Mutual Information based Feature Selection for Nurse Care Activity Recognition

Md. Hasan Tarek*,[1], Md. Eusha Kadir*,[1], Mahir Mahbub[1], Pritom Saha Akash[1], Amin Ahsan Ali[2], Mohammad Shoyaib[1]
[1]Institute of Information Technology, University of Dhaka,
Dhaka, Bangladesh
[2]Department of CSE, Independent University,
Dhaka, Bangladesh
Email: {bsse0818, bsse0708, bsse0807, bsse0604}@iit.du.ac.bd, aminali@iub.edu.bd, shoyaib@du.ac.bd

***Abstract*—Human activity recognition is a challenging task as performing the activities varies from person to person. For the last few years, many complex methods have been proposed to identify human activities from sensor readings. To date, several studies have been conducted successfully to identify simple activities and many commercial applications have also been developed. However, reliable recognition of complex activities is still an active research area. The nurse care activity dataset can be treated as a complex activity recognition dataset. Researchers have proposed many solutions to identify nurse care activity by extracting numerous handcrafted features or using the Spatio-temporal graph convolution method. However, some of these features may be noisy, redundant, or even distract the classifier performance. In this paper, we propose a feature selection strategy to select important features from the handcrafted features. We claim that the simple classifier can provide satisfactory performance once the important features are selected and the noisy ones are eliminated. Experiments demonstrate that our proposed approach achieves 87.93% accuracy and 87.97% f1-score in test data. This is a significant improvement over state-of-the-art approaches on this benchmark dataset and thereby establishing our claim.**

***Index Terms*—Activity Recognition, Nurse Care Activity, Mutual Information, Feature Selection**

I. Introduction

In computer vision, smart environments, and healthcare research, human activity recognition is an important task. With the rapid growth of technology and the pervasiveness of electronics (e.g. smartphones, sensors and cameras), human activity has become an active area of research. Because of recent success in automated human activity recognition, it opens many opportunities in different areas such as human-computer interaction, health care, surveillance system, and security [1], [2]. Human activity recognition can also be useful for an individual to determine the changes in daily routine. Activity can be recognized using external, wearable sensors [3], and computer vision-based techniques [4]. Because of the complexity and diversity in human activities, it is challenging to recognize these activities. For example, doing the same activity by different persons can be different, the data collection rate of sensors can be varied, the sensors can be displaced [2].

As often being treated as a supervised classification problem, in activity recognition, researchers have applied different machine learning and deep learning approaches to extract features from sensor readings and recognize these activities [3], [5]. In the past decade, the advent of reliable sensors, and the availability of benchmark datasets have led this field to tremendous progress. Existing models perform well in recognizing simple activities (e.g. running, walking, sitting) [6], [7] with great precision. Many commercial products have been developed to recognize these simple activities.

In contrast to simple activities, the performance of applications attempting to recognize complex activities (e.g. cooking, nurse care) is still not up to the mark. The lack of publicly available data might be one reason behind that. In the health care domain, nursing activities can be considered as complex activities as the movements associated with these activities not only depend on the nurses but also depend on the behaviors of patients. This nurse care activity recognition has many healthcare applications (e.g. the facilitation of nurse training, checking compliance with care routines for a given patient, automatic document creation). A benchmark dataset of nurse activity recognition has been presented in [8]. Addressing this problem, several approaches ranging from deep learning to machine learning methods have already been proposed [9].

Haque et al. presented a gated recurrent unit-based approach to recognize these activities [10]. A spatio-temporal graph convolution method is used in [11]. Kadir et al. have proposed an approach by extracting features from raw data, dividing the features into some handcrafted groups, and using simple machine learning models [12]. They also argued that this simple classification approach is able to identify complex nurse activities with good accuracy.

Although the authors in [12] use different handcrafted features to achieve good accuracy, our observation is that all the features are not equally important for identifying an activity. Moreover, some features may be irrelevant, redundant, or create noise, which should be removed during classification. To achieve better performance, we propose a feature selection method namely $JMI\chi$. This method uses information-theoretic criteria and a Chi-Square (χ^2) test based filter approach to select relevant and non-redundant features.

*Both authors contributed equally to this work.

We organized the rest of the paper as follows. In section II, related works are discussed. The data collection process and other details of the nurse care activity dataset are presented in section III. An overview of feature extraction is briefed in section IV. Section V presents our feature selection method. We describe the overall nurse activity recognition process in section VI. The experiments and obtained results are discussed in section VII. In section VIII, we finally conclude the paper.

II. Related Work

Human activity recognition initially was limited to vision-based approaches. Early approaches to human activity recognition used template matching [13], [14]. One major limitation of such approaches is the requirement of high-level representations of the human body which is difficult to extract reliably. Addressing this limitation, the use of local features and optical flow features have gained popularity [15], [16]. In [17], a dynamic programming approach was used to recognize activity sequences from videoes. The earlier approaches recognized human activity from video sequences but it fails to capture the daily living of an individual with proper privacy. To capture the activities of daily living, the recognition of human activities from wearable sensors has become familiar.

The idea of using wearable sensors to recognize an activity was initially established by the work of house automation [18]. Wearable sensors can easily be embedded in user's clothes, belts, shoes, or directly in the body. It has been found that, based on the nature of activities, different types of sensor information show effectiveness in classifying human activities. For example, in monitoring activities involving repetitive body monitoring such as walking, running [19], accelerometer sensors are unusually effective.

In ubiquitous and pervasive computing, sensor data collected from the user's movement, location, or human-object interaction are usually noisy. After collecting sensor data, these raw data are converted into features by conducting a pre-process, and then a model is learned to classify different activities. Human activity recognition models can broadly be characterized into two groups namely the logic-based approach and probabilistic approach [20]. In logic-based approaches, activities are represented by a set of first-order statements. These approaches are noise-sensitive which is the major disadvantage of such approaches. To handle the noise in sensor data, probabilistic approaches are more suitable.

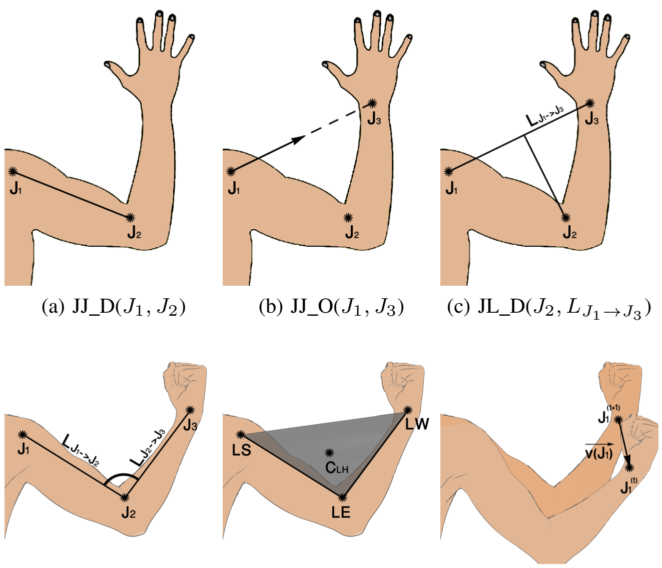
Probabilistic approaches classify human activities in a non-deterministic manner. Before applying probabilistic approaches, from the sensor data [19], various features are first extracted. Different machine learning models such as Naïve Bayes [6], [19], K-Nearest Neighbor (KNN) [21], [22], Decision tree [22], Support Vector Machine [23] are applied to recognize sequential activities in different settings. Moreover, hidden Markov models, dynamic belief networks, conditional random fields are also used to recognize human activities [24]–[26]. Deep learning approaches have also become prominent in this research area [27], [28]. A self-attention based neural network model has been published in [28]. In this model, recurrent architectures are used by applying different attention mechanisms for learning higher dimensional feature representation to classify activities.

A deep learning-based algorithm namely gated recurrent unit with attention mechanism to classify six different nurse care activities has been presented [10]. In this approach, the windowing method is used to extract features from location and motion capture data. However, this approach suffers from data overfitting and provides a lower classification accuracy in the test data. In [11], a spatial-temporal graph convolutional network has been presented for the same nurse care activities. They divide the time-series data into 20-second segments with a 10-second overlap and process 3D motion capture data using the spatial-temporal graph. Finally, the activity label is predicted based on the majority decision from each segment's output. To recognize the same nurse care activities, Kadir et al. have extracted some meaningful and reasonable features from location and motion capture data [12]. In this paper, they have grouped these features into four categories and proposed an ensemble of KNN to predict the activity label. This approach provides a reasonable classification accuracy but authors grouped extracted features into four groups intuitively which makes this approach less generalized. Moreover, these handcrafted features are not checked whether these are necessary for classification. This is necessary because there may exist some features that decrease the classification performance.

III. Dataset Overview

Nurse care activity recognition Dataset has been captured at the Smart Life Care Society Creation Unit in Kyutech, Japan [8]. A total of three days had been taken to collect the activity data performed by 8 professional nurses. The experimental room was organized with a desk, bed, and a wheeled cart. At the time of collecting data, nurses were provided with drip, gauze, diaper, and also other necessary things to perform the nurse activities. In the dataset, there are a total of 6 nurse activities which are C_1: Measurement of vital signs, C_2: Blood collection, C_3: Measurement of blood glucose, C_4: Indwelling drip retention and connection, C_5: Oral care, and C_6: Diaper exchange and cleaning of the area. With 5 repetitions by 8 nurses, in total 40 samples had been collected for each activity. Thus 240 activity sequences had been captured. Each activity sequences has been segmented into 60 seconds segment.

Data had been collected using three sensors namely Motion capture, location, and accelerometer sensor. The motion capture system used 29 body markers located in different parts of the body tracked by 16 infrared cameras to capture the data. Tri-axial data was collected with a rate of 100 samples per second. The location data was captured using a Bluetooth based Meditag sensor. A Bluetooth beacon had been placed on the right chest pocket of each nurse performing each activity. To receive the data from the beacon four receivers had been embedded in the laboratory which captured the locomotion data on both axes (x and y) at a rate of 20 Hz and also measure the air pressure. A smartphone is used to capture



(a) JJ_D(J_1, J_2) (b) JJ_O(J_1, J_3) (c) JL_D($J_2, L_{J_1 \to J_3}$)

(d) Line-Line Angle (e) Bary Center (f) Displacement Vector

Fig. 1: Extracted features from different joints

accelerometer sensor data. It had been placed in the right chest pocket of a nurse in the upright position to measure the tri-axial (horizontal, forward, backward movement) acceleration in m/s^2, and data collection rate was 4 Hz.

IV. Feature Extraction

Human activity is a collection of human motions. These motions can be identified by recognizing the movement of bones, joints, and muscles. Skeletal muscle is the active part of the human motion system. To capture human motions, geometric features (e.g. angles, position, orientation) can be extracted from different joints of the human body. The extraction of geometric features from different joints of the human body are discussed in literature [29]–[31]. Some of these geometric features are described below.

1) **Joint Coordinate** (J_C): $J\_C(J_1)$ represents the position of joint J_1 in a 3D coordinate system.
2) **Joint-Joint Distance** (JJ_D): The distance between joint J_1 and J_2 is represented by $JJ\_D(J_1, J_2)$.
3) **Joint-Joint Orientation** (JJ_O): $JJ\_O(J_1, J_2)$ denotes the unit vector directed from J_1 to J_2.
4) **Joint-Line Distance** (JL_D): $JL\_D(J_1, L_{J_2 \to J_3})$ represents the distance between joint J_1 to line $L_{J_2 \to J_3}$. Here, $L_{J_2 \to J_3}$ indicates the line between joint J_2 to J_3.
5) **Line-Line Angle** (LL_A): $LL\_A(L_{J_1 \to J_2}, L_{J_3 \to J_4})$ denotes the angle between line $L_{J_1 \to J_2}$ to $L_{J_3 \to J_4}$.
6) **Joint-Plane Distance** (JP_D): $JP\_D(J_1, P_{J_2 \to J_3 \to J_4})$ represents the euclidean distance between joint J_1 to plane $P_{J_2 \to J_3 \to J_4}$.

Besides these geometric features, action features based on human kinematics are also informative to represent human motions. A two-level hierarchical model was proposed to extract the action features [32], [33]. While extracting the first level features, the part of a human body is considered in an action (e.g. the action of the left hand as a unit). A barycenter of each unit is used to extract action features. The barycenter of a unit (e.g. left hand) can be measured using Eq. (1) where C_{LH} represents the barycenter of the left hand and LE, LW, LS present the 3D coordinates of left elbow, left wrist and left shoulder accordingly.

$$C_{LH} = \frac{LE + LW - 2 \times LS}{2} \tag{1}$$

From the barycenter of an unit (e.g. left hand), Range ($R(C_{LH})$), Mean ($M(C_{LH})$, Variance ($V(C_{LH})$ and Skewness ($S(C_{LH})$) are computed as first-level action features. To compute the second-level features, the local relative offset distance (displacement vector, $\overrightarrow{v(J)}$) on each axes between the current frame and the next frame is computed following Eq. (2) where $J\_C(J)^{(t)}$ represents the coordinates of joint J at frame t.

$$\overrightarrow{v(J)} = J\_C(J)^{(t+1)} - J\_C(J)^{(t)} \tag{2}$$

Action features and geometric features are needed to be extracted from sensor data to extract as these features carry more meaningful information about human motion.

V. Feature Selection

In the activity recognition, numerous features are extracted from sensor readings. According to the nature of activities, different types of features are important in different activity settings. For example, action features play an important role while classifying activities such as walking, running, swimming but these features may not be appropriate to classify other activities (e.g. sleeping, standing, sitting). Inappropriate features may create noise and mislead a classifier while identifying activities. In this regard, feature selection is important.

To select features, several filter methods based on information theory have been proposed [34], [35]. For selecting a

Algorithm 1: Feature Selection

Input : Set of features, F
Output: Selected subset of features, F_S
for $f_i \in F$ **do**
  Calculate MI for f_i with respect to C using Eq. (4)
end
$F \leftarrow$ Sort F in decreasing order based on their MI
$F_S \leftarrow f_1$
for $i = 2$ *to* $\mid F \mid$ **do**
  Calculate J_{JMI} for f_i using Eq. (3)
  if $J_{JMI} > \chi_c^2$ **then**
    $F_S \leftarrow F_S \cup f_i$
  end
end
return F_S

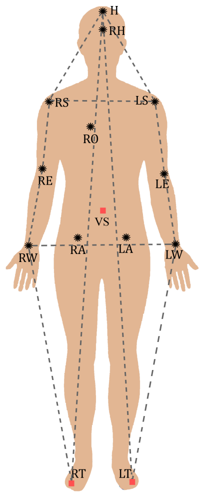


Fig. 2: Important body markers for feature extraction

feature different criteria such as correlation, mutual information (MI) have been proposed. Among them, MI-based feature selection methods are gaining popularity. In [34], a Joint Mutual Information (JMI) based feature selection approach has been presented. JMI computes the importance of a feature on the basis of the joint mutual information between the already-selected features and the class. The importance of a feature(f_i) can be calculated using Eq. (3).

$$J_{\text{JMI}} = I\left(f_{\text{i}}; C\right) - \frac{1}{|S|} \sum_{f_{\text{j}} \epsilon F_S} I\left(f_{\text{i}}; f_{\text{j}}\right) + \frac{1}{|S|} \sum_{f_{\text{j}} \epsilon F_S} I\left(f_{\text{i}}; f_{\text{j}} | C\right) \tag{3}$$

where C is the class label, F_S is the selected feature set, $I(X;Y)$ represents the mutual information between X and Y which can be formulated as follows.

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} P(x,y) \times \log \frac{P(x,y)}{P(x)P(y)} \tag{4}$$

Though JMI ranks all the features based on importance, it does not eliminate any features from the actual feature set. Note that, there are three terms representing relevancy (first term), redundancy (second term), and complimentary (third term) in Eq. (3). Each of these three terms follows χ^2 distribution [35]. Being motivated from that paper, we present an automatic threshold to decide either a feature will be selected or not. When $g = I(X;Y)$ follows χ^2 distribution, the critical value of $\chi^2_c(g)$ can be computed using Eq. (5).

$$\chi^2_c(g) = 2N \ln(2) \times I(X;Y) \tag{5}$$

where N is the total number of samples. We can compute the critical value for Eq. (3) in the similar fashion. Algorithm 1 describes our proposed feature selection approach namely $JMI\chi$. As can be seen in Line 8 of Algorithm 1, we compare the respective J_{JMI} with its corresponding critical value (χ^2_c) to take the decision regarding feature selection.

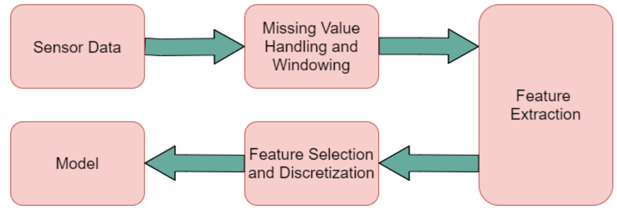


Fig. 3: Overview of overall activity recognition procedure

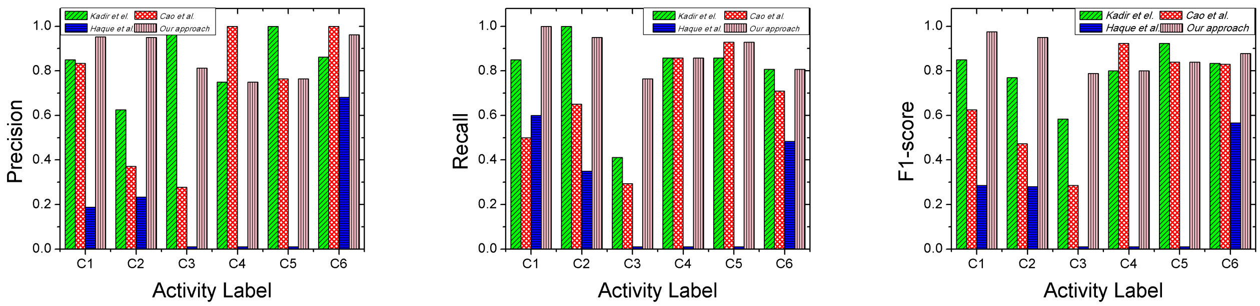
VI. Nurse Activity Recognition

In the nurse activity recognition dataset, three types of sensor data were provided. In this dataset, there are missing sensor readings. To overcome this situation, a simple imputation called simple linear interpolation has been performed. Sensor readings for a single frame help to build a human skeleton. A contiguous combination of such frames represents an activity. In the provided dataset, A 60-seconds segment represents an activity. We follow the windowing approach and divide each 60-seconds segments into 30 windows so that the duration of each window is 2 seconds.

After handling missing values and applying the windowing approach, we extract features from sensor data. To identify nurse care activities, we extract features from motion capture and location data. From location sensors, a 2-D position (x, y) of a nurse and air pressure data are available. For a window, we extract several features such as Mean position $(Mean(x), Mean(y))$, Standard Deviation $(SD(x), SD(y))$, Maximum $(M_{max}(x), M_{max}(y))$, Minimum $(M_{min}(x), M_{min}(y))$, Block ID where the nurse stays in that window, and Spatial and angular distance for x and y coordinates between first and last samples of a window. However, the extracted features from location sensors are not enough by itself to identify an activity. Thus we use motion capture sensor readings in this regard.

While extracting features from motion capture data, we have manually inspected that all these 29 body markers are not important. So, we use 14 important body markers represented as follows: J_1: Head (H), J_2: Rear Head (RH), J_3: Right Shoulder (RS), J_4: Right Elbow (RE), J_5: Right Wrist (RW), J_6: Left Shoulder (LS), J_7: Left Elbow (LE), J_8: Left Wrist (LW), J_9: Right Offset (RO), J_{10}: Right Asis (RA), J_{11}: Left Asis (LA), J_{12}: V Sacral (VS), J_{13}: Right Toe (RT), J_{14}: Left Toe (LT). From these 14 important body markers, we choose 11 joints, 12 lines, and 3 planes for feature extraction which are informative to identify nurse care activities. The chosen joints (marked with *) and lines are illustrated in Fig. 2.

From these chosen 11 joints, ${}^{11}C_2 = 55$ joint-joint distances are computed. Joint-joint orientations for the same combination are also extracted. From the chosen 12 lines, ${}^{12}C_2 = 66$ line-line angles are extracted. We also extract all combinations of joint-line distance from 11 joints and 12 lines which results in a total of 132 distances. We use three informative planes which are $P_{J_3 \to J_4 \to J_5}$, $P_{J_6 \to J_7 \to J_8}$, and $P_{J_2 \to J_9 \to J_{12}}$. Using these three planes, we extract 33 joint-plane distances and



(a) Activity-wise Precision

(b) Activity-wise Recall

(c) Activity-wise F1-score

Fig. 4: Activity-wise performance comparison among different approaches

consider each plane as a unit to extract action features. We also extract displacement vectors from all eleven joints.

To consider only important features and eliminate the unnecessary and noisy features, we use our proposed feature selection approach, $JMI\chi$. While applying $JMI\chi$, we perform equal-width discretization for all features. We simply apply KNN on the selected features which returns a probability vector $\left(Q^{(i)} = [P_{C_1}^{(i)} \ P_{C_2}^{(i)} \ \ldots \ P_{C_6}^{(i)}]\right)$ for the i^{th} window of an activity segment. As an activity segment is a contiguous sequence of some windows, we finally ensemble the results of all windows representing an activity segment and predict the activity label of this segment using Eq. (6) where $C = \{C_1, C_2, \ldots, C_6\}$ and m is the window size for an activity segment.

$$Y = arg \max_{C_j \in C} \sum_{i=1}^{m} P_{C_j}^{(i)} \tag{6}$$

VII. Experimental Results

In this section, we present and describe the experimental results of our proposed approach compared to other approaches. In the given dataset, train data consists of activities performed by six nurses and test data consists of activities performed by two other nurses. In our experiment, we fit our model with the training data and report the classification performance on the test data. As we apply KNN in our proposed method, we set $K = 10$ in this experimental setup.

Table I shows the overall accuracy and f1-score of different approaches. It is observed from Table I that, our proposed approach outperforms than other approaches in terms of overall accuracy and f1-score. Table I also demonstrates that the overall accuracy and f1-score of our proposed approach are 87.93% and 87.97% accordingly. Compared to the method described in [12], our proposed approach outperforms this even using the similar feature set. The reason behind this is the appropriate feature selection and inappropriate feature reduction. We even perform better with a single KNN than the ensemble of KNN using in [12].

We calculate the activity-wise precision and recall which are shown in Fig. 4a and 4b accordingly. It can be observed from Fig. 4a that, our proposed approach provides a reasonable precision score for all activities and performs better for activity C_1 and C_2 than other approaches in terms of precision. Our proposed approach also provides reasonable performance in terms of activity-wise recall which is shown in Fig. 4b.

TABLE I: Comparison among different approaches in terms of overall accuracy and f1-score

Method	Accuracy(%)	F1-score(%)
Kadir et al. [12]	80.17	79.53
Cao et al. [11]	64.66	66.56
Haque et al. [10]	29.31	24.88
Our approach	87.93	87.97

Apart from precision and recall, we also calculate f1-score for each activity. F1-score is the geometric average of precision and recall measures. It is also an important measurement metric for an imbalanced dataset. The activity-wise f1-score of different approaches is illustrated in Fig. 4c.

For a better understanding of the performance of our approach, we present the confusion matrix in Table II. Table II describes that our proposed approach classifies almost all activities effectively. Our proposed approach becomes confused while identifying C_3 and C_6. Some C_3 activity segments are identified as C_4 and some C_6 segments are predicted as C_5. The movements for activity groups (e.g. C_5 and C_6) are almost similar which might be a reason behind that. However, this performance is still considerable and satisfactory.

TABLE II: Confusion matrix of our proposed approach

		Predicted Label					
		C_1	C_2	C_3	C_4	C_5	C_6
Actual Label	C_1	20	0	0	0	0	0
	C_2	0	19	1	0	0	0
	C_3	0	0	13	4	0	0
	C_4	0	0	1	12	0	1
	C_5	0	1	0	0	13	0
	C_6	1	0	1	0	4	25

VIII. Conclusion

In this paper, we have proposed a mutual information-based feature selection algorithm. An activity recognition procedure using this feature selection algorithm is also presented in this paper. To recognize activities, we extract many necessary features from sensor data and select important features. We then apply a simple classifier (KNN) to recognize complex nurse activities. KNN algorithm followed by feature selection improves the nurse activity recognition result. We are planning to use this future selection strategy and extend our proposed approach for other complex activities in the future.

Acknowledgment

This research is supported by the University Grants Commission of Bangladesh.

References

[1] M. S. Ryoo, "Human activity prediction: Early recognition of ongoing activities from streaming videos," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 1036–1043.

[2] M. J. Roshtkhari and M. D. Levine, "Human activity recognition in videos using a single example," *Image and Vision Computing*, vol. 31, no. 11, pp. 864–876, 2013.

[3] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE communications surveys & tutorials*, vol. 15, no. 3, pp. 1192–1209, 2012.

[4] A. Bayat, M. Pomplun, and D. A. Tran, "A study on human activity recognition using accelerometer data from smartphones," *Procedia Computer Science*, vol. 34, pp. 450–457, 2014.

[5] M. Vrigkas, C. Nikou, and I. A. Kakadiaris, "A review of human activity recognition methods," *Frontiers in Robotics and AI*, vol. 2, p. 28, 2015.

[6] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *International conference on pervasive computing*. Springer, 2004, pp. 158–175.

[7] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, "Simple and complex activity recognition through smart phones," in *2012 Eighth International Conference on Intelligent Environments*. IEEE, 2012, pp. 214–221.

[8] S. Inoue, P. Lago, S. Takeda, A. Shamma, F. Faiz, N. Mairittha, and T. Mairittha, "Nurse care activity recognition challenge," 2019. [Online]. Available: http://dx.doi.org/10.21227/2cvj-bs21

[9] P. Lago, S. S. Alia, S. Takeda, T. Mairittha, N. Mairittha, F. Faiz, Y. Nishimura, K. Adachi, T. Okita, F. Charpillet *et al.*, "Nurse care activity recognition challenge: summary and results," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 746–751.

[10] M. N. Haque, M. Mahbub, M. H. Tarek, L. N. Lota, and A. A. Ali, "Nurse care activity recognition: A gru-based approach with attention mechanism," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 719–723.

[11] X. Cao, W. Kudo, C. Ito, M. Shuzo, and E. Maeda, "Activity recognition using st-gcn with 3d motion data," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 689–692.

[12] M. E. Kadir, P. S. Akash, S. Sharmin, A. A. Ali, and M. Shoyaib, "Can a simple approach identify complex nurse care activity?" in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 736–740.

[13] J. W. Davis and A. F. Bobick, "The representation and recognition of human movement using temporal templates," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1997, pp. 928–934.

[14] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 3, pp. 257–267, 2001.

[15] A. H. Shabani, D. A. Clausi, and J. S. Zelek, "Improved spatio-temporal salient feature detection for action recognition." in *BMVC*. Citeseer, 2011, pp. 1–12.

[16] Z. Lin, Z. Jiang, and L. S. Davis, "Recognizing actions by shape-motion prototype trees," in *2009 IEEE 12th international conference on computer vision*. IEEE, 2009, pp. 444–451.

[17] K. Kulkarni, G. Evangelidis, J. Cech, and R. Horaud, "Continuous action recognition based on sequence alignment," *International Journal of Computer Vision*, vol. 112, no. 1, pp. 90–114, 2015.

[18] M. C. Mozer, "The neural network house: An environment hat adapts to its inhabitants," in *Proc. AAAI Spring Symp. Intelligent Environments*, vol. 58, 1998.

[19] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *International conference on pervasive computing*. Springer, 2004, pp. 1–17.

[20] T. Gu, L. Wang, Z. Wu, X. Tao, and J. Lu, "A pattern mining approach to sensor-based human activity recognition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 9, pp. 1359–1372, 2010.

[21] B. Logan, J. Healey, M. Philipose, E. M. Tapia, and S. Intille, "A long-term evaluation of sensing modalities for activity recognition," in *International conference on Ubiquitous computing*. Springer, 2007, pp. 483–500.

[22] C. Lombriser, N. B. Bharatula, D. Roggen, and G. Tröster, "On-body activity recognition in a dynamic sensor network," in *2nd International Conference on Body Area Networks 2007*. Wearable Computing Lab, ETH Zurich, 2007.

[23] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *International workshop on ambient assisted living*. Springer, 2012, pp. 216–223.

[24] T. Huang, D. Koller, J. Malik, G. Ogasawara, B. S. Rao, S. J. Russell, and J. Weber, "Automatic symbolic traffic scene analysis using belief networks," in *AAAI*, vol. 94, 1994, pp. 966–972.

[25] D. L. Vail, M. M. Veloso, and J. D. Lafferty, "Conditional random fields for activity recognition," in *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, 2007, pp. 1–8.

[26] T. Van Kasteren, A. Noulas, G. Englebienne, and B. Kröse, "Accurate activity recognition in a home setting," in *Proceedings of the 10th international conference on Ubiquitous computing*, 2008, pp. 1–9.

[27] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert systems with applications*, vol. 59, pp. 235–244, 2016.

[28] S. Mahmud, M. T. H. Tonmoy, K. K. Bhaumik, A. M. Rahman, M. A. Amin, M. Shoyaib, M. A. H. Khan, and A. A. Ali, "Human activity recognition from wearable sensor data using self-attention," in *Proceedings of the Twenty-fourth European Conference on Artificial Intelligence*. IOS Press, 2020.

[29] L. Xia, C.-C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3d joints," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2012, pp. 20–27.

[30] R. Chaudhry, F. Ofli, G. Kurillo, R. Bajcsy, and R. Vidal, "Bio-inspired dynamic 3d discriminative skeletal features for human action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 471–478.

[31] S. Zhang, Y. Yang, J. Xiao, X. Liu, Y. Yang, D. Xie, and Y. Zhuang, "Fusing geometric features for skeleton-based action recognition using multilayer lstm networks," *IEEE Transactions on Multimedia*, vol. 20, no. 9, pp. 2330–2343, 2018.

[32] B. Su, H. Wu, and M. Sheng, "Human action recognition method based on hierarchical framework via kinect skeleton data," in *2017 International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 1. IEEE, 2017, pp. 83–90.

[33] B. Su, H. Wu, M. Sheng, and C. Shen, "Accurate hierarchical human actions recognition from kinect skeleton data," *IEEE Access*, vol. 7, pp. 52 532–52 541, 2019.

[34] H. Yang and J. Moody, "Feature selection based on joint mutual information," in *Proceedings of international ICSC symposium on advances in intelligent data analysis*. Citeseer, 1999, pp. 22–25.

[35] S. Sharmin, M. Shoyaib, A. A. Ali, M. A. H. Khan, and O. Chae, "Simultaneous feature selection and discretization based on mutual information," *Pattern Recognition*, vol. 91, pp. 162–174, 2019.