cardoso JM (2010) Preprocessing techniques for context recognition from accelerometer data. *Pers Ubiquitous Comput* 14:645–662
66. Ouyang Z, Sun X, Chen J, Yue D, Zhang T (2018) Multi-view stacking ensemble for power consumption anomaly detection in the context of industrial internet of things. *IEEE Access* 6:9623–9631
67. Zhu J, San-Segundo R, Pardo JM (2017) Feature extraction for robust physical activity recognition. *Human-centric Comput Inf Sci* 7:16
68. Nweke HF, Teh YW, Al-garadi MA, Alo UR (2018) Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. *Expert Syst Appl* 105:233–261
69. Adewole KS, Anuar NB, Kamsin A, Sangaiah AK (2017) SMSAD: a framework for spam message and spam account detection. *Multimedia Tools Appl* 78:3925–3960
70. Manurung H (2004) An evolutionary algorithm approach to poetry generation. Doctor of Philosophy PhD, University of Edinburgh
71. Alghamdi M, Al-Mallah M, Keteyian S, Brawner C, Ehrman J, Sakr S (2017) Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: the Henry Ford Exercise Testing (FIT) project. *PLoS ONE* 12:e0179805
72. Quinlan JR (1986) Induction of decision trees. *Mach Learning* 1:81–106
73. Lara ÓD, Pérez AJ, Labrador MA, Posada JD (2012) Centinela: a human activity recognition system based on acceleration and vital sign data. *Pervasive Mob Comput* 8:717–729

74. Zdravevski E, Lameski P, Trajkovic V, Kulakov A, Chorbev I, Goleva R et al (2017) Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering. *IEEE Access* 5:5262–5280. <https://doi.org/10.1109/access.2017.2684913>
75. Chapelle O, Vapnik V, Bousquet O, Mukherjee S (2002) Choosing multiple parameters for support vector machines. *Mach Learning* 46:131–159
76. Mujtaba G, Shuib L, Raj RG, Rajandram R, Shaikh K, Al-Garadi MA (2017) Automatic ICD-10 multi-class classification of cause of death from plaintext autopsy reports through expert-driven feature selection. *PLoS ONE* 12:e0170242
77. Adeniyi DA, Wei Z, Yongquan Y (2016) Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Appl Comput Inform* 12:90–108
78. Shoaib M, Bosch S, Incel O, Scholten H, Havinga P (2014) Fusion of smartphone motion sensors for physical activity recognition. *Sensors* 14:10146
79. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJM (2015) A survey of online activity recognition using mobile phones. *Sensors* 15:2059–2085
80. Bhasuran B, Murugesan G, Abdulkadhar S, Natarajan J (2016) Stacked ensemble combined with fuzzy matching for biomedical named entity recognition of diseases. *J Biomed Inform* 64:1–9
81. Gandhi H, Green D, Kounios J, Clark CM, Polikar R (2008) Stacked generalization for early diagnosis of Alzheimer's disease, pp 5350–5353
82. Gandhi I, Pandey M (2015) Hybrid ensemble of classifiers using voting. In: Green computing and internet of things (ICGCIoT), 2015 international conference on, pp 399–404
83. Banos O, Garcia R, Holgado-Terriza JA, Damas M, Pomares H, Rojas I, et al. (2014) mHealthDroid: a novel framework for agile development of mobile health applications. In: Pecchia L, Chen LL, Nugent C, Bravo J (eds) Edsambient assisted living and daily activities: 6th international work-conference, IWAAL 2014, Belfast, UK, December 2–5, 2014. Proceedings, Cham, Springer International Publishing, pp 91–98
84. Banos O, Villalonga C, Garcia R, Saez A, Damas M, Holgado-Terriza JA et al (2015) Design, implementation and validation of a novel open framework for agile development of mobile health applications. *Biomed Eng Online* 14:S6
85. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. *ACM SIGKDD Explor Newsl* 11:10–18
86. Nweke HF, Teh YW, Alo UR, Mujtaba G (2018) Analysis of multi-sensor fusion for mobile and wearable sensor based human activity recognition. In: Presented at the proceedings of the international conference on data processing and applications, Guangdong, China
87. Mujtaba G, Shuib L, Raj RG, Rajandram R, Shaikh K, Al-Garadi MA (2018) Classification of forensic autopsy reports through conceptual graph-based document representation model. *J Biomed Inform* 82:88–105
88. Cox DR, Hinkley DV (1979) *Theoretical statistics*. Chapman and Hall/CRC, London
89. Cao J, Li W, Ma C, Tao Z (2018) Optimizing multi-sensor deployment via ensemble pruning for wearable activity recognition. *Inf Fusion* 41:68–79

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
