ASOC: 108655 Model 5G pp. 1–11 (col. figs: 6)

ARTICLE IN PRESS

Applied Soft Computing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Fuzziness based semi-supervised multimodal learning for patient's activity recognition using RGBDT videos

Muhammed J.A. Patwary [a], Weipeng Cao [b], Xi-Zhao Wang [b,*], Mohammad Ahsanul Haque [c]

[a] Department of Computer Science and Engineering, International Islamic University Chittagong, Bangladesh
[b] College of Computer Science and Software Engineering, Shenzhen University, 518060, China
[c] Data Scientist and Machine Learning Developer RadioAnalyzer, Denmark

ARTICLE INFO

Article history:
Received 24 July 2021
Received in revised form 20 December 2021
Accepted 19 February 2022
Available online xxxx

Keywords:
Fuzziness
Semi-supervised learning
Patient activity recognition
Multimodal
RGBDT
Neural networks

ABSTRACT

Automatic recognition of bedridden patients' physical activity has important applications in the clinical process. Such recognition tasks are usually accomplished on visual data captured by RGB, depth, and/or thermal cameras by utilizing supervised machine learning. However, supervised machine learning requires a large amount of labeled data and the accuracy depends on extracting appropriate features based on the domain knowledge. A plausible solution to these issues is using semi-supervised learning that focuses less on labeled data and domain knowledge. In this paper, a novel fuzziness-based semi-supervised multimodal learning algorithm, called (FSSL-PAR) is proposed for bedridden patient activity recognition. We use a synergistic interaction on RGB, Depth, and Thermal videos to assess the effect of visual multimodality for the first time in this semi-supervised learning setting. Experiments are conducted on a dataset collected by mimicking the patients with Acute Brain Injury (ABI) from a neurorehabilitation center. The results exhibit the superiority of the proposed algorithm over the existing supervised learning algorithms. We also see a positive correlation between the performance and the size of the labeled data in the proposed FSSL-PAR.

1. Introduction

During the last couple of decades body motion analysis has been very popular due to its widespread application areas, such as sleep analysis, breathing, epilepsy, activity monitoring [1–4] etc. Because of the incorporation of technology in the health care sector, automatic Patient's Activity Recognition (PAR) is considered to be an active research area for the last few years [5–7]. PAR is very important because it helps to render timely and useful services to the patient without any human intervention. It is also capable of providing the appropriate recommendation to the patient on the bed which substantially improves the quality of services in the hospital environment.

The concept of body motion analysis is relatively broad because it considers the movements of the body across different directions. Most of the motions are regarded as a representation of the configuration changes of the body. Useful information about body motion can be obtained by assessing the activity of the patients. Tracking of the patient's activity, when lying on a flat surface, is subject to the detection of either the presence or absence of both motion and non-motion states, specifically in the temporal dimensions.

Sometimes it is difficult to obtain the information of real patients with the help of technological devices, such as different sensors mainly due to patients' health conditions and safety. To solve this problem, Haque et al. [8], collected a dataset, where, based on the experiences of the health workers, such as doctors, nurses, and other persons related to the health care services, some healthy volunteers were asked to exhibit some predefined activities. Then a series of videos were collected from those healthy volunteers based on their activities to mimic the actual hospital environment for the analysis of patients' activity. In this work, we used this dataset to investigate the issue of Patient Activity Recognition (PAR).

Actigraphy [9] is another widely used method of tracking Patient's Activity (PA) in human beings where accelerometer-based techniques are utilized in analyzing both the rest and activity cycles. The method is also utilized in monitoring sleep cycles in patients in the Intensive Care Unit (ICU), patients sustaining severe brain injuries, and in recumbent persons. In addition to actigraphy, the use of video-based analysis has attracted a lot of interest for the last few decades considering that it is cheap, inherent unobtrusiveness, and ubiquity [10].

* Corresponding author.
E-mail addresses: mjap@iiuc.ac.bd (M.J.A. Patwary), caoweipeng@szu.edu.cn (W. Cao), xizhaowang@ieee.org, xwang@szu.edu.cn (X.-Z. Wang), mohammad@radioanalyzer.com (M.A. Haque).

https://doi.org/10.1016/j.asoc.2022.108655
1568-4946/

The methodological challenges are mostly encountered in the process of patient activity tracking especially from videos. Three common methods have been used for a long time, namely, block matching, frame differencing, and optical flow. Both the optical flow and block matching find use in estimating the local displacement of the patient's activity while frame differencing is used to detect specific areas where motion is involved. There are notable choice differences for video sensors when using video data like in the case of RGB, thermal, or depth. Resource-related challenges are encountered due to the lack of well-designed databases that can be used to evaluate the methodologies used in assessing patient activity accurately.

When a huge volume of annotated data is available then the supervised learning model is the right choice for the machine learning researchers to build a model for the patient's activity recognition purpose [11–13]. To this end, usually unimodal data such as RGB, depth, or thermal data is used to build the model. The main challenge of this technique is that, it requires a huge amount of labeled data which is very rare, expensive, and time-consuming to obtain. Another potential challenge of patient's activity recognition is that it requires proper feature extraction [14, 15] technique which needs sufficient domain knowledge. The synergistic interaction of RGB, depth and thermal data to form a multimodal RGBDT video data has been successfully applied in [8], but realizing the unavailability of a huge amount of labeled data in this study a novel fuzziness-based semi-supervised learning algorithm, also called FSSL-PAR, is proposed to investigate the recognition of patient's activity.

The concept of fuzziness (also called vagueness) is different from ambiguity and it is describing the difficulty of making sharp or precise distinctions of the data samples [16–18]. Wang et al. in [19] reported the relationship between the fuzziness of a classifier and the misclassification rate of the classifier on a group of samples. They derived the output of a classifier as a membership vector and showed that samples with higher fuzziness in the output vector exhibited a higher probability of misclassification. Whereas, lower fuzziness in the output vector indicates higher accuracy of classification. Therefore, while developing the proposed multi-step machine learning model FSSL-PAR, we utilized the concept of classifier's fuzziness in terms of the output vector in order to select the better initial classifiers based on low fuzziness value. The details of the approach are described in the methodology section. While semi-supervised learning scenarios is discarding the need for large labeled data and domain knowledge, using fuzziness helps decide the best initial supervised classifiers with the first few training data to be used in the further step of unsupervised learning.

The major contributions of our work are listed below:

(1) A novel fuzziness based semi-supervised multimodal learning approach, called FSSL-PAR, is proposed to investigate the recognition of patients' activity.
(2) Unlike traditional approaches that use either the RGB, thermal, or depth sensor-based unimodal data, multimodal RGBDT video data are utilized in this work to attain higher performance.
(3) Finding the correlation between the performance of the proposed algorithm and the increase of the number of initial labeled training data.

The rest of the paper is organized as follows. Section 2 gives the detailed background of a study with a synopsis of semi-supervised learning. Section 3 describes the methodology of the body motion analysis technique. Section 4 discusses the rationality of fuzziness-based SSL (F-SSL). Section 5 familiarizes our proposed methodology of fuzziness-based semi-supervised learning algorithm for patient's activity recognition (FSSL-PAR). Section 6 gives a detailed discussion about the experimental results. At last, Section 7 completes the paper with some concrete guidelines for future research.

2. Background study

2.1. Related works

Multi-modal video analysis has been very popular due to its widespread application areas, such as object recognition, face recognition, activity recognition etc [20,21]. In most of the multimodal video data analysis, RGB video data analysis has many successful applications [22]. Combination of RGB and depth, i.e., RGB-D video is also applied in many application [23,24]. Mark Hamer et al. studied the longitudinal patterns in physical activity and sedentary behavior of a patient [25]. Risto Telama [26] conducted a review to study the tracking of physical activity of a person from his/her childhood to adulthood. Risto reported the tracking of physical activity was dependent on sex of the patient. In [27], authors report the research outcome of physical activity recognition of a patient after his/her hip fracture. In [6], authors give a detailed survey of visioned-based patient monitoring system. The authors discussed the challenges of the state-of-the-art visioned-based patient monitoring system with their potential advantages and disadvantages. In [28], Avi Sadeh, provides the updated research output of sleep analysis also called actigraphy. In [9], Biren et al. come up with the recent research results of sleep analysis of patients in a medical intensive care unit (ICU). In [1], Adriene et al. conducted a study to detect robust and sensitive video motion analysis for sleeping pattern of a patient. In [29], the author showed how biometric information can be recognized from the facial video data. To this end, heartbeat signal is used.

The multimodal RGBDT video, where thermal information also used along with RGB-D information, is used for the first time by [8] to investigate the patient's body motion. Here, the authors used some shallow supervised machine learning techniques. To the best of our knowledge, in our proposed technique, fuzziness based semi-supervised learning has been used to investigate the RGBDT videos for the first time.

2.2. Scenario and categorization of the Patient's Activities (PAs)

This study aims to develop a tool or a system to observe patients with ABI admitted to HNRC, Denmark. This is a specialized center for patients with ABI. This tool can be used efficiently to assist the health workers to observe or specifically prescribe naturalistic motor activities without visible nursing. Patient's on bed with consciousness abnormalities can also be monitored with the help of this tool. This is not only helpful for the prognosis of patients but also for rendering with the appropriate services in a timely manner. To investigate the system, we primary considered a group of healthy volunteers and classified their probable body movements which are usually observed by the health workers when they observe the patients are lying on bed. Then five groups of movements are considered based on the body parts, such as foot, wrist, head, leg, and arm. These groupings are further multiplied by considering the body parts are once covered and then uncovered. The scenario is well described by the Fig. 1.

2.3. A synopsis of semi-supervised learning

The main concept of Semi-Supervised Learning (SSL) is explained in the following Fig. 2:

Fig. 1. Different activities exhibited by patients on bed

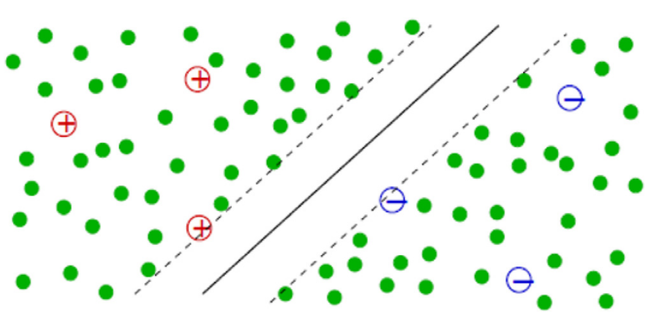

Fig. 2. The basic idea of SSL.

Suppose, the labeled and unlabeled data sets are respectively $(x_l, y_l) = \{x_{1:l}, y_{1:l}\}$ and $x_u = \{x_{l+1:n}\}$, where, x_l is the l^{th} sample and y_l is the corresponding label. A decision boundary (dotted line) is obtained when only labeled data is considered while a different decision boundary (solid line) is obtained considering both the labeled and unlabeled data.

SSL is a twofold technique. At its first step, a small amount of labeled data is used to build an initial classifier and in the second step that classifier is used to annotate huge amount of unlabeled data. Then these newly labeled data is added to the original initial classifier and retrain the model as long as the initial accuracy is improved or satisfy some specific stopping criteria.

2.3.1. Self-training

Self-training is possibly the oldest kind of semi-supervised learning method. The underlying idea of self-training is very simple. When there are a small quantity of labeled data is available while most of the data are unlabeled, self-training is a good choice as a learning algorithm. Self-training algorithm builds an initial classifier using a very small quantity of labeled data. Then, this initial classifier is used to give labels to the most unlabeled data in an iterative manner. It can also determine the ranks of the examples by confidence in their prediction and finally adds the most confident samples in the labeled dataset. The initial classifier is retrained with the help of the augmented training dataset and this process is repeated for a specific number of iterations or unless a given heuristic criteria has been satisfied. The accuracy of the classification improves only when all the classifiers are able to correctly label the given unlabeled data. In the real life applications, more neat and confident thresholds as well as several accuracy confident measures are used to restrain mislabeling of the given data. The main disadvantage of self-training is that it is sensitive to the outlier.

Self-training has been successfully applied in the following works: [30–32]

2.3.2. Co-training and multi-view learning

Some datasets can be seen from two different point of views. For example, a car be represented by either a picture or a bunch of text. Realizing the structure of the dataset, machine learning practitioners proposed another sort of SSL algorithm, known as co-training algorithm. Co-training uses two initial classifiers, whereas these two classifiers teaches one other to improve the performance of the classifiers. Co-training has been applied in the following works: [33,34]

Realizing the huge success of the co-training algorithm, machine learning researchers proposed multi-view learning. It is the natural extension of co-training algorithm. It uses more than two classifiers to teach each other to improve the overall performance. Multi-view learning has been used in the following application works: [35,36].

2.3.3. SSL with generative models

Generative model of SSL algorithm assumes that all the data follow the same parametric distribution, where the prior probability $p(y)$ is known. To calculate the conditional probability $p(x|y)$, $p(y)$ is used and it is also assumed that the number of components is known then the algorithm starts with an initial value of the parameters and iteratively fine tune the parameters and find the final value of the parameters. Here, labeled data are used to find the parameters. Based on those parameters class labels of unlabeled data are predicted. Following are some of the applications of generative models [37,38].

2.3.4. Semi-Supervised Support Vector Machines (S3VMs)

Let the training samples are partitioned into two disjoint subgroups: L for labeled sample and U for unlabeled sample. The main purpose of S3VM algorithm is to utilize the huge amount of unlabeled data for adjusting the boundary of decision which is basically created from a limited number of labeled data. That is, unlabeled data $U = x_{j_{j=1}}^{u}$ is used to construct the boundary from labeled data $L = (x_i, y_i)_{i=1}^{l}, y_i = \pm 1$ so that it traverse through the impenetrable area while preserving the labeled data accurately classified. S3VMs has been successfully applied in many applications. Such as follows: [39–41]

2.3.5. Graphs-based SSL

The central concept of a graph-based SSL algorithm is the existence of a graph G. This graph is consists of a set of vertices V and a set of edges E. Where, V denote the labeled and unlabeled training samples and E denote the linking between samples i and j. There is a weight, $w_{i,j}$ which indicates the closeness of samples x_i and x_j.

The graph-based semi-supervised learning method has been successfully applied in the following applications [42,43].

2.3.6. Fuzziness-based semi-supervised learning (FSSL)

The idea of fuzziness has been successfully exploited in the construction of SSL algorithm by Wang et al. [19] for the first time in 2015, since then it has been an active research area. As far as SSL algorithm construction is concerned, the accuracy of the initial classifier plays an important role. In [44], the authors studied the sensitivity analysis of the base classifier accuracy in a fuzziness-based SSL. The authors found that while the accuracy of the base classifier is between 70% and 80% then the maximum improvement of SSL algorithm is achieved. Many fuzziness measuring models can be used to estimate the fuzziness of a fuzzy set. Later, they [45] again studied the significance of the measures of fuzziness on the learning performance of SSL algorithm. Ashfaq et al. [46] used fuzziness-based SSL algorithm to detect intruder in an automatic system.

3. The detailed technique of body motion analysis

3.1. Body motion tracking

After eradication of noise, any existence of surface motion could be considered as a trigger to the existence of body motion [47]. For the categorization of 8 perceptual PA categories, we need to inspect local surface motions and the motion's intensity as well as global surface motion. The need to consider local motion and motion density was that the movements of different organs of the body can take place in the same block. For example, shifting of wrist and arm may occur in the same place. But in this case, the intensities are different. Thus, for the tracking of movement including intensity, a motion tracking approach which is based on point is introduced that includes a technique and a tracker namely GFT and LKT respectively, where GFT is used to identify proper pixels in the Region of Interest (ROI) and LKT is used to track those points [48]. The progression can be described as follows.

The pattern of pixel intensities varies on behalf of body shifting in successive video frames. An affine motion model was used to express the intensity change and also form a tracking algorithm. Assume two successive video frames as K and L. Then $K(X) = K(x,y)$ and $L(X) = L(x,y)$ represent magnitude of intensities of two images at (x,y) coordinate. The image point on a given frame K can be expressed as $[p_x, p_y]^T$ and feature tracking problem can be defined by tracking this point in the successive frame L. In the real case, the resemblance between the points in the two consecutive frames are measured by the window of pixels.

Thus, we can track a window of size $\omega_x \times \omega_y$ in the frame K to the frame L. This was done on the factor Θ by reducing f_{GFT} according to the following equation:

$$f_{GFT}(\Theta) = \sum_{x=p_x}^{p_x+\omega_x} \sum_{x=p_y}^{p_y+\omega_y} (I(x) - J(x+\Theta))^2 \tag{1}$$

Here, $(K(X)-L(X+\Theta))$ represents $(K(x,y)-L(x+\Theta_x, y+\Theta_y))$ and Θ was used as the image position x, and for tracking the variation of d played an important role even for small window size. Thus, the dislodgment of the same window could be varied. Under this situation to address this issue an affine motion field was introduced in [48] in the following way:

$$\Theta = \varphi x + \beta \tag{2}$$

$$\varphi = \begin{bmatrix} \varphi_{xx} & \varphi_{xy} \\ \varphi_{yx} & \varphi_{yy} \end{bmatrix}$$

where, φ is a deformation matrix and $\beta = [\beta_x, \beta_y]^T$ is the conversion of the feature windows. The tracing of a point between two images is determined by the 6 factors of φ and β. In [49], the authors introduced a lowering method of f_{GFT} through an LKT feature tracer [50]. Based on three factors namely, window size, the texturedness of the image frame, and the quantity of motion between frames the value of estimate by this tracker could be examined [3]. When we applied GFT in the whole ROI for tracking the points then a problem arose. The problem was that GFT automatically rejects low-textured areas as an awful feature pixel compare to its counterpart areas. To resolve this, the complete ROI was divided into (20×20) grid size and GFT was used to all the grid cells independently. As a result, we obtained some amount of feature points for tracing in every grid units defining fixed components of the body. Then LKT tracker was used to tracking the pixels of the consecutive video frames. A motion map having motions of higher intensity level was produced in ROI. This motion map contains three aspects. The first pole represents the video frame, the next pole comprises the traced feature points, and the last one contains the motion's intensity of the second feature over the first feature. To signify the strength of the body motion, the region of feature points are stored in the real map. Therefore, the discussed three poles represent respectively the temporal direction, the spatial direction and the intensity.

3.2. Different feature extraction techniques for the classification of activity

Haque [8], successfully extracted the following features (detailed discussion of how to extract these features can be found from the literature [8]).

(i) First Difference of the Motion Map (FDiff)
(ii) Principal Component Analysis (PCA)
(iii) Primitive Radon Features (PRadon)
(iv) Radon Distance Features (DiffRadon)

After the extraction of the features they successfully used those features in their shallow learning models such as KNN, LDA, SVM, DT, NB and GLM and obtained good results. In our proposed fuzziness-based semi-supervised learning algorithm we used the same set of features to investigate the patient's activity recognition problem.

4. Discussion on the rationality of F-SSL

In this section, we give some theoretical support for the rationality of the Fuzziness based Semi-Supervised Learning algorithm (F-SSL).

In practical applications, the true data distribution is usually unknown. To obtain a model with good generalization ability, the general method is to obtain the optimal parameters by minimizing the empirical error of the model on the limited training samples collected.

Therefore, given a training data set $\mathbf{D} = \{(x_1, y_1), \ldots, (x_N, y_N)\} \subset \mathbf{R}^d \times \mathbf{R}^m$, where N is the number of training samples, d is the dimension of the instance, and m is the number of classes. The training objective of machine learning algorithm can be expressed as follows:

$$\mathbf{f}_N = arg\ min \frac{1}{N} \sum_{i=1}^{N} \ell(y_i, f(x_i)) \tag{3}$$

where $\ell(\cdot)$ is the loss function and $f(x_i)$ is the prediction value of the model on the ith sample.

Let $R_n(f) = \frac{1}{N}\sum_{i=1}^{N} \ell(y_i, f(x_i))$, the VC dimension theory [51] states that as N approaches infinity, $\mathbf{f}_N$ can approach the true hypothesis function, that is,

$$\begin{aligned} arg\ minR_n(f) = \mathbf{f}_N \rightarrow \mathbf{f}^* = inf\ R(f) \\ s.t.\ N \rightarrow \infty. \end{aligned} \tag{4}$$

This inspired us that with the increase of the number of training samples, the generalization ability of the machine learning model will have a high probability to become better. In recent years, there are many theories and methods about how to augment the scale of the training set, such as performing the affine transformation on images and using Generative Adversarial Networks (GAN) to generate pseudo-samples.

The fuzziness based SSL (F-SSL) discussed in this paper is also a method to augment the valid training samples. Specifically, this method first uses a limited labeled training set X_l to train an initial classifier M_0, and then uses this classifier to predict the unlabeled samples set X_u. Based on the prediction results of the initial classifier, these unlabeled samples are divided into different clusters by using the fuzzy theory technology, for example, a cluster X_{u-H} composed of samples with high fuzziness and a cluster X_{u-L} composed of samples with low fuzziness. Then, according to a given threshold, the samples with high confidence are selected from X_{u-L} and then combined with their predicted labels to form a new labeled data set for augmenting the original training set. Now an augment training data set X_{l-A} is obtained.

According to the VC dimension theory [51], with the increase of the number of training samples, the performance of classifier is expected to become better. In other words, as $X_l \rightarrow X_{l-A}$, the prediction error of the model M_1 obtained by retraining with X_{l-A} will be lower than that of the M_0. This implies that the F-SSL strategy is feasible.

Moreover, the training process of F-SSL is iterative, that is, F-SSL iteratively uses the current classifier to predict unlabeled samples, then converts these samples to available labeled samples, and then retrains the classifier. Ideally, as the prediction ability of the current classifier gets better and better, its predictions for unlabeled samples become more and more accurate, the quality of the newly generated samples gets better and better, and the generalization ability of the final model is also expected to be better. Conversely, if the prediction ability of the initial classifier is not good, this iterative process may also cause the algorithm to fall into a vicious circle. Therefore, it is necessary to study the security guarantee technology of F-SSL. However, this issue is not the focus of our paper.

5. Proposed methodology of fuzziness based SSL for patient's activity recognition

The idea of fuzziness has been successfully used in the construction of our proposed algorithm. The basic notion is that, the given dataset is randomly divided into a training set, X_{tr} and a testing set, X_{te}. Then X_{tr} is further divided into labeled set, $X_{tr}(L)$ and unlabeled set, $X_{tr}(U)$. An initial classifier C is trained with $X_{tr}(L)$ according to some training algorithms such as KNN, ELM, NN etc. Then the initial classifier is used to get the fuzzy vector for each of the samples from $X_{tr}(U)$ and calculate their fuzziness. Based on the magnitude of fuzziness, samples are categorized into $X_{tr}(U)low$, $X_{tr}(U)midium$ and $X_{tr}(U)high$ fuzzy samples. Only the samples with low fuzziness, $X_{tr}(U)low$, are added to the original training set. Then the classifier is retrained with this new training samples. The detailed algorithm is given below (Algorithm 1).

Algorithm 1 FSSL-PAR

Input: Dataset, k (# of nearest neighbor for KNN or # of node in the hidden layer of ELM.)
Output: Maximum improvement of accuracy.

Randomly partition the dataset into a training dataset $\mathbf{X}_{tr}$ and a testing dataset $\mathbf{X}_{te}$.
Randomly partition the training set, $\mathbf{X}_{tr}$, into labeled dataset $\mathbf{X}_{tr}(L)$ and unlabeled dataset, $\mathbf{X}_{tr}(U)$.
$m \leftarrow 1$
$n \leftarrow 1$
while $m <= k$ **do**
  Train the classifier $\mathbf{C}$, according to a training algorithm, based on $\mathbf{X}_{tr}(L)$
  Get the training and testing accuracy respectively tr_{accB}, te_{accB}
  Get the fuzzy vector $A_i = \{\theta_1, \theta_2, \cdots, \theta_n\}$ for each sample in $\mathbf{X}_{tr}(U)$ by classifier $\mathbf{C}$.
  Calculate the fuzziness $P(A_i)$ of each sample in the $\mathbf{X}_{tr}(U)$ by $P(A_i) = -\frac{1}{n}\sum_{i=1}^{n} \theta_i\ log\theta_i + (1-\theta_i)\ log(1-\theta_i)$
  Sort the samples by the fuzziness $P(A_i)$, and group them into three fractions: $\mathbf{X}_{tr}(U)low$, $\mathbf{X}_{tr}(U)midium$ and $\mathbf{X}_{tr}(U)high$.
  Get a new training set $\mathbf{X}_{tr}new = \mathbf{X}_{tr}(L) \cup \mathbf{X}_{tr}(U)low$.
  Retrain a new classifier $\mathbf{C}_{new}$ according to the given training algorithm with $\mathbf{X}_{tr}new$.
  Again record the training and testing accuracy respectively tr_{accA}, te_{accA} by classifier $\mathbf{C}_{new}$ with $\mathbf{X}_{tr}new$.
  Record $diff[n] = te_{accA} - te_{accB}$
  $m = m + 1$
  $n = n + 1$
end while
Find the maximum of $diff$.

5.1. Non-iterative neural network with single hidden layer

Here, we will briefly describe the non-iterative neural network with exactly one hidden layer [52,53]. There are three essential parts in this algorithm namely, (i) input layer, (ii) hidden layer and (iii) output layer. Suppose, the training set, $\aleph = \{(x_i, t_i)|x_i \in \mathbb{R}^n, t_i \in \mathbb{R}^m, wherei = 1, 2, \ldots, N\}$, the activation function $g(x)$ and the # of hidden layer node is P.

Firstly, we randomly choose the weight w_i and the bias b_i where, $i = 1, 2, \ldots, P$, which are used between the input and the hidden layer. Secondly, based on w_i and b_i, hidden layer output matrix is calculated. Finally, the output weight, B is obtained by the following equation.

$$B = (H^T H)^{-1} H^T T \tag{5}$$

In the 1st experimental setup, we randomly assign the value of Q, n and m where Q is the # of hidden layer node, n and m are two integers such that n is smaller than m. We start the experiment by assigning a random value of n to Q and gradually change it and the final value is m. We change the value of hidden layer node to observe for which value of hidden layer we get maximum performance. The details of the 1st experimental setup is depicted in the Fig. 3:

5.2. Fuzzy KNN algorithm

KNN algorithm is one of the most important algorithms of machine learning. The underlying idea of KNN is very simple for implementation. However, James et al. [54] proposed the fuzzy version of the KNN algorithm for the first time which brought substantial improvements over the existing traditional crisp KNN. The major improvement of fuzzy KNN is that, it's generalization

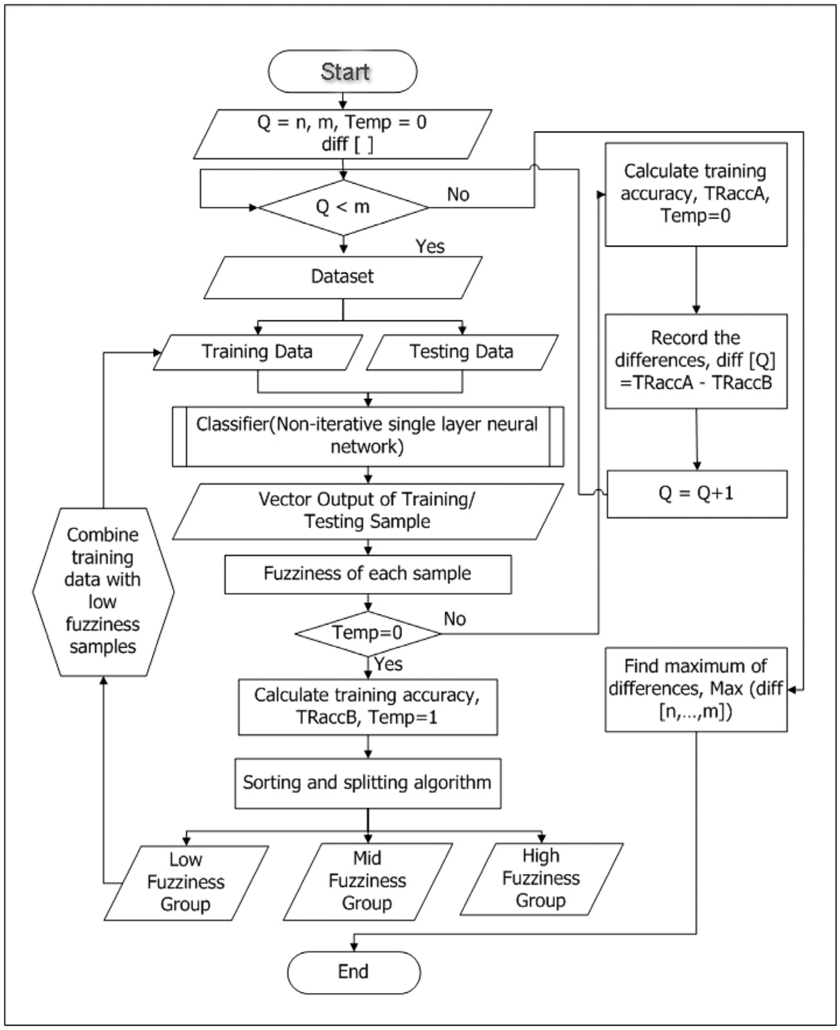

Fig. 3. Flowchart of the 1st experiment.

ability and it's ability to model the real world more perfectly. Fuzzy set theory had been successfully applied in the fuzzy KNN algorithm. As such, this new algorithm classify an object to a specific class with a degree of belief instead of classifying to a specific class with 100% belief. The degree of belief of fuzzy KNN lies between 0 and 1. There are numerous applications of fuzzy KNN to solve real-world problems. Considering the great success of fuzzy KNN, we use it in our proposed method for the patient's activity recognition problem.

In our 2*nd* experiment, the only parameter K has been chosen to get the maximum classification accuracy. In this regard, the value of K has been calibrated across a broad range. From this experiment, we can easily notice that, for which value of K maximum improvement is achieved.

6. Experimental results and discussion

6.1. Environment

The proposed algorithm was implemented in Python 3 software, and the experiments were conducted on a personal computer with Windows 10, 64-bit operating system, an Intel core i5-4590 CPU, 3.30 GHz processor and a 12 GB RAM. In the experiment, we consider the 10-fold cross-validation method to select the best experimental results for each method.

6.2. The database

The database was collected by Haque et al. [8] from the Hammel Neurorehabilitation and Research Center (HNRC), Denmark. The collection of the database from the real patient is difficult due to their health conditions. To this end, a group of volunteers both male and female, with sound health, acted to generate the database. To simulate the actual scenario, a patient's room was decorated at HNRC with all necessary stuffs, such as bed, Kinect camera, Thermal camera and IP camera. Volunteers were asked to lie on the bed so that his/her face remains upward. The lighting and the temperature were calibrated to represent the actual hospital condition. The cameras were set up on the ceiling of the room, so that the recordings are not interrupted by the hospital staffs. Although in the real hospital scenario, some staff members may appear during the monitoring time of the patient, but in our experimental set up we did not consider other staff members in the room. For the experimental simplicity we consider only patient. The overall scenario is depicted in the following figure:

Two possible conditions were considered while recording, such as patients can either covered with blanket or uncovered. Volunteers performed a series of movements. Half of the sequences were taken while the patients were covered by a blanket, other half were taken when patients were uncovered. These sequences were random. Then the experimenter asked the volunteers to lie on the bed and then the volunteer start to move

Table 1
Without the fusion of the features.

Features	Supervised learning algorithms						SSL(ELM)					SSL(KNN)				
	KNN	DT	NB	SVM	LDA	GLM	5%	10%	30%	50%	80%	5%	10%	30%	50%	80%
R-F1	35.82	42.47	**49.68**	49.29	30.14	45.96	45.19	46.24	46.08	47.16	**50.48**	46.54	48.50	48.89	50.16	**53.03**
R-F2	55.51	43.94	59.51	**60.00**	56.82	57.47	56.04	57.96	56.40	60.32	**60.45**	58.39	60.61	59.42	63.93	**65.77**
R-F3	68.25	61.45	12.50	**72.66**	66.74	71.57	69.42	69.40	71.43	72.41	**74.60**	70.87	72.69	75.11	76.32	**79.73**
R-F4	65.76	63.72	61.90	58.63	63.09	**68.02**	66.42	67.36	68.46	69.43	**70.44**	68.94	70.29	71.85	72.02	**74.60**
T-F1	36.21	32.65	**47.49**	44.83	33.38	45.56	45.40	45.64	46.46	**47.43**	**47.43**	44.75	44.92	48.99	49.90	**50.98**
T-F2	**56.57**	33.29	12.50	46.49	46.40	48.73	53.35	53.32	54.39	57.35	**58.43**	56.92	55.99	57.91	60.92	**61.75**
T-F3	58.52	63.34	12.50	63.18	41.37	**64.10**	62.41	62.55	64.38	64.41	**67.39**	65.13	65.28	68.00	67.83	**70.16**
T-F4	63.11	58.13	49.43	32.72	59.05	**64.95**	62.47	63.40	63.46	66.47	**67.46**	65.89	66.97	66.13	69.30	**70.13**
D-F1	18.22	24.51	**29.19**	28.54	21.68	26.70	40.46	41.43	40.44	43.44	**44.38**	41.78	43.59	42.77	45.09	**46.60**
D-F2	28.26	22.97	32.20	**37.83**	32.21	34.31	38.35	38.36	39.41	41.42	**42.41**	39.46	40.28	41.38	43.50	**45.53**
D-F3	38.38	42.32	12.50	46.19	**50.81**	46.67	45.35	46.36	48.37	53.38	**54.41**	43.62	44.68	50.38	56.05	**58.83**
D-F4	41.37	37.65	33.67	**47.12**	44.26	45.24	47.42	47.40	48.40	51.47	**51.49**	48.79	49.77	50.62	54.64	**56.76**

Note: Here, R = RGB, T = Thermal and D = Depth.
F1 = FDiff, F2 = PCA, F3 = PRadon and F4 = DiffRadon.
For example, R-F1 means RGB-FDiff.

Table 2
After the fusion of the features.

Features	Supervised learning algorithms						SSL(ELM)					SSL(KNN)				
	KNN	DT	NB	SVM	LDA	GLM	5%	10%	30%	50%	80%	5%	10%	30%	50%	80%
R-D-F1	30.73	**53.13**	49.3	50.02	34.28	47.72	52.75	53.92	54.99	55.90	**58.98**	53.29	57.85	56.86	61.57	**62.60**
R-D-F2	52.33	48.28	**63.01**	51.83	48.99	53.77	63.92	63.99	64.31	65.92	**66.38**	65.66	66.72	66.66	67.35	**68.37**
R-D-F3	59.2	59.34	12.5	**72.29**	70.3	70.26	74.33	74.28	77.68	76.28	**80.63**	77.36	77.58	78.27	80.93	**81.07**
R-D-F4	60.4	60.9	61.39	55.84	65.89	**70.20**	71.39	72.97	75.11	76.30	**78.31**	76.39	78.29	79.03	80.78	**82.49**
R-T-F1	39.7	44.22	**54.97**	49.29	40.32	50.1	54.95	55.90	55.99	56.90	**58.20**	56.69	57.40	58.63	61.74	**63.87**
R-T-F2	61.19	45.48	12.5	60.82	57.48	**62.62**	62.92	61.99	64.31	63.92	**65.25**	56.01	60.11	61.83	62.11	**64.75**
R-T-F3	69.38	60.39	12.5	74.93	71.24	**77.68**	75.61	78.68	68.00	78.93	**80.26**	70.58	75.93	78.21	80.46	**84.53**
R-T-F4	61.81	67.89	66.04	55.74	68.72	**72.41**	70.19	71.97	72.13	73.30	**75.13**	71.81	71.25	74.95	76.44	**78.68**
D-T-F1	27.81	37.33	42.81	**48.08**	31.17	42.62	48.75	49.92	49.99	49.90	**51.98**	51.10	53.62	55.56	**57.27**	56.85
D-T-F2	42.16	35.19	12.5	46.17	40.61	**47.03**	56.21	55.10	57.10	60.39	**61.18**	54.90	54.55	58.83	62.29	**64.14**
D-T-F3	50.5	54.17	12.5	59.4	51.95	**61.12**	65.43	65.28	67.00	67.83	**69.16**	63.11	64.98	68.22	69.50	**72.31**
D-T-F4	53.91	52.91	49.18	40.98	59.25	**63.96**	66.09	67.17	66.53	69.50	**70.73**	63.41	64.12	64.38	66.32	**68.47**

Note: Here, R = RGB, T = Thermal and D = Depth.
F1 = FDiff, F2 = PCA, F3 = PRadon and F4 = DiffRadon.
For example, R-D-F1 means RGB-Depth-FDiff.

their body according to the instructions of he experimenter. Seventy (70) possible movements were recorded. Each of five (5) body segments was moved in 7 different movement with covering and uncovering the patients. We used the frames which were collected from the difference between two consecutive movements.

Video cameras were set up on the wall shelf, Three video cameras were simultaneously used in this experiment. a Microsoft Kinect V2 for depth, an axis 214 PTZ RGB camera, and an axis Q1922 thermal camera. The database contained the videos annotated with eight PA categories. The database contains 1127 annotated videos from nine volunteers and each video contained one motion event.

6.3. Discussion

In this section, we would like to clarify the following questions give their answers:

(1) Is the FSSL strategy effective for dealing with the patient's activity recognition problem?

(2) Does the fusion of multi-modal features help improve the generalization ability of the model?

(3) Does the proportion of labeled samples in the data set affect the effectiveness of the FSSL strategy?

For question 1, we use ELM and KNN as the basic classifier and apply the above-mentioned FSSL strategy to compare the performance of the trained model with the existing models trained with the supervised training mechanism on the above data set. In this experiment, we set the proportion of labeled samples in the training set to 80%. The experimental results are shown in Table 1.

It can be observed from the Table 2, that the ELM and KNN models using the FSSL strategy have achieved higher prediction accuracy on all data sets than the models using the supervised learning strategy. This experimental phenomenon fully illustrates the effectiveness of the FSSL strategy. For question 2, we still use the above experimental configuration to compare the performance of each algorithm on the feature-fusion data sets. The experimental results are shown in Tables 2 and 3. Tables 2 and 3 show the performance comparison between models trained with the FSSL strategy and models trained with the supervised learning mechanism (with feature fusion)

Comparing Tables 1 and 2, it can be observed that using the feature-fusion data sets to train the models corresponding to the above algorithms can improve their prediction accuracy in most cases. This phenomenon shows that multi-modal feature fusion can provide more useful information for the model training, which helps the models to make more accurate decisions, thereby improving the generalization ability of the models. Moreover, one can observe that on most feature sets (10/16), ELM and KNN models using the FSSL strategy can achieve higher prediction accuracy than models using the supervised learning strategy, which once again verifies the effectiveness of the FSSL strategy. For question 3, we evaluated the performance of the F-ELM and F-KNN on the data sets with different proportions (5%,10%,30%,50%,80%,) of labeled samples. The experimental results are shown in Figs. 4–7.

It can be seen from Figs. 2–5 that the prediction performance of the classifier will increase monotonously with the increase of the proportion of labeled samples in the data set, whether or not the data features are fused. Moreover, this experimental phenomenon has nothing to do with the type of classifier. It can be

Table 3
After the fusion of the features (continued).

Features	Supervised learning algorithms						SSL(ELM)					SSL(KNN)				
	KNN	DT	NB	SVM	LDA	GLM	5%	10%	30%	50%	80%	5%	10%	30%	50%	80%
R-D-T-F1	32.62	46.68	**52.56**	48.24	38.43	47.49	52.75	52.92	54.99	55.90	**58.98**	54.22	57.77	54.30	57.91	**61.03**
R-D-T-F2	53.57	41.67	12.5	57.29	52.03	**57.75**	58.59	55.99	57.91	60.92	**61.75**	55.85	56.85	56.88	62.69	**64.07**
R-D-T-F3	61.31	66.6	12.5	**75.28**	72.62	74.58	65.13	65.28	68.00	67.83	**70.16**	60.45	62.62	63.31	66.73	**68.21**
R-D-T-F4	69.53	66.25	67.18	51.4	**73.10**	71.46	65.89	66.97	66.13	69.30	**70.13**	69.05	70.86	71.42	72.23	**75.54**

Note: Here, R = RGB, T = Thermal and D = Depth.
F1 = FDiff, F2 = PCA, F3 = PRadon and F4 = DiffRadon.
For example, R-D-F1 means RGB-Depth-FDiff.

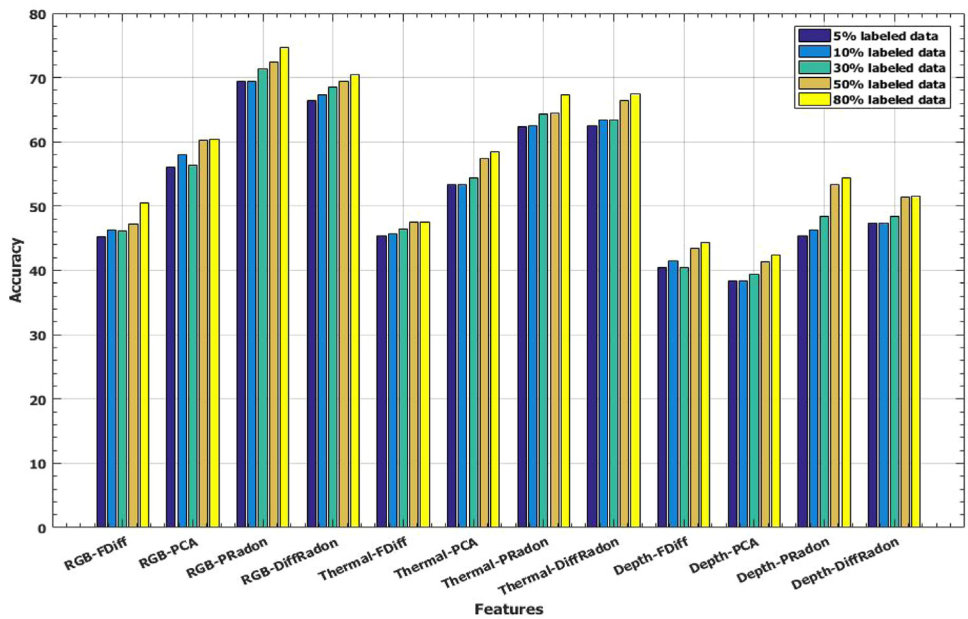

Fig. 4. Without the fusion of the features ELM.

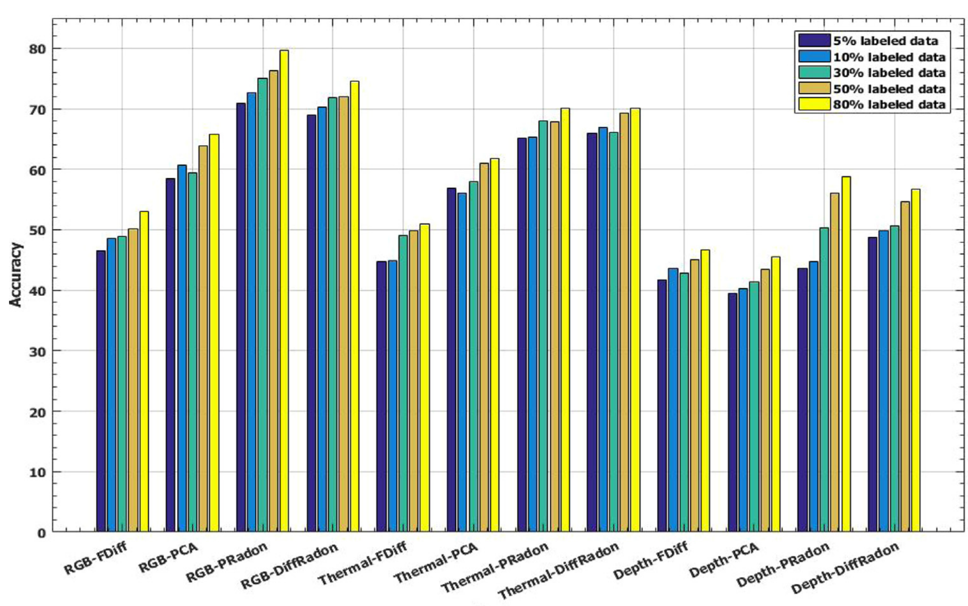

Fig. 5. Without the fusion of the features KNN.

concluded that to maximize the advantages of the FSSL strategy, it is also necessary to obtain as many labeled samples as possible. For the above experimental phenomenon, our explanation is as follows: given more labeled training samples, it means that the initial classifier is more likely to achieve a higher prediction accuracy, and then when we use the initial model to predict the output fuzziness of unlabeled samples, the corresponding results will be more accurate. In this way, the availability of unlabeled

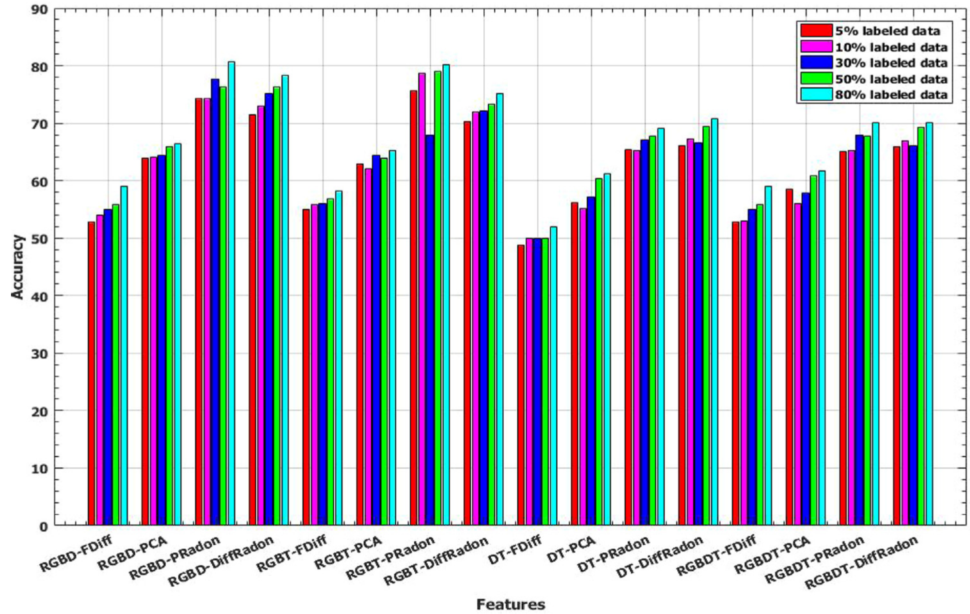

Fig. 6. After fusion of the features ELM.

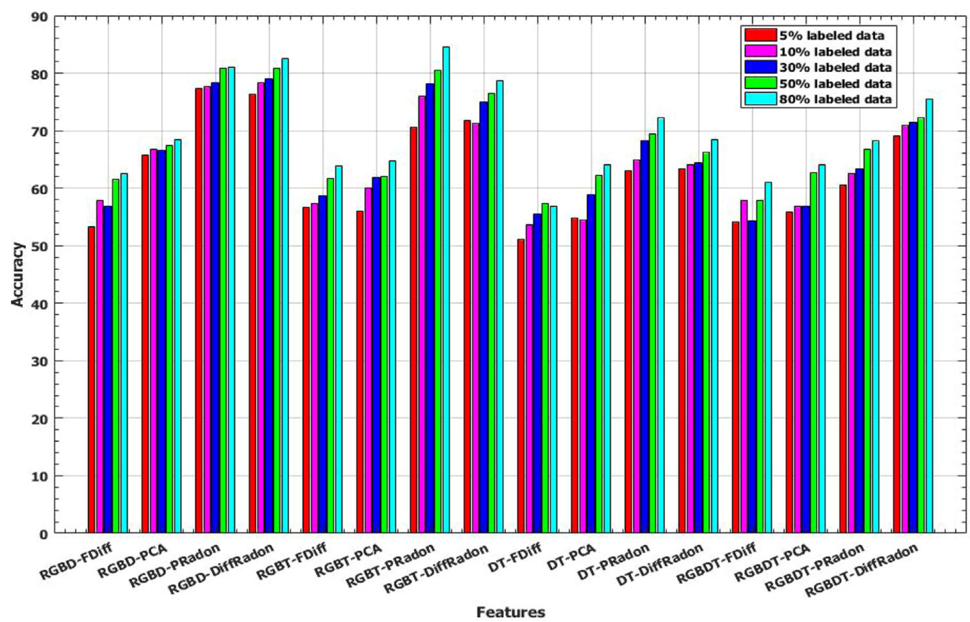

Fig. 7. After fusion of features KNN.

samples selected by the FSSL strategy is also higher, which is beneficial for the algorithm to get a better model after retraining.

Remark: From the above three experiments, we can confirm the following facts: (1) the FSSL strategy is still effective for dealing with the patient's activity recognition problem, which can effectively improve the generalization ability of the classifier; (2) the fusion of multi-modal features helps to improve the generalization ability of the model, and this phenomenon has nothing to do with the classification algorithm in most cases; (3) the ratio of labeled samples in the data set has a great influence on the effectiveness of the FSSL strategy. Based on our experimental results, we suggest that one should collect as many labeled samples as possible.

7. Conclusions and future works

In the modern technology-based health care system, automatic recognition of patients' activity is very important for monitoring and providing the proper services to the patients. Traditional supervised learning algorithms in the literature consider unimodal data and require a large amount of annotated data which is sometimes difficult to obtain in practice. In this paper, a multimodal RGBDT video database was used which was collected from a real hospital in Denmark. Besides, a novel fuzziness-based SSL algorithm was proposed for patient activity recognition. Experimental results proved the superiority of our proposed algorithm over some fully supervised learning techniques.

In the future, we will compare our proposed algorithm with the other state-of-the-art SSL algorithms. Besides, as the performance of the proposed SSL looks promising while using traditional handcrafted features, a further study may exhibit better performance if the proposed method is combined with deep learning based features. Future works may also include the performance analysis of the proposed SSL in the other application areas where a large amount of labeled data is difficult to obtain and subject-specific domain knowledge has a large impact on the traditional supervised learning settings. Incorporation of the proposed SSL model in a federated learning scenario where data is different in different test sites for distributed machine learning can also open a new direction of research in the future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Grant 61772344,Grant 62106150, and Grant 61732011), Natural Science Foundation of SZU (Grant 827-000140, Grant 827-000230, and Grant 2017060), Guangdong Province 2014GKXM054, CCF-NSFOCUS (Grant 2021001), and CAAC Key Laboratory of Civil Aviation Wide Surveillance and Safety Operation Management and Control Technology (Grant 2021001).

References

[1] A. Heinrich, D. Geng, D. Znamenskiy, J.P. Vink, G. de Haan, Robust and sensitive video motion detection for sleep analysis, IEEE J. Biomed. Health Inf. 18 (3) (2013) 790–798.
[2] M.A. Haque, R.B. Bautista, F. Noroozi, K. Kulkarni, C.B. Laursen, R. Irani, M. Bellantonio, S. Escalera, G. Anbarjafari, K. Nasrollahi, et al., Deep multimodal pain recognition: a database and comparison of spatio-temporal visual modalities, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2018, IEEE, 2018, pp. 250–257.
[3] M.A. Haque, R. Irani, K. Nasrollahi, T.B. Moeslund, Heartbeat rate measurement from facial video, IEEE Intell. Syst. 31 (3) (2016) 40–48.
[4] M.S.B. Alam, M.J. Patwary, M. Hassan, Birth mode prediction using bagging ensemble classifier: A case study of Bangladesh, in: 2021 International Conference on Information and Communication Technology for Sustainable Development, ICICT4SD, IEEE, 2021, pp. 95–99.
[5] B. Khaertdinov, E. Ghaleb, S. Asteriadis, Contrastive self-supervised learning for sensor-based human activity recognition, in: 2021 IEEE International Joint Conference on Biometrics, IJCB, IEEE, 2021, pp. 1–8.
[6] S. Sathyanarayana, R.K. Satzoda, S. Sathyanarayana, S. Thambipillai, Vision-based patient monitoring: a comprehensive review of algorithms and technologies, J. Ambient Intell. Humaniz. Comput. 9 (2) (2018) 225–251.
[7] J. Klonovs, M.A. Haque, V. Krueger, K. Nasrollahi, K. Andersen-Ranberg, T.B. Moeslund, E.G. Spaich, Distributed computing and monitoring technologies for older patients, Springer, 2016.
[8] M.A. Haque, S.S. Kjeldsen, F.G. Arguissain, I. Brunner, K. Nasrollahi, O.K. Andersen, J.F. Nielsen, T.B. Moeslund, A. Jørgensen, Patient's body motion study using multimodal RGBDT videos, in: International Symposium on Visual Computing, Springer, 2018, pp. 552–564.
[9] B.B. Kamdar, D.J. Kadden, S. Vangala, D.A. Elashoff, M.K. Ong, J.L. Martin, D.M. Needham, Feasibility of continuous actigraphy in patients in a medical intensive care unit, Am. J. Crit. Care 26 (4) (2017) 329–335.
[10] A. Heinrich, F. van Heesch, B. Puvvula, M. Rocque, Video based actigraphy and breathing monitoring from the bedside table of shared beds, J. Ambient Intell. Humaniz. Comput. 6 (1) (2015) 107–120.
[11] P. Bota, J. Silva, D. Folgado, H. Gamboa, A semi-automatic annotation approach for human activity recognition, Sensors 19 (3) (2019) 501.
[12] S. An, A. Medda, M.N. Sawka, C.J. Hutto, M.L. Millard-Stafford, S. Appling, K.L. Richardson, O.T. Inan, AdaptNet: Human activity recognition via bilateral domain adaptation using semi-supervised deep translation networks, IEEE Sens. J. 21 (18) (2021) 20398–20411.
[13] P. Gupta, P. Caleb-Solly, A framework for semi-supervised adaptive learning for activity recognition in healthcare applications, in: International Conference on Engineering Applications of Neural Networks, Springer, 2018, pp. 3–15.
[14] A. Saha, S. Das, Clustering of fuzzy data and simultaneous feature selection: A model selection approach, Fuzzy Sets and Systems 340 (2018) 1–37.
[15] M. Jamshed, S. Parvin, S. Akter, Significant HOG-histogram of oriented gradient feature selection for human detection, Int. J. Comput. Appl. 132 (17) (2015) 20–24.
[16] S. Hossain, M.Z. Hasan, M.J. Patwary, M.S. Uddin, An expert system to determine systemic lupus erythematosus under uncertainty, in: Proceedings of International Joint Conference on Advances in Computational Intelligence, Springer, 2021, pp. 117–130.
[17] M.J.A. Patwary, M.O. Rahman, M.S. Hossain, Uncertainty handling in ship assessment: a case study of Bangladesh, J. Invest. Manag. 4 (5) (2015) 152–161.
[18] M.J.A. Patwary, S. Akter, T. Mahmud, An expert system to detect uterine cancer under uncertainty, IOSR J. Comput. Eng. (IOSR-JCE), E-ISSN (2014) 36–47.
[19] X.-Z. Wang, R.A.R. Ashfaq, A.-M. Fu, Fuzziness based sample categorization for classifier performance improvement, J. Intell. Fuzzy Systems 29 (3) (2015) 1185–1196.
[20] M. Liu, M. Zhou, T. Zhang, N. Xiong, Semi-supervised learning quantization algorithm with deep features for motor imagery EEG recognition in smart healthcare application, Appl. Soft Comput. 89 (2020) 106071.
[21] X. Shi, Y. Jin, Q. Dou, P.-A. Heng, Semi-supervised learning with progressive unlabeled data excavation for label-efficient surgical workflow recognition, Med. Image Anal. 73 (2021) 102158.
[22] P. Singh, P. Agrawal, H. Karki, A. Shukla, N.K. Verma, L. Behera, Vision-based guidance and switching-based sliding mode controller for a mobile robot in the cyber physical framework, IEEE Trans. Ind. Inf. 15 (4) (2019) 1985–1997.
[23] K.I. Withanage, I. Lee, R. Brinkworth, S. Mackintosh, D. Thewlis, Fall recovery subactivity recognition with RGB-D cameras, IEEE Trans. Ind. Inf. 12 (6) (2016) 2312–2320.
[24] J. Xiao, R. Stolkin, Y. Gao, A. Leonardis, Robust fusion of color and depth data for RGB-D target tracking using adaptive range-invariant depth models and spatio-temporal consistency constraints, IEEE Trans. Cybern. 48 (8) (2018) 2485–2499.
[25] M. Hamer, M. Kivimaki, A. Steptoe, Longitudinal patterns in physical activity and sedentary behaviour from mid-life to early old age: a substudy of the whitehall II cohort, J. Epidemiol Community Health 66 (12) (2012) 1110–1115.
[26] R. Telama, Tracking of physical activity from childhood to adulthood: a review, Obes. Facts 2 (3) (2009) 187–195.
[27] J.B. Talkowski, E.J. Lenze, M.C. Munin, C. Harrison, J.S. Brach, Patient participation and physical activity during rehabilitation and future functional outcomes in patients after hip fracture, Arch. Phys. Med. Rehabil. 90 (4) (2009) 618–622.
[28] A. Sadeh, The role and validity of actigraphy in sleep medicine: an update, Sleep Med. Rev. 15 (4) (2011) 259–267.
[29] M.A. Haque, K. Nasrollahi, T.B. Moeslund, Heartbeat signal from facial video for biometric recognition, in: Scandinavian Conference on Image Analysis, Springer, 2015, pp. 165–174.
[30] H. Qian, S.J. Pan, C. Miao, Distribution-based semi-supervised learning for activity recognition (AAAI-19), 2019.
[31] I. Misra, A. Shrivastava, M. Hebert, Watch and learn: Semi-supervised learning for object detectors from video, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3593–3602.
[32] M.J. Patwary, J.N. Liu, H. Dai, Recent advances of statistics in computational intelligence (RASCI), Springer, 2018.
[33] S. Qiao, W. Shen, Z. Zhang, B. Wang, A. Yuille, Deep co-training for semi-supervised image recognition, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 135–152.
[34] Y. Cheng, X. Zhao, R. Cai, Z. Li, K. Huang, Y. Rui, et al., Semi-supervised multimodal deep learning for RGB-D object recognition, 2016.
[35] C. Gong, D. Tao, S.J. Maybank, W. Liu, G. Kang, J. Yang, Multi-modal curriculum learning for semi-supervised image classification, IEEE Trans. Image Process. 25 (7) (2016) 3249–3260.
[36] Z. Zhang, F. Ringeval, B. Dong, E. Coutinho, E. Marchi, B. Schüller, Enhanced semi-supervised learning for multimodal emotion recognition, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2016, pp. 5185–5189.
[37] Y. Tu, Y. Lin, J. Wang, J.-U. Kim, Semi-supervised learning with generative adversarial networks on digital signal modulation classification, Comput. Mater. Continua 55 (2) (2018) 243–254.

[38] A. Diaz-Pinto, A. Colomer, V. Naranjo, S. Morales, Y. Xu, A.F. Frangi, Retinal image synthesis and semi-supervised learning for glaucoma assessment, IEEE Trans. Med. Imaging (2019).

[39] M.P. Kumar, M.K. Rajagopal, Detecting facial emotions using normalized minimal feature vectors and semi-supervised twin support vector machines classifier, Appl. Intell. (2019) 1–25.

[40] M.J. Patwary, S. Akter, M.B. Alam, A.R. Karim, Bank deposit prediction using ensemble learning, Artif. Intell. Evol. (2021) 42–51.

[41] M.T. Uddin, M.J.A. Patwary, T. Ahsan, M.S. Alam, Predicting the popularity of online news from content metadata, in: 2016 International Conference on Innovations in Science, Engineering and Technology, ICISET, IEEE, 2016, pp. 1–5.

[42] Y. Hu, B. Wang, Y. Sun, J. An, Z. Wang, Graph-based semi-supervised learning for activity labeling in health smart home, IEEE Access 8 (2020) 193655–193664.

[43] Z.-J. Zha, T. Mei, J. Wang, Z. Wang, X.-S. Hua, Graph-based semi-supervised learning with multiple labels, J. Vis. Commun. Image Represent. 20 (2) (2009) 97–103.

[44] M.J. Patwary, X.-Z. Wang, Sensitivity analysis on initial classifier accuracy in fuzziness based semi-supervised learning, Inform. Sci. 490 (2019) 93–112.

[45] M.J. Patwary, X.-Z. Wang, D. Yan, Impact of fuzziness measures on the performance of semi-supervised learning, Int. J. Fuzzy Syst. 21 (5) (2019) 1430–1442.

[46] R.A.R. Ashfaq, X.-Z. Wang, J.Z. Huang, H. Abbas, Y.-L. He, Fuzziness based semi-supervised learning approach for intrusion detection system, Inform. Sci. 378 (2017) 484–497.

[47] S. Cabon, F. Porée, A. Simon, M. Ugolin, O. Rosec, G. Carrault, P. Pladys, Motion estimation and characterization in premature newborns using long duration video recordings, IRBM 38 (4) (2017) 207–213.

[48] J. Shi, et al., Good features to track, in: 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 1994, pp. 593–600.

[49] J.-Y. Bouguet, et al., Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm, Intel Corp. 5 (1–10) (2001) 4.

[50] S. Baker, I. Matthews, Lucas-kanade 20 years on: A unifying framework, Int. J. Comput. Vis. 56 (3) (2004) 221–255.

[51] M. Csikós, N.H. Mustafa, A. Kupavskii, Tight lower bounds on the VC-dimension of geometric set systems, J. Mach. Learn. Res. 20 (2019) 1–8.

[52] J. Liu, M.J. Patwary, X. Sun, K. Tao, An experimental study on symbolic extreme learning machine, Int. J. Mach. Learn. Cybern. 10 (4) (2019) 787–797.

[53] W. Cao, M.J. Patwary, P. Yang, X. Wang, Z. Ming, An initial study on the relationship between meta features of dataset and the initialization of NNRW, in: 2019 International Joint Conference on Neural Networks, IJCNN, IEEE, 2019, pp. 1–8.

[54] J.M. Keller, M.R. Gray, J.A. Givens, A fuzzy k-nearest neighbor algorithm, IEEE Trans. Syst. Man Cybern. (4) (1985) 580–585.