

Genetic Algorithm-Based Classifiers Fusion for Multisensor Activity Recognition of Elderly People

Saisakul Chernbumroong, Shuang Cang, and Hongnian Yu, *Member, IEEE*

Abstract—Activity recognition of an elderly person can be used to provide information and intelligent services to health care professionals, carers, elderly people, and their families so that the elderly people can remain at homes independently. This study investigates the use and contribution of wrist-worn multisensors for activity recognition. We found that accelerometers are the most important sensors and heart rate data can be used to boost classification of activities with diverse heart rates. We propose a genetic algorithm-based fusion weight selection (GAFW) approach which utilizes GA to find fusion weights. For all possible classifier combinations and fusion methods, the study shows that 98% of times GAFW can achieve equal or higher accuracy than the best classifier within the group.

Index Terms—Ambient intelligence, genetic algorithm (GA), neural networks, sensor fusion, smart homes, support vector machine (SVM).

I. INTRODUCTION

THE number of people aged 65 and over has increased significantly over the years. In 2050, the number will reach to 1.9 billion people [1]. The increase has a significant effect on health care. Issues such as high demand in long-term care, rise in health care cost, and ineffective and insufficient care are expected. One of the ways to lessen the issues is to promote home-based care. In this study, we develop an activity recognition model that can be used to recognize or detect an activity of a person. The detected activity can be used to provide information and intelligent services to health care professionals, carers, elderly people, and their families so that the elderly people can remain at home as long as possible.

In activity recognition, a vast variety of sensors have been explored and used such as accelerometer [2]–[4], gyroscope [5], light [4], [6], motion sensor [7], magnetometer [5], [6], microphone [4], [6], [7], barometer [8], temperature [2], [6], RFID [9], etc. Based on sensor types, activity recognition can be divided into two approaches. In the on-object sensor-based activity recognition approach, the sensors are attached to objects in inhabitant area. For example, sensors were installed on furniture to infer activities [10]. RFID was used to identify the objects to

infer detailed activities such as putting on lotion, taking photo, etc. [9]. The approach exploits the semantic relationships between objects and activities to classify activities. However, a large number of sensors are required which is an infeasible and time-consuming process; uncertainty of sensors such as false start and inability to detect object could also result in a poor recognition rate. An on-body sensor-based activity recognition approach, on the other hand, collects data, e.g., movement from sensors distributed on human body. Common on-body sensors include accelerometer and inertia sensor. The sensors are attached to body locations such as chest [10], wrist [2], waist [3], etc. However, sensors are required to be worn all the time which may interrupt or reduce mobility of a user or even obstruct daily activities routine. Especially in elderly care applications, the sensors may be perceived as stigmatized. It is important that the activity recognition model is highly accurate and practical. Recent work [2], [4] showed an activity recognition model, which is practical and highly accurate based on wrist-worn sensors. Accelerometer, temperature sensor, and altimeter were used and 90% accuracy was achieved.

A limited number of studies have been carried out on wrist-worn-based activity recognition. For example, multisensor wrist-worn equipment was used to detect walking, walking upstairs, walking downstairs, sitting, and running [4]. In this study, we use accelerometer, temperature sensor, and altimeter as appeared in [2]. In addition, heart rate monitor, barometer, and light sensor are investigated. Heart rate can be used to measure physical activities indirectly as heart rate is proportional to the intensity of movement and oxygen supplied to skeletal muscles [11]. A combination of acceleration and heart rate improves the accuracy of estimation of energy expenditure by 1.4% [12]. However, the study concluded that the small improvement was not worth it as the user needs to wear the heart rate monitor at all times. Accelerometer and barometer (air pressure differential) were used to detect ambulatory movements considering vertical position shifts [8]. Combining barometer and accelerometer improved classification accuracy in child activities [3].

Temperature could be used to indicate changes in environment when performing certain activities, e.g., washing dishes and brushing teeth may involve a use of water or when ironing, the temperature maybe higher than normal. Several works used the temperature sensor as part of their activity recognition systems [6], [7]. For example, the difference of temperature of 15 min was used to determine the use of a shower [7]. Work in [4] and [6] used light sensors as part of their activity recognition systems. Gyroscope can be used to estimate the orientation and rotation of the movement. After gyroscope and magnetometer were added to the accelerometer, the accuracy was increased

Manuscript received September 4, 2013; revised January 17, 2014; accepted March 12, 2014. Date of publication; date of current version. (*Corresponding author: H. Yu.*)

S. Chernbumroong and H. Yu are with the Faculty of Science and Technology, Bournemouth University, Poole, Dorset BH12 5BB, U.K. (e-mail: schernbumroong@bournemouth.ac.uk; yuh@bournemouth.ac.uk).

S. Cang is with the School of Tourism, Bournemouth University, Poole, Dorset BH12 5BB, U.K. (e-mail: scang@bournemouth.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JBHI.2014.2313473

by 17% [5]. From the literature, it can be seen that combining these sensors could improve classification accuracy. However, these sensors have not yet been combined on the wrist location. This study investigates the contribution of different sensors in the system and what features should be extracted from these sensors.

Many machine learning algorithms such as multilayer perceptron [2], [6], support vector machine (SVM) [2], [3], [7], [12], and decision tree [3], [6], [12] are used in activity classification. Previous studies also showed that combining multiple classifiers help improve classification accuracy. An investigation on several techniques for classifier fusion and fusion weights is carried out. Fusion strategies explored includes majority voting, product, sum, min, max, ranking, and weighted average. Furthermore, since some classifiers may perform better than others, it is sensible to incorporate weights to the classifiers. Fusion weight functions including weighted accuracy (WACC) and techniques used in forecasting domain [13], i.e., simple average (SA), variance-covariance (VACO), and discounted mean square forecast error (DMSFE) are investigated. These techniques offer easy weight determination. However, they may not yield the optimal results for all fusion methods as they are deterministic.

We use genetic algorithm (GA) to determine the fusion weights. Studies indicated GA could improve the classifier fusion accuracy [14], [15]. For example, classifier combination using eight to ten ensembles generated from different techniques was studied in [14]. Weight combination using GA to combine several Bayesian classifiers was investigated in [15]. However, there are some limitations on these studies. First, most of them focused only on the fusion of all classifiers. For example, they produced six classifiers then used GA to combine them. Based on this, the conclusion that GA could improve classifier combination accuracy is not always true as all possible combinations have not been tested. Second, fitness functions such as a function that reflects on the classifier combination functions, e.g., sum, min, max, product, ranking, and weighted average have not been investigated before. Finally, some of these results were often compared with the mean accuracy of a set of classifiers rather than to the best individual classifier. The mean accuracy is always equal to or less than the accuracy of the best individual classifier (equal accuracy is only occurred if and only if all classifiers have the same accuracy). For example, if there are three classifiers with 90%, 85%, 95%, the mean accuracy is 90% which is less than the best individual (95%). This weakens the conclusion that the classifier combination is better than a single classifier.

In this study, we collected a multisensor dataset of 13 activities of elderly persons in a real home. We propose a multisensor activity recognition which fuses information at feature and classifier levels. We investigate the use of GA for fusion weight that addresses the previous limitations.

II. SYSTEM ARCHITECTURE AND ALGORITHMS

A. Multisensor Activity Recognition Framework

Data from multisensor are fused at feature and classifier levels. The reason for performing two-level data fusion is that some of the sensors are not useful in recognizing activities by them-

selves, e.g., altimeter and temperature. Therefore, these sensors should be combined at feature level so that they can be used to provide more information for the classification. Second, results from literature works showed that there is no best classifier for all tasks. Therefore, fusing the data at classifier level would improve the classification accuracy.

Here, we describe the multisensor activity recognition platform (see Fig. 1). First, the system receives data from multiple sensors. For n sensors, the framework raw input is defined as $\{x_i, y_i\}$, where $x = S_1, S_2, \dots, S_n$ and y is the output of K possible activities. The raw inputs are then preprocessed using the weighted moving average (WMA). WMA is used to smoothen the signal using $A_t = w_1 A_t + w_2 A_{t-1}$ where A is the signal at time t . Next, the data are scaled to [0 1] range. The feature set F is extracted (see Section II-D) and fed into feature selection process using the feature combination (FC) technique (see Section II-E) resulting in a feature set S . This is the first level of data fusion where features from different sensors are combined and used in a classifier. The feature set S is passed through a multiclassification block (see Section II-F) which produces class posterior probability $\hat{P}_i^{(j)}$. This is the second level of data fusion where decisions from multiple classifiers are combined. The classifiers are combined and the fusion weights are determined using GA (see Sections II-I and II-G) to produce the final prediction.

B. Multisensor Platform

Accelerometer, temperature sensor, and altimeter are embedded on the CC430F6137 Microcontroller with the MSP430 CPU from Texas Instruments. The accelerometer can measure 3-D acceleration at a range of up to $\pm 2G$ ($G = 9.81 \text{ m/s}^2$) with a resolution of 18 mG. The pressure sensor's measuring range is of 30–120 kPa with 6 Pa resolution. Gyroscope, barometer, and light sensor are implemented on Gadgeteer FEZ Cerberus board with the 168-MHz 32 bit Cortex M4 processor. The gyroscope can measure up to $\pm 2000^\circ/\text{s}$ and 14.375 LSBs per $^\circ/\text{s}$ sensitivity. The Barometer measures the 300–1100 hPa absolute Pressure Range. We used the heart rate monitor chest strap from BlueRobin. The accelerometer and gyroscope are sampling at 33 Hz, while other sensors are sampling at 1 Hz.

A heart rate monitor is worn over the chest using the chest strap. Accelerometer, gyroscope, light, and barometer sensors are worn on the dominant wrist and temperature sensor and altimeter are worn on the other wrist. The sensors are separated between the two wrists due to the limitation on the hardware. The separation is designed in such a way that it should not interfere with the activity recognition. The sensors that are related to the movement, i.e., accelerometer and gyroscope, are worn on the dominant wrist in order to capture the activity's movement. Also, barometer and light sensors are also worn on the dominant wrist as they are parts of the Microsoft Gadgeteer platform. The temperature sensor that captures the body temperature and altimeter are worn on the nondominant wrist. Acceleration and heart rate are transmitted wirelessly to PC over 868 MHz. Temperature and altitude are stored on a 8-kB flash on the watch,

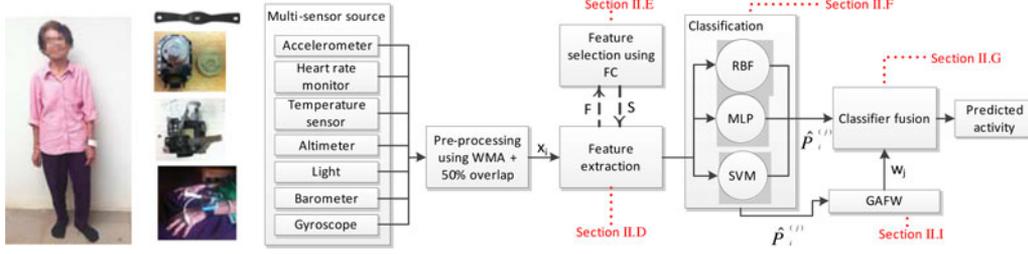


Fig. 1. Multisensor activity recognition of an elderly person framework.

while data from gyroscope, barometer, and light sensors are stored on a 2-GB SD card.

C. Application of the Proposed Framework Scenario

In this section, we provide a scenario of how the proposed framework will be used in detecting activities of an elderly person in home. The proposed activity recognition model is to be developed offline and stored on the PC of the elderly person's home. In real applications, we expect that the sensors will be embedded on a single watch. The elderly person wears the watch that contains multiple sensors on the dominant wrist. The sensor data from the watch are sent wirelessly to the PC. The activity recognition model performs a classification and stores the results on PC. The predicted activity is encrypted and sent over the Internet to the relevant stakeholders such as health professionals, carer, and families. The stakeholders' PCs or mobiles should contain specialized models for further analysis, e.g., to generate behavior pattern for health professionals, to show a day activity report for family members, etc. The detected activity can also be used to provide services in homes. For example, if the sleeping activity is detected, the bedroom's temperature and light can be automatically adjusted.

D. Feature Extraction

Twelve raw sensor data are extracted including 3-D acceleration, heart rate, temperature, altitude, light, barometer temperature, barometer pressure, and 3-D rotation. Also, the magnitudes $\sqrt{x^2 + y^2 + z^2}$ of both acceleration and rotation are calculated. In total, there are 14 input data. For each input, features including mean, standard deviation (STD), maximum, minimum, median, mode, kurtosis, skewness, intensity, difference, root-mean-square (RMS), energy, entropy, and key coefficient are extracted. The key coefficient is the summation of the signal coefficients from 0.5 to 3 Hz. Also, correlations between each acceleration axis and gyroscope axis are calculated. In total, 202 features are extracted.

E. Feature Selection Algorithm

In this study, we use the FC proposed in [2]. FC that is based on neural network and Clamping technique [16] measures the impact of the clamped features within the network and selects a group of features which as a whole achieve the best result. Given $S = \{\}$ where S is a set of selected features, $F = \{f_1, f_2, \dots, f_N\}$ where F is a set of N features, and $g()$ is

the generalized performance of the network. The FC algorithm is as follows:

- 1) First calculate the feature importance for all features f_i in the feature space F using $Im(f_i) = 1 - \frac{g(F|f_i=\bar{f}_i)}{g(F)}$.
 - 2) Select the feature f_s in F which has the maximum impact $f_s = \max_{f_i \in F} Im(f_i)$.
 - 3) If and only if $g(S \cup f_s) \geq g(S)$, then update S and F using $S = S \cup \{f_s\}$ and $F = F \setminus \{f_s\}$.
- Repeat steps 2 and 3 for $N - 1$ times.

F. Classification Algorithm

This study focuses on three algorithms which are widely used in the sensor-based activity recognition research.

- 1) *Multilayer perceptron neural network (MLP)*: MLP uses the concept of connectionist where several input nodes are connected with associated weights to several outputs nodes. The network output can be calculated from the summation function $o_i = \phi(\sum_i W_i x_i)$, where W_i is the weight used for adjusting input x_i and ϕ is the activation function [17]. MLP learns the classification error through a back propagation algorithm and tries to find the weights to minimize that error.
- 2) *Radial basis function neural network (RBF)*: RBF [17] is a neural network that uses RBF as an activation function. For N hidden neurons, the activation function is $f(x) = \sum_{i=1}^N W_i \varphi(\|x - c_i\|)$, where c_i is the center vector for neuron i and φ is a kernel function.
- 3) *Support vector machine*: SVM [18] constructs decision boundaries by solving the optimization objective $\min_{W,b,\xi} \frac{1}{2} W^T W + C \sum_{i=1}^m \xi_i$ subject to $y_i(W^T f(x_i) + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$. The slack term ξ_i is used to relax the constraints allowing misclassified examples. The associated cost parameter C is used for penalizing ξ_i . $f()$ is a function which transforms the input x_i into a higher dimensional space. This study used an RBF kernel function $f(x_i) = \exp(-\frac{1}{(2\sigma^2)} \|x_i - x_j\|^2)$, where σ is the width of the Gaussian kernel. For multiclass classification, we constructed K binary classifiers and applied one-vs-all classification.

Posterior probability can be produced by cooperating some functions, e.g., softmax or by solving optimization function, i.e., $\min_p \frac{1}{2} \sum_{i=1}^k \sum_{j:j \neq i} (r_{ji} p_i - r_{ij} p_j)^2$, where r_{ij} is the pairwise class probabilities between class i and j and p_i is the probability of class i .

TABLE I
FUSION WEIGHT FUNCTION STUDIED IN THIS PAPER

Fusion function	Equation
MV	$pred_i = mode_J \{pred_i^{(j)}\}$
Product	$pred_i = \max_K \left\{ \frac{1}{p(C_k)^{J-1}} \prod_{j=1}^J (\hat{P}_{ik}^{(j)})^{w_j} \right\}$
Sum	$pred_i = \max_K \left\{ \frac{1}{J} \sum_{j=1}^J (\hat{P}_{ik}^{(j)})^{w_j} \right\}$
Min	$pred_i = \max_K \left\{ \frac{\min_J (\hat{P}_{ik}^{(j)})^{w_j}}{\sum_{k=1}^K \min_J (\hat{P}_{ik}^{(j)})^{w_j}} \right\}$
Max	$pred_i = \max_K \left\{ \frac{\max_J (\hat{P}_{ik}^{(j)})^{w_j}}{\sum_{k=1}^K \max_J (\hat{P}_{ik}^{(j)})^{w_j}} \right\}$
Ranking	$pred_i = \max_K \sum_{j=1}^J w_j rank_{ik}^{(j)}$
WA	$pred_i = \max_K \sum_{j=1}^J w_j \hat{P}_{ik}^{(j)}$

G. Multimodel Fusion Methods

In this study, seven classifier fusion methods which are widely used in the classifier combination context are investigated [14], [15]. The idea of the majority vote (MV) is to combine all the votes given by each models selecting the class which has the highest vote. For product method, the probabilities are combined using vector product. The product rule is more sensitive to objection than support where the class with low probability is more influenced to the decision than the class with high probability. For the sum technique, the probabilities are combined using sum function and the class which has the highest maximum average probability is selected. The sum function generates the result based on the average decisions of all classifiers which is similar to majority voting. However, the sum method utilizes class probabilities. The maximum method decides the result based on the most confident classifier where it selects the class with the highest probability from all the models. The min method combines classifiers' results by selecting the class which is least objection by all the models. For the ranking method, first the probability $\hat{P}_i^{(j)}$ is converted to ranks where the maximum rank score is K and minimum is 1. The class with the maximum rank is selected. Weighted Average (WA) associate classifiers' decisions with weights. WA is the same as sum when the weights are 1.

Given that $pred_i^j$ is the prediction of input x_i using classifier model j , $\hat{P}_{ik}^{(j)}$ is the posterior probability that x_i belongs to class k and w_j is the weight for classifier model j , the prediction of the multi-model fusion can be calculated using equations in Table I. In case of equal scores, the model selects the result based on the best classifier.

H. Fusion Weight

Since each classification model may be superior to others, it is common to incorporate weights to the models to reflect this. Six weight functions are studied. SA gives the average weights to all classifiers. VACO uses the mean square error to calculate the weights. In this study, we modified the VACO equation to suit a classification problem by utilizing class probabilities. DMSFE is the modified version of VACO where a parameter β is used to discount weights of the instances. Unit weight gives all classifier weights 1 which means all classifiers are associated with weights. WACC uses the weighted accuracy of each model as the weights. Note that all calculated weights must

TABLE II
FUSION WEIGHT FUNCTION STUDIED IN THIS PAPER

Fusion weight	Equation
SA	$w_j = \frac{1}{J}$
VACO	$w_j = \frac{[\sum_{i=1}^m (1 - \hat{P}_{iK}^{(j)})]^{-1}}{\sum_{j=1}^J [\sum_{i=1}^m (1 - \hat{P}_{iK}^{(j)})]^{-1}}$
DMSFE	$w_j = \frac{[\sum_{i=1}^m \beta^{m-i+1} (1 - \hat{P}_{iK}^{(j)})]^{-1}}{\sum_{j=1}^J [\sum_{i=1}^m \beta^{m-i+1} (1 - \hat{P}_{iK}^{(j)})]^{-1}}, \beta = [0.95, 0.9, 0.85, 0.80]$
Unit weight	$w_j = 1$
WACC	$w_j = \frac{Accuracy^{(j)}}{\sum_{j=1}^J Accuracy^{(j)}}$

be summed to 1, i.e., $\sum_{j=1}^J w_j = 1$. This is except for the unit weight function where all the weights are 1.

Given m training examples and J models, and $\hat{P}_{iK}^{(j)}$ is the probability that model j predicts that data x_i belongs to class K , given that the true class is K , the weight for each classifier model j can be calculated using equations in Table II.

I. Genetic Algorithm Based Fusion Weight (GAFW)

In this study, we propose to use GA to find weights for classifiers. GA [19] has been commonly used to solve an optimization problem. The advantage of GA over other optimization techniques is that instead of starting at a single point to find the solution, a population of points is created. It mimics natural selection in which the population is modified over time. Individuals are randomly selected as parents to produce children of the next generation.

1) *Fitness Function*: GA is used to find the weights that minimize the mean square of the combination error. The classification error is defined as follows:

$$error_i = \begin{cases} 1, & \text{if } tru_i = pred_i \\ 0, & \text{otherwise.} \end{cases}$$

We propose to use the fitness function (ff) according to the fusion method. Given the fusion method (fm) as any function described in Section II-G, the fitness function is defined as

$$ff(w_j) = \frac{1}{2m} \sum_{i=1}^m error(tru_i, fm_i(w_j))^2.$$

The linear weight fitness function (GA-Linear) is also explored:

$$ff(w_j) = \frac{1}{2m} \sum_{i=1}^m error \left(tru_i, \max_K \left\{ \frac{1}{J} \sum_{j=1}^J (\hat{P}_{ik}^{(j)}) * w_j \right\} \right)^2.$$

2) *Population Initialization*: The weight for each classifier is represented in each bit of a chromosome. For each combination, we have J bits. Each bit is represented by a real number between 0 and 1. In order to make sure that the weight obtained will result in higher classification accuracy, a population that covers the search space and near a possible optimum point is necessary. We propose to use the following strategy to initialize the population. First, one of the populations must contain weighted average accuracy chromosomes. Second, the weights are randomly generated from a uniform distribution and the highest weight is assigned to the best model. Note that the weight for the best model within the group is generated randomly between

$\frac{1}{J}$ and 1. The initial population of $20 \times J$ chromosomes are generated.

3) *Crossover and Mutation*: The crossover rate is set to 0.8, same as used in [15]. Adaptive mutation is used where it randomly generates directions that are adaptive with respect to the last generation. The feasible region is bounded by the constraint ($0 \leq w_j \leq 1$). A mutant is checked so that linear constraints ($\sum_{j=1}^J w_j = 1$) and bounds are satisfied.

To control the experimental time, GA is set to run for 5 min. This is repeated two times as prior experiments found similar results from various runs. The weight is selected based on the lowest error on the validation set.

J. Contribution of a Sensor in the Network

We use two techniques to investigate the importance of the sensor, i.e., mutual information (MI) to measure the importance of the sensor to the classification and Clamping to measure the importance of the sensor within the model.

- 1) *Mutual information* [20]: MI is based on information theory. It is used for defining the dependency between variables. Given two variables, x, y , the MI can be calculated as $I(x; y) = \int \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy$.
- 2) *Clamping* [16]: MLP is constructed using several sensors based on the feature selection process. Features of each sensor are substituted using their mean values. If the sensor is important in the network, removing it would result in lower network performance. Assuming all features within a sensor give equal significance, the contribution of a particular sensor is $con(S) = 1 - \frac{g(F|S=\bar{S})}{g(F)}$, where F is a set of features, S is the set of features of a particular sensor, $g(F|S=\bar{S})$ is the performance of the network when the values of S are substituted by their mean values, and $g(F)$ is the generalized performance of the network.

III. EXPERIMENTAL RESULTS

We collected data of 13 activities including brushing teeth, exercising, feeding, ironing, reading, scrubbing, sleeping, using stairs, sweeping, walking, washing dishes, watching TV, and wiping. Note that for exercise activity, the participants carry out exercise using an elastic stretching band. The project was approved by the Faculty of Computing, Engineering and Technology Academic Ethics Team, Staffordshire University, U.K. The participants were first interviewed on their gender, age, and health issue to evaluate their suitability for participation. Twelve participants were recruited for the study including two males and ten females aged 72.55 ± 4.321 years. The participants were asked to perform each activity for 10 min. They were allowed to perform the activities in any order and could take breaks during activities.

A total of 33.75 h of activity of elderly people data were collected. All missing data were removed. Also, to keep the balance between classes, sweeping floor activity data were removed as after removing missing data it only constitutes to 3% of the dataset. The data were preprocessed using WMA, where w_1 was set to 0.2 and w_2 was set to 0.8 in the experiment.

TABLE III
TOP TEN FEATURES

Source	Feature
Acceleration Y axis	Max, mean, median, min, mode, RMS
Norm gyro	RMS, mean
Acceleration Z axis	Min, mode

TABLE IV
FEATURES WITH MI OVER 75% QUARTILE

Sensor	Data	Feature
Accelerometer	X - axis	Max, RMS
	Y - axis	Mean, max, min, median, mode, RMS
	Z - axis	Mean, max, min, median, mode
	$\sqrt{X^2 + Y^2 + Z^2}$	Mean, max, median, intensity, RMS
Light	Light intensity	Max
Barometer	Temperature	Max
	Pressure	Max
Gyroscope	X - axis	STD, min, mode, intensity, RMS
	Y - axis	STD
	Z - axis	STD, max, intensity, RMS
	$\sqrt{X^2 + Y^2 + Z^2}$	Mean, STD, median, RMS

The data were segmented at 3.88 s with 50% overlapping. The dataset contains 39 328 samples. The features are calculated as mentioned in Section II-D. NaN and constant valued features were removed. To reduce the feature space, we examined the MI of each feature. Using a cutoff point at 3% of MI, the number of features is reduced from 185 to 141. All experiments in this study used tenfold cross validation where eight folds were used for training, one for validation, and one for testing. The data were randomly selected with equal class distributions. All experiments were repeated for ten runs.

A. Importance of Sensors and Suggested Features

The importance of sensors and features was analyzed using MI. The result shows that accelerometer was the most important sensor. Thirty-four percent of accelerometer features contained over third quartile of MI about the classes. Altimeter and temperature sensors were the least important sensors. Gyroscope, barometer, and light were also among the most important sensors containing useful information in classifying 12 activities. Accelerometer and gyroscopes produce the top ten MI (see Table III). MI of some of the features calculated from these sensors were in the third quartile or higher. Also, it can be seen in Table IV that the time-domain features provide more useful information than frequency-domain features. Maximum, RMS, mean, median, STD, mode, minimum, and intensity were the most important features, respectively.

The feature selection was performed using FC. The truncation at 24 features was selected as the accuracy started to remain constant. Features from accelerometer, altimeter, heart rate monitor, light, and barometer were selected. Also, 16 features were used to conform to previous study. Next, the contributions of sensors in our model (with 24 features) were investigated. The result shows that accelerometer was the most important sensor in the model. This is followed by altimeter, heart rate monitor (HR), barometer, gyroscope, and light, respectively. The top three features with the highest importance in the model were mean acceleration on Z-axis, maximum barometer pressure, and minimum altitude, respectively.

TABLE V
AVERAGE CLASSIFICATION RESULTS USING DIFFERENT CLASSIFIER FUSION METHODS AND FUSION WEIGHT FUNCTIONS

Fusion method	Accuracy (%)	Improve (%)	<BI (%)	=BI (%)	>BI (%)
Best individual	96.9662 ± 0.4158	0.0000	0.0000	100.0000	0.0000
Majority vote	97.1522 ± 0.4832	0.1918	7.0175	26.3158	66.6667
Product	97.2690 ± 0.4889	0.3028	6.1404	0.0000	93.8596
Sum	97.3096 ± 0.5030	0.3435	4.2105	0.0000	95.7895
Min	97.1419 ± 0.5040	0.1757	15.9649	0.5263	83.5088
Max	97.1506 ± 0.5169	0.1781	12.6706	0.1949	87.1345
Ranking	96.9364 ± 0.5984	-0.0298	30.6043	11.3060	58.0897
WA	97.1314 ± 0.5356	0.1652	10.8772	0.1754	88.9474

TABLE VI
AVERAGE CLASSIFICATION RESULTS USING DIFFERENT FUSION WEIGHT FUNCTIONS

Weight function	Accuracy (%)	Improve (%)	<BI (%)	=BI (%)	>BI (%)
SA	97.2078 ± 0.5391	0.2417	13.4503	0.0000	86.5497
VACO	97.2141 ± 0.5723	0.2479	6.1404	3.5088	90.3509
DMSFE-0.80	96.9824 ± 0.6428	0.0162	28.3626	0.5848	71.0526
DMSFE-0.85	97.0493 ± 0.6115	0.0831	26.0234	0.5848	73.3918
DMSFE-0.90	97.1451 ± 0.5780	0.1789	16.3743	0.5848	83.0409
DMSFE-0.95	97.2298 ± 0.5610	0.2636	12.5731	0.2924	87.1345
No weight	97.1974 ± 0.5428	0.2312	14.0351	0.0000	85.9649
Weighted accuracy	97.2924 ± 0.4870	0.3262	7.0175	0.0000	92.9825
GA-function	97.1178 ± 0.3346	0.1516	2.9240	12.5731	84.5029
GA-linear	97.1357 ± 0.3093	0.1695	0.5848	14.3275	85.0877

B. Classifier Fusion

Classification was performed using three algorithms with 16 and 24 features. In total, six classification models were produced which gave mean accuracy between 94.85% and 97.20% with STD between 0.3088 and 0.4186. As expected, SVM performance was superior to other algorithms. However, according to precision and recall of each classifier, some classifiers were better than SVM in some of the activities.

Next, classifier fusion was performed. Data from training and validation set were used to determine the weight for SA, VACO, DMSFE, and WACC techniques, whereas in GAFW, the training set was used in the fitness function and the validation set was used to select the weight. There are 57 possible combinations that were generated from the six classifiers. The results of the classifier fusion on the test data are presented in Tables III-B and VI. The classifiers fusion result is compared with the best individual classifier (BI) within the fusion group. The improvement column shows the percentage of mean difference between classifier fusion and BI. It can be seen that classifier fusion that utilized posterior probability achieved better results comparing to fusing the class output directly. Among seven classifier fusion methods, sum was the best fusion technique. It improved classification accuracy by 0.3435% on average comparing to using only the best individual classifier. 95.79% of all possible combinations using the sum method achieved equal or higher accuracy than using the best classifier. This is followed by product, MV, weighted average, max, min, and ranking, respectively. In terms of the fusion weight determination technique, in general, 98.25% of combination using GA achieved equal or higher accuracy than using one best classifier. VACO also achieved very good result of 93.86% accuracy equal to or higher than BI followed by WACC, DMSFE-0.95, SA, unit weight, DMSFE-0.90, DMSFE-0.85, and DMSFE-0.80, respectively.

TABLE VII
COMPUTATIONAL COST ON DIFFERENT FUSION WEIGHT FUNCTIONS AND DIFFERENT CLASSIFIER COMBINATION METHODS

Weight function	Cost (s)	Weight function	Cost (s)
GA sum	122.7658	VACO	0.00941
GA min	150.9691	DSMFE-0.80	0.1387
GA max	156.4008	DSMFE-0.85	0.1479
GA rank	71.9906	DSMFE-0.90	0.1378
GA prod	125.4562	DSMFE-0.95	0.1514
GA linear	119.0677	Unit weight	0.0001
SA	0.0105	WACC	0.1213

Fusion methods	Cost (s)	Fusion methods	Cost (s)
MV	0.012245	Max	0.000121
Product	0.001735	Ranking	0.001716
Sum	0.000847	WA	0.000067
Min	0.000115		

C. Computational Cost of Data Fusion

The computational cost of determining fusion weights using different method in Section II-H was calculated. The cost is based on the time used to find the weights for combining two classifiers using training dataset of 2 016 000 samples. For different classifier combination functions, the cost is based on using the function to combine the result of two classifiers per sample. The results are shown in Table VII.

IV. DISCUSSION

The result of the study indicates that accelerometer is the most important sensor for activity recognition. This confirms that accelerometer has ability of measuring human activity quantitatively, fast reaction to changes in movement and reflects type of activity well [6]. We also find that the new sensors introduced including gyroscope, barometer, and light contain useful information about human activities. Similar to accelerometer, gyroscope can reflect changes in activity well. We also observe that data obtained from gyroscope are similar to those from accelerometer. Barometer and light can be used to differentiate activities such as using stairs and sleeping.

Interestingly, although gyroscope, barometer, and light are shown to be very important sensors on their own, this is not the case when they are combined together. In our model of 24 feature selected using FC, only two gyroscope features are selected. Also, its contribution to the network is not as high as other sensors. This may be explained that although gyroscope is a good sensor on its own, when it is used with accelerometer, many of its features become redundant. The result also indicates that heart rate has significant contribution to the model. Using heart rate in the model increases the accuracy by 1.74%. The statistical tests showed that the improvement is significant ($p < 0.05$). This may be due to the fact that majority of activities studied in [12] are exercise-related activities, e.g., cycling, running, rowing, etc. Although the authors reported that heart rate help improve exercise activities, due to the similarity in these activities and large number of classes, the overall improvement is not as high as they expected. On the other hand, our study contains activities that are rather different, e.g., walking, sleeping, exercise; large difference in heart rate between these activities is expected, thus resulting in heart rate having a significant impact on our model.

Several classifier fusion and fusion weight techniques were investigated. The results show that the sum is the most effective fusion method and when used with SA, WACC, or GA, improvement on all combinations can be achieved. As the sum technique uses the average probability, the result is not heavily affected when some classifiers are over confident. On the contrary, the min method selects the class that has the minimum objection by all classifiers. As min is sensitive toward objection, it is affected when some inaccurate classifiers always produce low probability. Similarly for the max technique, if the system contains bad classifiers that produce high probability, the system accuracy is affected. The results show that ranking is the worst fusion method. Although ranking reduces the bias caused by some classifiers being over confident, converting probabilities into rank also loses some information. Thus, fusing classifiers by ranking could produce conflict or wrong prediction if there are many inaccurate classifiers in the group. The product technique shows the best result when unit weight is used.

In terms of fusion weight techniques, we find that, in general, GA performs the best compared to others. The improvement over BI is significant ($p < 0.05$). Although this improvement is lower comparing to other techniques, the results show that by using GA-linear, 99.42% of the combination can achieve equal or better accuracy than using only the best classifier. Also, better performance is expected if GA converges. The results also revealed that using linear function as a fitness function can find better weights, especially for min and max classifier combination functions. However, GA with ranking fitness function is better when uses with ranking classifier combination. When we observed that the cases that used GA-linear function fail to improve the accuracy, it was found that the calculated weights were totally different. For example, GA-linear gave 0.3 and 0.7 weight, while GA with ranking gave 0.7 and 0.3 weight. This is because the class probability has been converted into ranks that have different data representation than that used in linear function. VACO also obtained very good results while using the low computational cost. The DMSFE technique performs the worst. The results revealed that β value nearer to 1 achieves better combination accuracy which is similar to previous work [13]. This is because DMSFE uses β to give different weights for each error which was not suitable in the classification problem.

The results of the study also showed that using GA to find the fusion weight uses a much higher computational cost than other functions especially when trying to optimize min and max functions. Therefore, the proposed GAFW should be appropriate in the activity recognition model that will be developed offline. For other system that needs to update the fusion weights in real time, other functions such as VACO and WACC should be used. For the classifier combination function, the computational cost is very low and can be applied in both online and offline applications.

The proposed GAFW can be applied with any systems aiming to combine multiple classifiers. This study has demonstrated that 98% of classifier fusion using GA achieves higher accuracy than using only the best classifier. While other fusion weight techniques cannot guarantee accuracy improvement, we show

that GAFW is a more suitable method for determining fusion weight regardless of which fusion techniques are used.

V. CONCLUSION

In this study, a multisensor dataset of 13 activities of elderly persons in a real home has been collected. A multisensor activity recognition framework using two-level fusion was proposed. The results indicate that gyroscope, barometer, light, and HR provide valuable information for activity recognition. Gyroscope and accelerometer exhibit similar characteristics and some of their features are overlapped. HR is useful when classifying activities which have diversity in heart rate data, e.g., walking, sleeping, and exercise. We studied the use of GA to find fusion weights. Unlike previous works, different fitness function were investigated and performances were compared with BI on all possible classifier combinations. The results show that for all possible classifier combinations and fusion methods, 98% of times GAFW can achieve higher or equal accuracy to the best classifier within the group.

ACKNOWLEDGMENT

The authors would like thank P. Ayumak, J. Tubtimsri, and S. Chernbumroong for their support on data collection.

REFERENCES

- [1] P. D. of the Department of Economic and S. A. of the United Nations Secretariat, "World population prospects: The 2012 revision," 2012.
- [2] S. Chernbumroong, S. Cang, A. Atkins, and H. Yu, "Elderly activities recognition and classification for applications in assisted living," *Expert Syst. Appl.*, vol. 40, no. 5, pp. 1662–1674, 2013.
- [3] Y. Nam and J. Park, "Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor," *IEEE J. Biomed. Health Inf.*, vol. 17, no. 2, pp. 420–426, Mar. 2013.
- [4] U. Maurer, A. Rowe, A. Smailagic, and D. Siewiorek, "Location and activity recognition using ewatch: A wearable sensor platform," in *Ambient Intelligence in Everyday Life*. Heidelberg, Germany: Springer-Verlag, 2006, pp. 86–102.
- [5] H. Gjoreski and M. Gams, "Activity/posture recognition using wearable sensors placed on different body locations," in *Proc. Int. Conf. Artif. Intell. Soft Comput.*, 2011.
- [6] J. Parkka, M. Ermes, P. Korpiainen, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 119–128, Jan. 2006.
- [7] A. Fleury, M. Vacher, and N. Noury, "SVM-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 274–283, Mar. 2010.
- [8] Y. Ohtaki, H. Inooka, K. Sagawa, A. Suzuki, X. Zhan, M. Okutsu, and R. Nagatomi, "Recognition of daily ambulatory movements utilizing accelerometer and barometer," in *Proc. 2nd IASTED Int. Conf. Biomech.*, Hawaii, USA, Aug. 2004, pp. 18–21.
- [9] Y.-J. Hong, I.-J. Kim, S. C. Ahn, and H.-G. Kim, "Mobile health monitoring system based on activity recognition using accelerometer," *Simul. Model. Practice Theory*, vol. 18, no. 4, pp. 446–455, 2010.
- [10] P. Barsocchi, "Position recognition to support bedsores prevention," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 1, pp. 53–59, Jan. 2013.
- [11] J. Booyens and G. Hervey, "The pulse rate as a means of measuring metabolic rate in man," *Can. J. Biochem. Physiol.*, vol. 38, no. 11, pp. 1301–1309, 1960.
- [12] E. Munguia Tapia, "Using machine learning for real-time activity recognition and estimation of energy expenditure," Ph.D. dissertation, Dept. Archit., Massachusetts Inst. Technol., Cambridge, MA, USA, 2008.
- [13] S. Cang, "A non-linear tourism demand forecast combination model," *Tourism Econ.*, vol. 17, no. 1, pp. 5–20, 2011.

- [14] S. Günter and H. Bunke, "Optimization of weights in a multiple classifier handwritten word recognition system using a genetic algorithm," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 3, no. 1, pp. 25–41, 2004.
- [15] Y. Maghsoudi, A. Alimohammadi, M. J. Valadan Zoej, and B. Mojaradi, "Weighted combination of multiple classifiers for the classification of hyperspectral images using a genetic algorithm," presented at the SPRS Commiss. VII, Mid-term Symp., Remote Sens.: Pixels Processes, Enschede, The Netherlands, 2006.
- [16] W. Wang, P. Jones, and D. Partridge, "Assessing the impact of input features in a feedforward neural network," *Neural Comput. Appl.*, vol. 9, no. 2, pp. 101–112, 2000.
- [17] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford Univ. Press, 1995.
- [18] C.-C. Chang and C.-J. Lin. (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* [Online]. 2, pp. 27:1–27:27. Software Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [19] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Mach. Learn.*, vol. 3, no. 2, pp. 95–99, 1988.
- [20] C. E. Shannon, "A mathematical theory of communication," *SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, Jan. 2001.



Saisakul Chernbumroong received the B.Eng. degree in industrial engineering from Chiang Mai University, Chiang Mai, Thailand, and the M.Sc. degree in computer science from the University of Hertfordshire, Hertfordshire, U.K. She is currently working toward the Ph.D. degree at the School of Design, Engineering and Computing, Bournemouth University, Bournemouth, U.K.

Her research interests include sensor-based activity recognition, RFID sensors, and intelligent system for smart health. Her current research is focused on

multisensor-based activity recognition and classification to support elderly care.



Shuang Cang received the B.Sc. (first class Hons.) degree in mathematics from Heilongjiang University, Harbin, China, the M.Sc. (with distinction) degree in applied mathematics from King's College London, U.K. and the Ph.D. degree in applied mathematics from the University of Abertay Dundee, Dundee, U.K.

She is currently a Senior Lecturer at the School of Tourism, Bournemouth University, Bournemouth, U.K. She worked in a U.K. leading Software Company for about two and half years. Then, she worked in the Department of Computer Sciences, Exeter University and University of Wales (Aberystwyth). She spent more than two years as a Senior Statistician/Senior Analyst in the U.K. Government Research Laboratory and U.K. Government Department, where she applied statistical and pattern recognition techniques to solve real and complex problems. She is currently managing three EU funded projects, Erasmus Mundus cLINK (€ 2.5M) and FUSION (€ 3.05M) projects, an EU Marie Curie IRSES RABOT project as the BU local coordinator. Her research interests include data mining, artificial intelligence, pattern recognition, multivariate statistics, forecasting, and segmentations.



Hongnian Yu received the B.Eng. degree in electrical and electronic engineering from Harbin Institute of Technology, Harbin, China, the M.Sc. degree in control engineering from Northeast Heavy Machinery Institute, Heilongjiang, China, and the Ph.D. degree in Robotics from King's College London, U.K.

He is currently a Professor of computer science. His research interests include robotics, wireless networked control systems, RFID and its applications, mobile computing, modeling, scheduling, planning, and simulations of large discrete event dynamic systems

with applications to manufacturing systems, supply chains, transportation networks and computer networks. He has published more than 200 research papers. He has held several research grants worth about three million pounds from EPSRC, the Royal Society, and the EU, AWM, as well as from industry. He has successfully completed two EU funded projects, Asia-Link project (Euro-Asia Collaborations and Networking in Information Engineering System Technology) and eLINK (east-west Link for Innovation, Networking and Knowledge exchange, 5.5 million Euro) project. Currently, he is supervising a large Erasmus Mundus project (FUSION, € 3.05M) as a project coordinator. He is also managing an EU Marie Curie IRSES RABOT project as a coordinator. He is also supervising two EU funded projects, Sustainable ETourism (2.5 million Euro) and cLINK (2.5 million Euro) as a local coordinator.

Prof. Yu was a General Chair of the International conference on Software Knowledge Information Management and Applications in 2006 and 2012, respectively, and is serving on various other conferences and academic societies.

Genetic Algorithm-Based Classifiers Fusion for Multisensor Activity Recognition of Elderly People

Saisakul Chernbumroong, Shuang Cang, and Hongnian Yu, *Member, IEEE*

Abstract—Activity recognition of an elderly person can be used to provide information and intelligent services to health care professionals, carers, elderly people, and their families so that the elderly people can remain at homes independently. This study investigates the use and contribution of wrist-worn multisensors for activity recognition. We found that accelerometers are the most important sensors and heart rate data can be used to boost classification of activities with diverse heart rates. We propose a genetic algorithm-based fusion weight selection (GAFW) approach which utilizes GA to find fusion weights. For all possible classifier combinations and fusion methods, the study shows that 98% of times GAFW can achieve equal or higher accuracy than the best classifier within the group.

Index Terms—Ambient intelligence, genetic algorithm (GA), neural networks, sensor fusion, smart homes, support vector machine (SVM).

I. INTRODUCTION

THE number of people aged 65 and over has increased significantly over the years. In 2050, the number will reach to 1.9 billion people [1]. The increase has a significant effect on health care. Issues such as high demand in long-term care, rise in health care cost, and ineffective and insufficient care are expected. One of the ways to lessen the issues is to promote home-based care. In this study, we develop an activity recognition model that can be used to recognize or detect an activity of a person. The detected activity can be used to provide information and intelligent services to health care professionals, carers, elderly people, and their families so that the elderly people can remain at home as long as possible.

In activity recognition, a vast variety of sensors have been explored and used such as accelerometer [2]–[4], gyroscope [5], light [4], [6], motion sensor [7], magnetometer [5], [6], microphone [4], [6], [7], barometer [8], temperature [2], [6], RFID [9], etc. Based on sensor types, activity recognition can be divided into two approaches. In the on-object sensor-based activity recognition approach, the sensors are attached to objects in inhabitant area. For example, sensors were installed on furniture to infer activities [10]. RFID was used to identify the objects to

infer detailed activities such as putting on lotion, taking photo, etc. [9]. The approach exploits the semantic relationships between objects and activities to classify activities. However, a large number of sensors are required which is an infeasible and time-consuming process; uncertainty of sensors such as false start and inability to detect object could also result in a poor recognition rate. An on-body sensor-based activity recognition approach, on the other hand, collects data, e.g., movement from sensors distributed on human body. Common on-body sensors include accelerometer and inertia sensor. The sensors are attached to body locations such as chest [10], wrist [2], waist [3], etc. However, sensors are required to be worn all the time which may interrupt or reduce mobility of a user or even obstruct daily activities routine. Especially in elderly care applications, the sensors may be perceived as stigmatized. It is important that the activity recognition model is highly accurate and practical. Recent work [2], [4] showed an activity recognition model, which is practical and highly accurate based on wrist-worn sensors. Accelerometer, temperature sensor, and altimeter were used and 90% accuracy was achieved.

A limited number of studies have been carried out on wrist-worn-based activity recognition. For example, multisensor wrist-worn equipment was used to detect walking, walking upstairs, walking downstairs, sitting, and running [4]. In this study, we use accelerometer, temperature sensor, and altimeter as appeared in [2]. In addition, heart rate monitor, barometer, and light sensor are investigated. Heart rate can be used to measure physical activities indirectly as heart rate is proportional to the intensity of movement and oxygen supplied to skeletal muscles [11]. A combination of acceleration and heart rate improves the accuracy of estimation of energy expenditure by 1.4% [12]. However, the study concluded that the small improvement was not worth it as the user needs to wear the heart rate monitor at all times. Accelerometer and barometer (air pressure differential) were used to detect ambulatory movements considering vertical position shifts [8]. Combining barometer and accelerometer improved classification accuracy in child activities [3].

Temperature could be used to indicate changes in environment when performing certain activities, e.g., washing dishes and brushing teeth may involve a use of water or when ironing, the temperature maybe higher than normal. Several works used the temperature sensor as part of their activity recognition systems [6], [7]. For example, the difference of temperature of 15 min was used to determine the use of a shower [7]. Work in [4] and [6] used light sensors as part of their activity recognition systems. Gyroscope can be used to estimate the orientation and rotation of the movement. After gyroscope and magnetometer were added to the accelerometer, the accuracy was increased

Manuscript received September 4, 2013; revised January 17, 2014; accepted March 12, 2014. Date of publication; date of current version. (*Corresponding author: H. Yu.*)

S. Chernbumroong and H. Yu are with the Faculty of Science and Technology, Bournemouth University, Poole, Dorset BH12 5BB, U.K. (e-mail: schernbumroong@bournemouth.ac.uk; yuh@bournemouth.ac.uk).

S. Cang is with the School of Tourism, Bournemouth University, Poole, Dorset BH12 5BB, U.K. (e-mail: scang@bournemouth.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JBHI.2014.2313473

by 17% [5]. From the literature, it can be seen that combining these sensors could improve classification accuracy. However, these sensors have not yet been combined on the wrist location. This study investigates the contribution of different sensors in the system and what features should be extracted from these sensors.

Many machine learning algorithms such as multilayer perceptron [2], [6], support vector machine (SVM) [2], [3], [7], [12], and decision tree [3], [6], [12] are used in activity classification. Previous studies also showed that combining multiple classifiers help improve classification accuracy. An investigation on several techniques for classifier fusion and fusion weights is carried out. Fusion strategies explored includes majority voting, product, sum, min, max, ranking, and weighted average. Furthermore, since some classifiers may perform better than others, it is sensible to incorporate weights to the classifiers. Fusion weight functions including weighted accuracy (WACC) and techniques used in forecasting domain [13], i.e., simple average (SA), variance-covariance (VACO), and discounted mean square forecast error (DMSFE) are investigated. These techniques offer easy weight determination. However, they may not yield the optimal results for all fusion methods as they are deterministic.

We use genetic algorithm (GA) to determine the fusion weights. Studies indicated GA could improve the classifier fusion accuracy [14], [15]. For example, classifier combination using eight to ten ensembles generated from different techniques was studied in [14]. Weight combination using GA to combine several Bayesian classifiers was investigated in [15]. However, there are some limitations on these studies. First, most of them focused only on the fusion of all classifiers. For example, they produced six classifiers then used GA to combine them. Based on this, the conclusion that GA could improve classifier combination accuracy is not always true as all possible combinations have not been tested. Second, fitness functions such as a function that reflects on the classifier combination functions, e.g., sum, min, max, product, ranking, and weighted average have not been investigated before. Finally, some of these results were often compared with the mean accuracy of a set of classifiers rather than to the best individual classifier. The mean accuracy is always equal to or less than the accuracy of the best individual classifier (equal accuracy is only occurred if and only if all classifiers have the same accuracy). For example, if there are three classifiers with 90%, 85%, 95%, the mean accuracy is 90% which is less than the best individual (95%). This weakens the conclusion that the classifier combination is better than a single classifier.

In this study, we collected a multisensor dataset of 13 activities of elderly persons in a real home. We propose a multisensor activity recognition which fuses information at feature and classifier levels. We investigate the use of GA for fusion weight that addresses the previous limitations.

II. SYSTEM ARCHITECTURE AND ALGORITHMS

A. Multisensor Activity Recognition Framework

Data from multisensor are fused at feature and classifier levels. The reason for performing two-level data fusion is that some of the sensors are not useful in recognizing activities by them-

selves, e.g., altimeter and temperature. Therefore, these sensors should be combined at feature level so that they can be used to provide more information for the classification. Second, results from literature works showed that there is no best classifier for all tasks. Therefore, fusing the data at classifier level would improve the classification accuracy.

Here, we describe the multisensor activity recognition platform (see Fig. 1). First, the system receives data from multiple sensors, the framework raw input is defined as $\{x_i, y_i\}$, where $x = S_1, S_2, \dots, S_n$ and y is the output of K possible activities. The raw inputs are then preprocessed using the weighted moving average (WMA). WMA is used to smoothen the signal using $A_t = w_1 A_t + w_2 A_{t-1}$ where A is the signal at time t . Next, the data are scaled to [0 1] range. The feature set F is extracted (see Section II-D) and fed into feature selection process using the feature combination (FC) technique (see Section II-E) resulting in a feature set S . This is the first level of data fusion where features from different sensors are combined and used in a classifier. The feature set S is passed through a multiclassification block (see Section II-F) which produces class posterior probability $\hat{P}_i^{(j)}$. This is the second level of data fusion where decisions from multiple classifiers are combined. The classifiers are combined and the fusion weights are determined using GA (see Sections II-I and II-G) to produce the final prediction.

B. Multisensor Platform

Accelerometer, temperature sensor, and altimeter are embedded on the CC430F6137 Microcontroller with the MSP430 CPU from Texas Instruments. The accelerometer can measure 3-D acceleration at a range of up to $\pm 2G$ ($G = 9.81 \text{ m/s}^2$) with a resolution of 18 mG. The pressure sensor's measuring range is of 30–120 kPa with 6 Pa resolution. Gyroscope, barometer, and light sensor are implemented on Gadgeteer FEZ Cerberus board with the 168-MHz 32 bit Cortex M4 processor. The gyroscope can measure up to $\pm 2000^\circ/\text{s}$ and 14.375 LSBs per $^\circ/\text{s}$ sensitivity. The Barometer measures the 300–1100 hPa absolute Pressure Range. We used the heart rate monitor chest strap from BlueRobin. The accelerometer and gyroscope are sampling at 33 Hz, while other sensors are sampling at 1 Hz.

A heart rate monitor is worn over the chest using the chest strap. Accelerometer, gyroscope, light, and barometer sensors are worn on the dominant wrist and temperature sensor and altimeter are worn on the other wrist. The sensors are separated between the two wrists due to the limitation on the hardware. The separation is designed in such a way that it should not interfere with the activity recognition. The sensors that are related to the movement, i.e., accelerometer and gyroscope, are worn on the dominant wrist in order to capture the activity's movement. Also, barometer and light sensors are also worn on the dominant wrist as they are parts of the Microsoft Gadgeteer platform. The temperature sensor that captures the body temperature and altimeter are worn on the nondominant wrist. Acceleration and heart rate are transmitted wirelessly to PC over 868 MHz. Temperature and altitude are stored on a 8-kB flash on the watch,

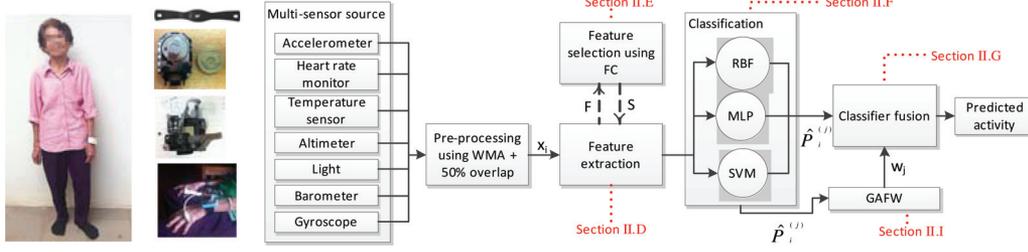


Fig. 1. Multisensor activity recognition of an elderly person framework.

while data from gyroscope, barometer, and light sensors are stored on a 2-GB SD card.

C. Application of the Proposed Framework Scenario

In this section, we provide a scenario of how the proposed framework will be used in detecting activities of an elderly person in home. The proposed activity recognition model is to be developed offline and stored on the PC of the elderly person's home. In real applications, we expect that the sensors will be embedded on a single watch. The elderly person wears the watch that contains multiple sensors on the dominant wrist. The sensor data from the watch are sent wirelessly to the PC. The activity recognition model performs a classification and stores the results on PC. The predicted activity is encrypted and sent over the Internet to the relevant stakeholders such as health professionals, carer, and families. The stakeholders' PCs or mobiles should contain specialized models for further analysis, e.g., to generate behavior pattern for health professionals, to show a day activity report for family members, etc. The detected activity can also be used to provide services in homes. For example, if the sleeping activity is detected, the bedroom's temperature and light can be automatically adjusted.

D. Feature Extraction

Twelve raw sensor data are extracted including 3-D acceleration, heart rate, temperature, altitude, light, barometer temperature, barometer pressure, and 3-D rotation. Also, the magnitudes $\sqrt{x^2 + y^2 + z^2}$ of both acceleration and rotation are calculated. In total, there are 14 input data. For each input, features including mean, standard deviation (STD), maximum, minimum, median, mode, kurtosis, skewness, intensity, difference, root-mean-square (RMS), energy, entropy, and key coefficient are extracted. The key coefficient is the summation of the signal coefficients from 0.5 to 3 Hz. Also, correlations between each acceleration axis and gyroscope axis are calculated. In total, 202 features are extracted.

E. Feature Selection Algorithm

In this study, we use the FC proposed in [2]. FC that is based on neural network and Clamping technique [16] measures the impact of the clamped features within the network and selects a group of features which as a whole achieve the best result. Given $S = \{\}$ where S is a set of selected features, $F = \{f_1, f_2, \dots, f_N\}$ where F is a set of N features, and $g()$ is

the generalized performance of the network. The FC algorithm is as follows:

- 1) First calculate the feature importance for all features f_i in the feature space F using $Im(f_i) = 1 - \frac{g(F|f_i=\bar{f}_i)}{g(F)}$.
 - 2) Select the feature f_s in F which has the maximum impact $f_s = \max_{f_i \in F} Im(f_i)$.
 - 3) If and only if $g(S \cup f_s) \geq g(S)$, then update S and F using $S = S \cup \{f_s\}$ and $F = F \setminus \{f_s\}$.
- Repeat steps 2 and 3 for $N - 1$ times.

F. Classification Algorithm

This study focuses on three algorithms which are widely used in the sensor-based activity recognition research.

- 1) *Multilayer perceptron neural network (MLP)*: MLP uses the concept of connectionist where several input nodes are connected with associated weights to several outputs nodes. The network output can be calculated from the summation function $o_i = \phi(\sum_i W_i x_i)$, where W_i is the weight used for adjusting input x_i and ϕ is the activation function [17]. MLP learns the classification error through a back propagation algorithm and tries to find the weights to minimize that error.
- 2) *Radial basis function neural network (RBF)*: RBF [17] is a neural network that uses RBF as an activation function. For N hidden neurons, the activation function is $f(x) = \sum_{i=1}^N W_i \varphi(\|x - c_i\|)$, where c_i is the center vector for neuron i and φ is a kernel function.
- 3) *Support vector machine*: SVM [18] constructs decision boundaries by solving the optimization objective $\min_{W,b,\xi} \frac{1}{2} W^T W + C \sum_{i=1}^m \xi_i$ subject to $y_i(W^T f(x_i) + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$. The slack term ξ_i is used to relax the constraints allowing misclassified examples. The associated cost parameter C is used for penalizing ξ_i . $f()$ is a function which transforms the input x_i into a higher dimensional space. This study used an RBF kernel function $f(x_i) = \exp(-\frac{1}{(2\sigma^2)} \|x_i - x_j\|^2)$, where σ is the width of the Gaussian kernel. For multiclass classification, we constructed K binary classifiers and applied one-vs-all classification.

Posterior probability can be produced by cooperating some functions, e.g., softmax or by solving optimization function, i.e., $\min_p \frac{1}{2} \sum_{i=1}^k \sum_{j:j \neq i} (r_{ji} p_i - r_{ij} p_j)^2$, where r_{ij} is the pairwise class probabilities between class i and j and p_i is the probability of class i .

TABLE I
FUSION WEIGHT FUNCTION STUDIED IN THIS PAPER

Fusion function	Equation
MV	$pred_i = mode_J \{pred_i^{(j)}\}$
Product	$pred_i = \max_K \left\{ \frac{1}{p(C_k)} \prod_{j=1}^J (\hat{P}_{ik}^{(j)})^{w_j} \right\}$
Sum	$pred_i = \max_K \left\{ \frac{1}{J} \sum_{j=1}^J (\hat{P}_{ik}^{(j)})^{w_j} \right\}$
Min	$pred_i = \max_K \left\{ \frac{\min_J (\hat{P}_{ik}^{(j)})^{w_j}}{\sum_{k=1}^K \min_J (\hat{P}_{ik}^{(j)})^{w_j}} \right\}$
Max	$pred_i = \max_K \left\{ \frac{\max_J (\hat{P}_{ik}^{(j)})^{w_j}}{\sum_{k=1}^K \max_J (\hat{P}_{ik}^{(j)})^{w_j}} \right\}$
Ranking	$pred_i = \max_K \sum_{j=1}^J w_j rank_{ik}^{(j)}$
WA	$pred_i = \max_K \sum_{j=1}^J w_j \hat{P}_{ik}^{(j)}$

G. Multimodel Fusion Methods

In this study, seven classifier fusion methods which are widely used in the classifier combination context are investigated [14], [15]. The idea of the majority vote (MV) is to combine all the votes given by each models selecting the class which has the highest vote. For product method, the probabilities are combined using vector product. The product rule is more sensitive to objection than support where the class with low probability is more influenced to the decision than the class with high probability. For the sum technique, the probabilities are combined using sum function and the class which has the highest maximum average probability is selected. The sum function generates the result based on the average decisions of all classifiers which is similar to majority voting. However, the sum method utilizes class probabilities. The maximum method decides the result based on the most confident classifier where it selects the class with the highest probability from all the models. The min method combines classifiers' results by selecting the class which is least objection by all the models. For the ranking method, first the probability $\hat{P}_i^{(j)}$ is converted to ranks where the maximum rank score is K and minimum is 1. The class with the maximum rank is selected. Weighted Average (WA) associate classifiers' decisions with weights. WA is the same as sum when the weights are 1.

Given that $pred_i^j$ is the prediction of input x_i using classifier model j , $\hat{P}_{ik}^{(j)}$ is the posterior probability that x_i belongs to class k and w_j is the weight for classifier model j , the prediction of the multi-model fusion can be calculated using equations in Table I. In case of equal scores, the model selects the result based on the best classifier.

H. Fusion Weight

Since each classification model may be superior to others, it is common to incorporate weights to the models to reflect this. Six weight functions are studied. SA gives the average weights to all classifiers. VACO uses the mean square error to calculate the weights. In this study, we modified the VACO equation to suit a classification problem by utilizing class probabilities. DMSFE is the modified version of VACO where a parameter β is used to discount weights of the instances. Unit weight gives all classifier weights 1 which means all classifiers are associated with weights. WACC uses the weighted accuracy of each model as the weights. Note that all calculated weights must

TABLE II
FUSION WEIGHT FUNCTION STUDIED IN THIS PAPER

Fusion weight	Equation
SA	$w_j = \frac{1}{J}$
VACO	$w_j = \frac{(\sum_{i=1}^m (1 - \hat{P}_{iK}^{(j)})^{-1})^{-1}}{\sum_{j=1}^J (\sum_{i=1}^m (1 - \hat{P}_{iK}^{(j)})^{-1})^{-1}}$
DMSFE	$w_j = \frac{(\sum_{i=1}^m \beta^{m-i+1} (1 - \hat{P}_{iK}^{(j)})^{-1})^{-1}}{\sum_{j=1}^J (\sum_{i=1}^m \beta^{m-i+1} (1 - \hat{P}_{iK}^{(j)})^{-1})^{-1}}, \beta = [0.95, 0.9, 0.85, 0.80]$
Unit weight	$w_j = 1$
WACC	$w_j = \frac{Accuracy^{(j)}}{\sum_{j=1}^J Accuracy^{(j)}}$

be summed to 1, i.e., $\sum_{j=1}^J w_j = 1$. This is except for the unit weight function where all the weights are 1.

Given m training examples and J models, and $\hat{P}_{iK}^{(j)}$ is the probability that model j predicts that data x_i belongs to class K , given that the true class is K , the weight for each classifier model j can be calculated using equations in Table II.

I. Genetic Algorithm Based Fusion Weight (GAFW)

In this study, we propose to use GA to find weights for classifiers. GA [19] has been commonly used to solve an optimization problem. The advantage of GA over other optimization techniques is that instead of starting at a single point to find the solution, a population of points is created. It mimics natural selection in which the population is modified over time. Individuals are randomly selected as parents to produce children of the next generation.

1) *Fitness Function*: GA is used to find the weights that minimize the mean square of the combination error. The classification error is defined as follows:

$$error_i = \begin{cases} 1, & \text{if } tru_i = pred_i \\ 0, & \text{otherwise.} \end{cases}$$

We propose to use the fitness function (ff) according to the fusion method. Given the fusion method (fm) as any function described in Section II-G, the fitness function is defined as

$$ff(w_j) = \frac{1}{2m} \sum_{i=1}^m error(tru_i, fm_i(w_j))^2.$$

The linear weight fitness function (GA-Linear) is also explored:

$$ff(w_j) = \frac{1}{2m} \sum_{i=1}^m error \left(tru_i, \max_K \left\{ \frac{1}{J} \sum_{j=1}^J (\hat{P}_{ik}^{(j)}) * w_j \right\} \right)^2.$$

2) *Population Initialization*: The weight for each classifier is represented in each bit of a chromosome. For each combination, we have J bits. Each bit is represented by a real number between 0 and 1. In order to make sure that the weight obtained will result in higher classification accuracy, a population that covers the search space and near a possible optimum point is necessary. We propose to use the following strategy to initialize the population. First, one of the populations must contain weighted average accuracy chromosomes. Second, the weights are randomly generated from a uniform distribution and the highest weight is assigned to the best model. Note that the weight for the best model within the group is generated randomly between

$\frac{1}{J}$ and 1. The initial population of $20 \times J$ chromosomes are generated.

3) *Crossover and Mutation*: The crossover rate is set to 0.8, same as used in [15]. Adaptive mutation is used where it randomly generates directions that are adaptive with respect to the last generation. The feasible region is bounded by the constraint ($0 \leq w_j \leq 1$). A mutant is checked so that linear constraints ($\sum_{j=1}^J w_j = 1$) and bounds are satisfied.

To control the experimental time, GA is set to run for 5 min. This is repeated two times as prior experiments found similar results from various runs. The weight is selected based on the lowest error on the validation set.

J. Contribution of a Sensor in the Network

We use two techniques to investigate the importance of the sensor, i.e., mutual information (MI) to measure the importance of the sensor to the classification and Clamping to measure the importance of the sensor within the model.

- 1) *Mutual information* [20]: MI is based on information theory. It is used for defining the dependency between variables. Given two variables, x, y , the MI can be calculated as $I(x; y) = \int \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy$.
- 2) *Clamping* [16]: MLP is constructed using several sensors based on the feature selection process. Features of each sensor are substituted using their mean values. If the sensor is important in the network, removing it would result in lower network performance. Assuming all features within a sensor give equal significance, the contribution of a particular sensor is $con(S) = 1 - \frac{g(F|S=\bar{S})}{g(F)}$, where F is a set of features, S is the set of features of a particular sensor, $g(F|S=\bar{S})$ is the performance of the network when the values of S are substituted by their mean values, and $g(F)$ is the generalized performance of the network.

III. EXPERIMENTAL RESULTS

We collected data of 13 activities including brushing teeth, exercising, feeding, ironing, reading, scrubbing, sleeping, using stairs, sweeping, walking, washing dishes, watching TV, and wiping. Note that for exercise activity, the participants carry out exercise using an elastic stretching band. The project was approved by the Faculty of Computing, Engineering and Technology Academic Ethics Team, Staffordshire University, U.K. The participants were first interviewed on their gender, age, and health issue to evaluate their suitability for participation. Twelve participants were recruited for the study including two males and ten females aged 72.55 ± 4.321 years. The participants were asked to perform each activity for 10 min. They were allowed to perform the activities in any order and could take breaks during activities.

A total of 33.75 h of activity of elderly people data were collected. All missing data were removed. Also, to keep the balance between classes, sweeping floor activity data were removed as after removing missing data it only constitutes to 3% of the dataset. The data were preprocessed using WMA, where w_1 was set to 0.2 and w_2 was set to 0.8 in the experiment.

TABLE III
TOP TEN FEATURES

Source	Feature
Acceleration Y axis	Max, mean, median, min, mode, RMS
Norm gyro	RMS, mean
Acceleration Z axis	Min, mode

TABLE IV
FEATURES WITH MI OVER 75% QUARTILE

Sensor	Data	Feature
Accelerometer	X - axis	Max, RMS
	Y - axis	Mean, max, min, median, mode, RMS
	Z - axis	Mean, max, min, median, mode
	$\sqrt{X^2 + Y^2 + Z^2}$	Mean, max, median, intensity, RMS
Light	Light intensity	Max
Barometer	Temperature	Max
	Pressure	Max
Gyroscope	X - axis	STD, min, mode, intensity, RMS
	Y - axis	STD
	Z - axis	STD, max, intensity, RMS
	$\sqrt{X^2 + Y^2 + Z^2}$	Mean, STD, median, RMS

The data were segmented at 3.88 s with 50% overlapping. The dataset contains 39 328 samples. The features are calculated as mentioned in Section II-D. NaN and constant valued features were removed. To reduce the feature space, we examined the MI of each feature. Using a cutoff point at 3% of MI, the number of features is reduced from 185 to 141. All experiments in this study used tenfold cross validation where eight folds were used for training, one for validation, and one for testing. The data were randomly selected with equal class distributions. All experiments were repeated for ten runs.

A. Importance of Sensors and Suggested Features

The importance of sensors and features was analyzed using MI. The result shows that accelerometer was the most important sensor. Thirty-four percent of accelerometer features contained over third quartile of MI about the classes. Altimeter and temperature sensors were the least important sensors. Gyroscope, barometer, and light were also among the most important sensors containing useful information in classifying 12 activities. Accelerometer and gyroscopes produce the top ten MI (see Table III). MI of some of the features calculated from these sensors were in the third quartile or higher. Also, it can be seen in Table IV that the time-domain features provide more useful information than frequency-domain features. Maximum, RMS, mean, median, STD, mode, minimum, and intensity were the most important features, respectively.

The feature selection was performed using FC. The truncation at 24 features was selected as the accuracy started to remain constant. Features from accelerometer, altimeter, heart rate monitor, light, and barometer were selected. Also, 16 features were used to conform to previous study. Next, the contributions of sensors in our model (with 24 features) were investigated. The result shows that accelerometer was the most important sensor in the model. This is followed by altimeter, heart rate monitor (HR), barometer, gyroscope, and light, respectively. The top three features with the highest importance in the model were mean acceleration on Z-axis, maximum barometer pressure, and minimum altitude, respectively.

TABLE V
AVERAGE CLASSIFICATION RESULTS USING DIFFERENT CLASSIFIER FUSION METHODS AND FUSION WEIGHT FUNCTIONS

Fusion method	Accuracy (%)	Improve (%)	<BI (%)	=BI (%)	>BI (%)
Best individual	96.9662 ± 0.4158	0.0000	0.0000	100.0000	0.0000
Majority vote	97.1522 ± 0.4832	0.1918	7.0175	26.3158	66.6667
Product	97.2690 ± 0.4889	0.3028	6.1404	0.0000	93.8596
Sum	97.3096 ± 0.5030	0.3435	4.2105	0.0000	95.7895
Min	97.1419 ± 0.5040	0.1757	15.9649	0.5263	83.5088
Max	97.1506 ± 0.5169	0.1781	12.6706	0.1949	87.1345
Ranking	96.9364 ± 0.5984	-0.0298	30.6043	11.3060	58.0897
WA	97.1314 ± 0.5356	0.1652	10.8772	0.1754	88.9474

TABLE VI
AVERAGE CLASSIFICATION RESULTS USING DIFFERENT FUSION WEIGHT FUNCTIONS

Weight function	Accuracy (%)	Improve (%)	<BI (%)	=BI (%)	>BI (%)
SA	97.2078 ± 0.5391	0.2417	13.4503	0.0000	86.5497
VACO	97.2141 ± 0.5723	0.2479	6.1404	3.5088	90.3509
DMSFE-0.80	96.9824 ± 0.6428	0.0162	28.3626	0.5848	71.0526
DMSFE-0.85	97.0493 ± 0.6115	0.0831	26.0234	0.5848	73.3918
DMSFE-0.90	97.1451 ± 0.5780	0.1789	16.3743	0.5848	83.0409
DMSFE-0.95	97.2298 ± 0.5610	0.2636	12.5731	0.2924	87.1345
No weight	97.1974 ± 0.5428	0.2312	14.0351	0.0000	85.9649
Weighted accuracy	97.2924 ± 0.4870	0.3262	7.0175	0.0000	92.9825
GA-function	97.1178 ± 0.3346	0.1516	2.9240	12.5731	84.5029
GA-linear	97.1357 ± 0.3093	0.1695	0.5848	14.3275	85.0877

B. Classifier Fusion

Classification was performed using three algorithms with 16 and 24 features. In total, six classification models were produced which gave mean accuracy between 94.85% and 97.20% with STD between 0.3088 and 0.4186. As expected, SVM performance was superior to other algorithms. However, according to precision and recall of each classifier, some classifiers were better than SVM in some of the activities.

Next, classifier fusion was performed. Data from training and validation set were used to determine the weight for SA, VACO, DMSFE, and WACC techniques, whereas in GAFW, the training set was used in the fitness function and the validation set was used to select the weight. There are 57 possible combinations that were generated from the six classifiers. The results of the classifier fusion on the test data are presented in Tables III-B and VI. The classifiers fusion result is compared with the best individual classifier (BI) within the fusion group. The improvement column shows the percentage of mean difference between classifier fusion and BI. It can be seen that classifier fusion that utilized posterior probability achieved better results comparing to fusing the class output directly. Among seven classifier fusion methods, sum was the best fusion technique. It improved classification accuracy by 0.3435% on average comparing to using only the best individual classifier. 95.79% of all possible combinations using the sum method achieved equal or higher accuracy than using the best classifier. This is followed by product, MV, weighted average, max, min, and ranking, respectively. In terms of the fusion weight determination technique, in general, 98.25% of combination using GA achieved equal or higher accuracy than using one best classifier. VACO also achieved very good result of 93.86% accuracy equal to or higher than BI followed by WACC, DMSFE-0.95, SA, unit weight, DMSFE-0.90, DMSFE-0.85, and DMSFE-0.80, respectively.

TABLE VII
COMPUTATIONAL COST ON DIFFERENT FUSION WEIGHT FUNCTIONS AND DIFFERENT CLASSIFIER COMBINATION METHODS

Weight function	Cost (s)	Weight function	Cost (s)
GA sum	122.7658	VACO	0.00941
GA min	150.9691	DSMFE-0.80	0.1387
GA max	156.4008	DSMFE-0.85	0.1479
GA rank	71.9906	DSMFE-0.90	0.1378
GA prod	125.4562	DSMFE-0.95	0.1514
GA linear	119.0677	Unit weight	0.0001
SA	0.0105	WACC	0.1213

Fusion methods	Cost (s)	Fusion methods	Cost (s)
MV	0.012245	Max	0.000121
Product	0.001735	Ranking	0.001716
Sum	0.000847	WA	0.000067
Min	0.000115		

C. Computational Cost of Data Fusion

The computational cost of determining fusion weights using different method in Section II-H was calculated. The cost is based on the time used to find the weights for combining two classifiers using training dataset of 2 016 000 samples. For different classifier combination functions, the cost is based on using the function to combine the result of two classifiers per sample. The results are shown in Table VII.

IV. DISCUSSION

The result of the study indicates that accelerometer is the most important sensor for activity recognition. This confirms that accelerometer has ability of measuring human activity quantitatively, fast reaction to changes in movement and reflects type of activity well [6]. We also find that the new sensors introduced including gyroscope, barometer, and light contain useful information about human activities. Similar to accelerometer, gyroscope can reflect changes in activity well. We also observe that data obtained from gyroscope are similar to those from accelerometer. Barometer and light can be used to differentiate activities such as using stairs and sleeping.

Interestingly, although gyroscope, barometer, and light are shown to be very important sensors on their own, this is not the case when they are combined together. In our model of 24 feature selected using FC, only two gyroscope features are selected. Also, its contribution to the network is not as high as other sensors. This may be explained that although gyroscope is a good sensor on its own, when it is used with accelerometer, many of its features become redundant. The result also indicates that heart rate has significant contribution to the model. Using heart rate in the model increases the accuracy by 1.74%. The statistical tests showed that the improvement is significant ($p < 0.05$). This may be due to the fact that majority of activities studied in [12] are exercise-related activities, e.g., cycling, running, rowing, etc. Although the authors reported that heart rate help improve exercise activities, due to the similarity in these activities and large number of classes, the overall improvement is not as high as they expected. On the other hand, our study contains activities that are rather different, e.g., walking, sleeping, exercise; large difference in heart rate between these activities is expected, thus resulting in heart rate having a significant impact on our model.

Several classifier fusion and fusion weight techniques were investigated. The results show that the sum is the most effective fusion method and when used with SA, WACC, or GA, improvement on all combinations can be achieved. As the sum technique uses the average probability, the result is not heavily affected when some classifiers are over confident. On the contrary, the min method selects the class that has the minimum objection by all classifiers. As min is sensitive toward objection, it is affected when some inaccurate classifiers always produce low probability. Similarly for the max technique, if the system contains bad classifiers that produce high probability, the system accuracy is affected. The results show that ranking is the worst fusion method. Although ranking reduces the bias caused by some classifiers being over confident, converting probabilities into rank also loses some information. Thus, fusing classifiers by ranking could produce conflict or wrong prediction if there are many inaccurate classifiers in the group. The product technique shows the best result when unit weight is used.

In terms of fusion weight techniques, we find that, in general, GA performs the best compared to others. The improvement over BI is significant ($p < 0.05$). Although this improvement is lower comparing to other techniques, the results show that by using GA-linear, 99.42% of the combination can achieve equal or better accuracy than using only the best classifier. Also, better performance is expected if GA converges. The results also revealed that using linear function as a fitness function can find better weights, especially for min and max classifier combination functions. However, GA with ranking fitness function is better when uses with ranking classifier combination. When we observed that the cases that used GA-linear function fail to improve the accuracy, it was found that the calculated weights were totally different. For example, GA-linear gave 0.3 and 0.7 weight, while GA with ranking gave 0.7 and 0.3 weight. This is because the class probability has been converted into ranks that have different data representation than that used in linear function. VACO also obtained very good results while using the low computational cost. The DMSFE technique performs the worst. The results revealed that β value nearer to 1 achieves better combination accuracy which is similar to previous work [13]. This is because DMSFE uses β to give different weights for each error which was not suitable in the classification problem.

The results of the study also showed that using GA to find the fusion weight uses a much higher computational cost than other functions especially when trying to optimize min and max functions. Therefore, the proposed GAFW should be appropriate in the activity recognition model that will be developed offline. For other system that needs to update the fusion weights in real time, other functions such as VACO and WACC should be used. For the classifier combination function, the computational cost is very low and can be applied in both online and offline applications.

The proposed GAFW can be applied with any systems aiming to combine multiple classifiers. This study has demonstrated that 98% of classifier fusion using GA achieves higher accuracy than using only the best classifier. While other fusion weight techniques cannot guarantee accuracy improvement, we show

that GAFW is a more suitable method for determining fusion weight regardless of which fusion techniques are used.

V. CONCLUSION

In this study, a multisensor dataset of 13 activities of elderly persons in a real home has been collected. A multisensor activity recognition framework using two-level fusion was proposed. The results indicate that gyroscope, barometer, light, and HR provide valuable information for activity recognition. Gyroscope and accelerometer exhibit similar characteristics and some of their features are overlapped. HR is useful when classifying activities which have diversity in heart rate data, e.g., walking, sleeping, and exercise. We studied the use of GA to find fusion weights. Unlike previous works, different fitness function were investigated and performances were compared with BI on all possible classifier combinations. The results show that for all possible classifier combinations and fusion methods, 98% of times GAFW can achieve higher or equal accuracy to the best classifier within the group.

ACKNOWLEDGMENT

The authors would like thank P. Ayumak, J. Tubtimsri, and S. Chernbumroong for their support on data collection.

REFERENCES

- [1] P. D. of the Department of Economic and S. A. of the United Nations Secretariat, "World population prospects: The 2012 revision," 2012.
- [2] S. Chernbumroong, S. Cang, A. Atkins, and H. Yu, "Elderly activities recognition and classification for applications in assisted living," *Expert Syst. Appl.*, vol. 40, no. 5, pp. 1662–1674, 2013.
- [3] Y. Nam and J. Park, "Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor," *IEEE J. Biomed. Health Inf.*, vol. 17, no. 2, pp. 420–426, Mar. 2013.
- [4] U. Maurer, A. Rowe, A. Smailagic, and D. Siewiorek, "Location and activity recognition using ewatch: A wearable sensor platform," in *Ambient Intelligence in Everyday Life*. Heidelberg, Germany: Springer-Verlag, 2006, pp. 86–102.
- [5] H. Gjoreski and M. Gams, "Activity/posture recognition using wearable sensors placed on different body locations," in *Proc. Int. Conf. Artif. Intell. Soft Comput.*, 2011.
- [6] J. Parkka, M. Ermes, P. Korpiainen, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 119–128, Jan. 2006.
- [7] A. Fleury, M. Vacher, and N. Noury, "SVM-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 274–283, Mar. 2010.
- [8] Y. Ohtaki, H. Inooka, K. Sagawa, A. Suzuki, X. Zhan, M. Okutsu, and R. Nagatomi, "Recognition of daily ambulatory movements utilizing accelerometer and barometer," in *Proc. 2nd IASTED Int. Conf. Biomech.*, Hawaii, USA, Aug. 2004, pp. 18–21.
- [9] Y.-J. Hong, I.-J. Kim, S. C. Ahn, and H.-G. Kim, "Mobile health monitoring system based on activity recognition using accelerometer," *Simul. Model. Practice Theory*, vol. 18, no. 4, pp. 446–455, 2010.
- [10] P. Barsocchi, "Position recognition to support bedsores prevention," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 1, pp. 53–59, Jan. 2013.
- [11] J. Booyens and G. Hervey, "The pulse rate as a means of measuring metabolic rate in man," *Can. J. Biochem. Physiol.*, vol. 38, no. 11, pp. 1301–1309, 1960.
- [12] E. Munguia Tapia, "Using machine learning for real-time activity recognition and estimation of energy expenditure," Ph.D. dissertation, Dept. Archit., Massachusetts Inst. Technol., Cambridge, MA, USA, 2008.
- [13] S. Cang, "A non-linear tourism demand forecast combination model," *Tourism Econ.*, vol. 17, no. 1, pp. 5–20, 2011.

- [14] S. Günter and H. Bunke, "Optimization of weights in a multiple classifier handwritten word recognition system using a genetic algorithm," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 3, no. 1, pp. 25–41, 2004.
- [15] Y. Maghsoudi, A. Alimohammadi, M. J. Valadan Zoej, and B. Mojaradi, "Weighted combination of multiple classifiers for the classification of hyperspectral images using a genetic algorithm," presented at the SPRS Commiss. VII, Mid-term Symp., Remote Sens.: Pixels Processes, Enschede, The Netherlands, 2006.
- [16] W. Wang, P. Jones, and D. Partridge, "Assessing the impact of input features in a feedforward neural network," *Neural Comput. Appl.*, vol. 9, no. 2, pp. 101–112, 2000.
- [17] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford Univ. Press, 1995.
- [18] C.-C. Chang and C.-J. Lin. (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* [Online]. 2, pp. 27:1–27:27. Software Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [19] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Mach. Learn.*, vol. 3, no. 2, pp. 95–99, 1988.
- [20] C. E. Shannon, "A mathematical theory of communication," *SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, Jan. 2001.



Saisakul Chernbumroong received the B.Eng. degree in industrial engineering from Chiang Mai University, Chiang Mai, Thailand, and the M.Sc. degree in computer science from the University of Hertfordshire, Hertfordshire, U.K. She is currently working toward the Ph.D. degree at the School of Design, Engineering and Computing, Bournemouth University, Bournemouth, U.K.

Her research interests include sensor-based activity recognition, RFID sensors, and intelligent system for smart health. Her current research is focused on multisensor-based activity recognition and classification to support elderly care.



Shuang Cang received the B.Sc. (first class Hons.) degree in mathematics from Heilongjiang University, Harbin, China, the M.Sc. (with distinction) degree in applied mathematics from King's College London, U.K. and the Ph.D. degree in applied mathematics from the University of Abertay Dundee, Dundee, U.K.

She is currently a Senior Lecturer at the School of Tourism, Bournemouth University, Bournemouth, U.K. She worked in a U.K. leading Software Company for about two and half years. Then, she worked in the Department of Computer Sciences, Exeter University and University of Wales (Aberystwyth). She spent more than two years as a Senior Statistician/Senior Analyst in the U.K. Government Research Laboratory and U.K. Government Department, where she applied statistical and pattern recognition techniques to solve real and complex problems. She is currently managing three EU funded projects, Erasmus Mundus cLINK (€ 2.5M) and FUSION (€ 3.05M) projects, an EU Marie Curie IRSES RABOT project as the BU local coordinator. Her research interests include data mining, artificial intelligence, pattern recognition, multivariate statistics, forecasting, and segmentations.



Hongnian Yu received the B.Eng. degree in electrical and electronic engineering from Harbin Institute of Technology, Harbin, China, the M.Sc. degree in control engineering from Northeast Heavy Machinery Institute, Heilongjiang, China, and the Ph.D. degree in Robotics from King's College London, U.K.

He is currently a Professor of computer science. His research interests include robotics, wireless networked control systems, RFID and its applications, mobile computing, modeling, scheduling, planning, and simulations of large discrete event dynamic systems with applications to manufacturing systems, supply chains, transportation networks and computer networks. He has published more than 200 research papers. He has held several research grants worth about three million pounds from EPSRC, the Royal Society, and the EU, AWM, as well as from industry. He has successfully completed two EU funded projects, Asia-Link project (Euro-Asia Collaborations and Networking in Information Engineering System Technology) and eLINK (east-west Link for Innovation, Networking and Knowledge exchange, 5.5 million Euro) project. Currently, he is supervising a large Erasmus Mundus project (FUSION, € 3.05M) as a project coordinator. He is also managing an EU Marie Curie IRSES RABOT project as a coordinator. He is also supervising two EU funded projects, Sustainable ETourism (2.5 million Euro) and cLINK (2.5 million Euro) as a local coordinator.

Prof. Yu was a General Chair of the International conference on Software Knowledge Information Management and Applications in 2006 and 2012, respectively, and is serving on various other conferences and academic societies.