Human Activity Recognition through Probabilistic Data Fusion

Article

Accepted version

F. Batool, G. Lo Re, M. Morana, G. Rizzo.

Proceedings of the 2025 IEEE International Conference on Smart Computing (SMARTCOMP)

DOI: https://doi.org/10.1109/SMARTCOMP65954.2025.00100

Publisher: IEEE

Human Activity Recognition through Probabilistic Data Fusion

Farwa Batool*, Giuseppe Lo Re†, Marco Morana† and Giuseppe Rizzo†
*Scuola IMT Alti Studi Lucca, Lucca, Italy
farwa.batool@imtlucca.it
†Department of Engineering, University of Palermo, Palermo, Italy.
giuseppe.lore@unipa.it, marco.morana@unipa.it, giuseppe.rizzo24@unipa.it

***Abstract*—The increasing availability of smart devices in people's daily lives is constantly driving the design of novel services aimed to support the users by leveraging data provided by sensors embedded in their devices. In this paper, we present a scenario where data generated by wearable devices, such as smartphones and smartwatches, are analyzed to perform Human Activity Recognition (HAR). Given the different nature of the devices, using a single classifier may lead to inconsistent performance, especially for tasks that are semantically complex. Conversely, a distributed approach to activity recognition, where independent classifiers are used on each device, would be more computationally demanding and challenging to maintain. To address these issues, we present a probabilistic data fusion approach to integrate measurements from multiple devices while improving the overall system accuracy. Experiments performed on real data acquired from different devices show the effectiveness of our approach, especially in the recognition of complex activities.**

I. Introduction and Related Work

The compact, powerful and affordable nature of modern computing systems has seamlessly integrated a variety of distributed smart devices, such as smartphones and smartwatches, into users' daily life. Together with the rise of the Internet of Things (IoT)[1], personal devices have become indispensable, facilitating continuous communication, efficient activity management, proactive health monitoring, and intelligent home automation. This has led to Ubiquitous Sensing [2], where data collected anytime and anywhere is analyzed by Artificial intelligence (AI) and Machine Learning (ML) algorithms [3] to detect patterns, identify trends, and take data-driven decisions.

In this context, Human Activity Recognition (HAR) has emerged as an area of significant interest to infer user behavior from collected data. In particular, HAR methods based on raw sensory data, e.g., obtained from accelerometers, gyroscopes, and magnetometers embedded in smart devices, are of great importance for pervasive computing and smart environments managements.

HAR systems must be adaptable to different activity contexts, often employing machine learning techniques to extract meaningful features from raw data using overlapping sliding window approaches. Most techniques use sensors such as cameras [4], accelerometers and gyroscopes [5], thus enabling the collection of information, essential in identifying the position, orientation, movement and other insights regarding users. While much of the research in HAR has focused on collecting data from a single device [6], such an approach may fail to capture subtle movements of different body parts. Moreover, distinguishing between different actions that share similar sensory patterns remains challenging, especially as sensor measurements often carry inherent uncertainties due to factors such as instrument accuracy, environmental noise, and external interference.

To overcome these challenges, a shift towards distributed approaches is required. This involves using multiple sensing devices in a decentralized manner to increase data robustness, minimize bias and improve detection accuracy. By distributing detection across multiple devices, each with its own sensors and classifiers, the system gains multiple perspectives on the same phenomenon, thereby increasing reliability and mitigating sensor-specific inaccuracies. The authors of [7], for instance, proposed a multi-device HAR framework using the fog computing paradigm, while a distributed approach based on federated learning is presented in [8], enabling collaborative model training across devices while preserving privacy.

Distributed approaches also present challenges, such as increased computational requirements and the need for effective device coordination. In order to make the recognition performed by different distributed devices consistent, the output of multiple classifiers can be integrated through *Data Fusion* [9], [10] techniques, which also help in making the overall system more robust to the noise and uncertainties that are intrinsic to sensors. The primary goal of these approaches is to derive a comprehensive and accurate representation of the phenomenon under study. For example, Ehatisham-Ul-Haq et al. [11] used data fusion to develop a HAR system that merged features from accelerometers, gyroscopes, depth sensors, and RGB cameras, and then integrated KNN and SVM classifiers, achieving higher accuracy - albeit with increased processing time. Other applications highlight the pivotal role of data fusion in improving the accuracy, reliability and versatility of HAR systems in a variety of domains, such as passenger activity detection [12] or heart disease prevention using wearable sensors [13].

In this paper, we propose a Bayesian-based data fusion system that allows to integrate sensory information from multiple smart devices, e.g., smartwatches and smartphones, to classify human activities. The primary goal is to mitigate the inherent uncertainty associated with individual sensor readings by using the complementary data provided by every device, thereby improving the overall classification accuracy.

The remainder of this paper is organized as follows: Section II presents a detailed explanation of the proposed methodology. Section III discusses the experimental setup and results, and finally conclusions are drawn in Section IV.

II. System overview

The primary motivation underlying the proposed method lies in the difficulty of sensor-based HAR systems to successfully recognize different types of activities. In fact, although they are generally capable of distinguishing simple activities, e.g., *walking*, *running*, *standing*, or *remaining stationary*, the recognition algorithms may fail when the semantic of the activities - and thus the related patterns - are more complex, such as *working at the PC* or *driving*. Hence, the idea is to exploit together data acquired from multiple devices to compensate for the unreliability of individual sensors. For example, when working at PC, a smartphone (kept in a pocket or placed on the desk) might lead the system to infer that the user is stationary, while the smartwatch might suggest that the user's arms are actually moving.

Each device considered in our system performs three main phases, namely, *data acquisition*, *pre-processing* and *classification*. After that, the outputs of the individual (local) classifiers are further analyzed by a master device (indifferently one of those involved in the previous phases, for example the most powerful) for the final data fusion and classification stage. An overview of the proposed architecture is provided in Fig. 1. For the sake of clarity, only two devices are considered in the example scenario presented here, namely the smartphone (SP) and the smartwatch (SW). However, the architecture proposed can easily be adopted for a larger number of devices.

Following the acquisition of raw data from gyroscopes, accelerometers, and magnetometers embedded on the devices, as first pre-processing step, *data cleaning* is performed [14] to remove noise and sensor-dependent artifacts. In addition, ensuring that acquired data are aligned with respect to the same reference system, regardless of their position at the time of acquisition, is essential to ensure the proper functioning of the subsequent recognition phases. This is achieved by aligning each device's reference system with the terrestrial frame: the X-axis points east, the Y-axis is tangential to the ground and points towards the magnetic North Pole, and the Z-axis is perpendicular to the ground and points skyward, preventing errors caused by axis inversion. The transformation relies on azimuth, pitch and roll [15] to define the three-dimensional orientation. The final two stages of the pre-processing consist of signal *phasing* and *feature extraction*. The former ensures synchronization of data sequences across devices, e.g., SP and SW, minimizing time lags and maintaining temporal alignment for more accurate analysis. In order to achieve the desired time synchronization efficiently, we employed a sliding window approach [16] with overlapping windows to capture temporal relationships; larger windows with high overlap were found to capture activity patterns more effectively as will be described in Section III. Then, feature extraction is performed by calculating key statistical metrics for each sensor, including mean, standard deviation, maximum, minimum and root mean square (RMS). For both the accelerometer and the gyroscope, a total of 15 features were obtained by calculating these five metrics across the three sensor axes. Preliminary analysis of the raw magnetometer data, instead, showed that using the magnitude of the vector as a unified metric is more effective than evaluating its individual X, Y and Z components. Therefore, we evaluate the L_2 norm of the vector $M = (X, Y, Z)$, which provides a low variance measurement for this sensor.

In order to allow the system to decouple the feature vectors' recognition process from the semantics of the analyzed activities, before training the classifier we leverage K-Means clustering as a form of representation learning. The resulting clusters serve as a *dictionary*, implementing a Bag of Words approach for activities. By capturing recurring patterns as semantic units, this vocabulary-based representation organizes the feature space, enhancing both classifier performance and interpretability [17]. We then use an ensemble model consisting of n atomic classifiers per device, implemented as neural networks and trained according to a one-to-many approach, with n being the number of activities observed. Each classifier learns to establish a mapping between real activities and the corresponding labels resulting from the clustering process, identifying the clusters related to a specific activity while ignoring unrelated ones. In this phase, the training data consists of the original feature set, while the labels are inferred from the clustering itself, thus allowing the model to refine its representation of the activity patterns and improve the classification performance.

Starting from the probabilities provided by the local classifiers during inference, a new dataset composed of triples is built, where the first element represents the original activity, while the second and third elements encapsulate the evidences produced by the classifiers for the smartphone (SP) and smartwatch (SW), respectively. On this basis, the proposed data fusion technique exploits a Bayesian network [18] consisting of two parent nodes, *PriorSP* and *PriorSW*, representing the probabilistic evidence from the two devices respectively, and a child node, *Activity*, which captured the likelihood of a given activity occurring based on the combined evidence from both sources. The Maximum Likelihood Estimator (MLE) [19], is used to estimate the network parameters, thus, allowing the model to uncover correlations between sensor readings and improve classification accuracy. By defining probability thresholds, the system can dynamically adjust its classification confidence to meet specific application requirements. Furthermore, the inclusion of the Bayesian Data Fusion Model adds negligible computational overhead, as the underlying structure remains compact, and the parameter estimation scales linearly with the number of observations, making it an efficient solution for integrating multiple data sources.

To assess the robustness of the system, we defined a metric that quantifies the proportion of discarded predictions relative

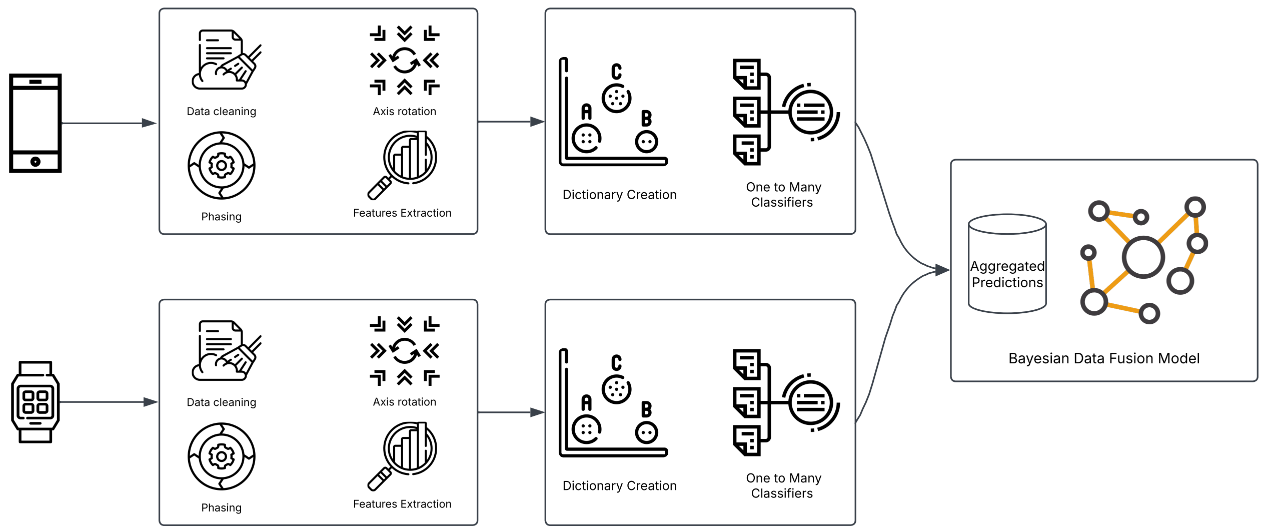

Fig. 1: Data captured from the devices are pre-processed separately before being passed to the classification module, which is also responsible for creating the dictionary. Finally, the results of the local classifiers are aggregated and evaluated by the data fusion module.

to the total number of predictions after applying the threshold:

$$N_D = \frac{N_T^*}{N_T}, \tag{1}$$

where θ indicates the predefined threshold, N_T the total number of predictions, and N_T^* the number of predictions N_T whose scores fall below θ. This measure provides valuable insight into the trade-off between classification confidence and prediction retention, offering a systematic approach to balancing accuracy and adaptability in real-world scenarios.

III. Experimental Analysis and Discussion

In order to carry out an in-depth experimental analysis of the performance of all architectural components, from the pre-processing modules to the individual classifiers up to the data fusion algorithm, a dataset was collected. In fact, to the best of our knowledge, datasets containing all the required information, i.e., simple and complex activities acquired simultaneously from multiple devices, are not available in the literature. To this aim, 13 people were asked to perform 5 simple and 2 complex activities while wearing a smartwatch and carrying their smartphones. As a result, 260.000 samples were obtained. Unintentionally, a certain imbalance was introduced during data collection, which became more pronounced after data cleaning process, in particular, the *Walking* class contained significantly more samples ($\approx 125,000$) compared to other activities. In order to achieve balanced learning, we down-sampled the dataset to match the number of instances in the minority class (*Stopped*, $\approx 11,000$, short for *Remaining Stationary*).

In order to time-align the extracted feature vectors, we adopted a sliding window approach, whose parameters were tuned using stratified 10-fold cross-validation [20], allowing for a robust and accurate assessment of classifier performance

Window size (sec)	**Overlap** (sec)	**Smartphone**		**Smartwatch**	
		Accuracy	**M-F1**	**Accuracy**	**M-F1**
10	*9*	*.88*	*.88*	*.86*	*.86*
9	8	.85	.85	.86	.85
8	7	.83	.83	.85	.85
7	6	.81	.82	.84	.84
6	5	.81	.81	.83	.82
10	8	.81	.81	.83	.83
9	7	.80	.80	.82	.82
8	6	.79	.79	.81	.81
5	4	.79	.79	.81	.81
10	7	.78	.78	.80	.80

TABLE I: The top 10 configurations of sliding windows (best values first), evaluated in terms of Accuracy and Macro-F1 values for each of the two devices.

	Smartphone		**Smartwatch**	
Activity	*TPR*	*F1*	*TPR*	*F1*
Walking	.81	.80	.74	.75
Running	.90	.90	.94	.96
Stopped	1	1	1	1
Go-Upstairs	.83	.83	.79	.78
Go-Downstairs	.85	.86	.85	.83

TABLE II: True Positive Rate (TPR) and F1 Score for each activity, evaluated using the window configuration (10,9).

as the window size and stride varied. The optimal configuration was found to be a 10-second window with a 9-second overlap (see Table I), whose corresponding activity recognition performance are reported in Table II in terms of TPR and F1-score values.

Another key element in the evaluation process is the tuning of the *dictionary*, derived from a K-means clustering [21] process. While K-means was chosen for its simplicity and effectiveness, determining the optimal number of clusters was a critical step in ensuring the best possible results. To this aim, we systematically evaluated the performance of the classifier

	Walking	Running	Stopped	Go-Upstairs	Go-Downstairs
Walking	.81	.05	-	.09	.05
Running	.04	.90	-	.03	.03
Stopped	-	-	1	-	-
Go-Upstairs	.1	.02	-	.83	.05
Go-Downstairs	.07	.02	-	.06	.85

(a) Smartphone (**SP**) - simple

	Walking	Running	Stopped	Go-Upstairs	Go-Downstairs
Walking	.74	.01	-	.14	.11
Running	.03	.94	-	.01	.02
Stopped	-	-	1	-	-
Going Upstairs	.13	.01	-	.79	.07
Going Downstairs	.06	-	-	.09	.85

(b) Smartwatch (**SW**) - simple

	Walking	Running	Stopped	Go-Upstairs	Go-Downstairs	Driving	At-PC
Walking	.85	.05	-	.03	.06	.01	-
Running	.04	.93	-	.01	.02	-	-
Stopped	-	-	1	-	-	-	-
Go-Upstairs	.14	.05	-	.71	.07	.03	-
Go-Downstairs	.06	.04	-	.03	.85	.02	-
Driving	.15	.34	-	.15	.3	.06	-
At-PC	-	-	1	-	-	-	-

(c) Smartphone (**SP**) - complex

	Walking	Running	Stopped	Go-Upstairs	Go-Downstairs	Driving	At-PC
Walking	.74	.01	-	.1	.12	.02	.01
Running	.03	.94	-	.01	.02	-	-
Stopped	-	-	1	-	-	-	-
Go-Upstairs	.15	-	-	.7	.11	-	.04
Go-Downstairs	.04	-	-	.07	.89	-	-
Driving	-	.16	-	.1	.68	.06	-
At-PC	-	-	-	-	.82	-	.18

(d) Smartwatch (**SW**) - complex

Fig. 2: Smartphone (SP) and smartwatch (SW) recognition accuracy for (a)(b) simple and (c)(d) complex activities, respectively.

Hyperparameters	
Number of layers	4
Number of neurons per layer	128, 128, 128, n_clusters + 1
Activation function	ReLU (middle layer), Softmax (output layer)
Optimizer	Adam
Learning rate	0.001
Loss function	Sparse Categorical Cross-Entropy
Training epochs	15
Batch size	32

TABLE III: Hyperparameters of the classifiers adopted in smartphone and smartwatch devices.

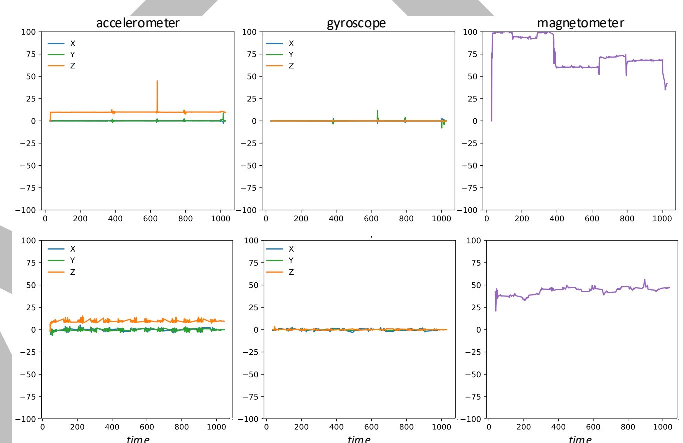

Fig. 3: Examples of data from *Stopped* (first row) and *Running* (second row) activities as captured by a smartphone device.

using different values of k_{SP} and k_{SW}, which represent the chosen number of clusters in the smartphone and in the smartwatch, respectively. For the sake of brevity, the results concerning the selection of k_{SP} and k_{SW} will be presented later in this Section, in conjunction with the evaluation of the data fusion method. In that context, these parameters will be analyzed in relation to the Bayesian classifier's acceptance threshold θ.

After the dictionary was created, the local classifiers were trained according to the hyperparameters reported in Table III. In particular, the classifier employed in each device is an ensemble of neural networks, each consisting of four layers: 128 neurons in the first three layers and $n_{clusters}+1$ neurons in the output layer, where $n_{clusters}$ is the number of clusters associated with a specific activity in the dictionary. ReLU activation is used for the hidden layers, while Softmax is used for the output. It is trained using the Adam optimizer with a learning rate of 0.001, for 15 epochs and a batch size of 32, using Sparse Categorical Cross-Entropy as the loss function.

As a preliminary step, we assessed the performance of the networks from both devices independently, without incorporating the data fusion module. The results achieved for the recognition of simple activities are shown in the first row of Fig. 2. As can be observed in Fig. 3, the *Stopped* activity is the easiest to recognize because of its distinct stationary pattern; similarly, due to a completely opposite pattern, high accuracy is also obtained for *Running* (**SP**: 90%, **SW**: 94%). On the other hand, moderately dynamic activities, such as *Walking*, *Go-Upstairs* and *Go-Downstairs* are more difficult to distinguish, regardless of the device considered, with average recognition rates around 83%.

The introduction of more complex activities further affects

k_{SP}	k_{SW}	$\theta = 0.6$			$\theta = 0.7$			$\theta = 0.8$			$\theta = 0.9$			$\theta = 1$		
		Acc	**M-F1**	N_D	**Acc**	**M-F1**	N_D	**Acc**	**M-F1**	N_D	**Acc**	**M-F1**	N_D	**Acc**	**M-F1**	N_D
10	10	.72	.72	.58	.74	.74	.59	.75	.75	.61	.80	.80	.63	1.00	1.00	.66
10	30	.82	.82	.42	.93	.93	.56	.95	.95	.57	.99	.99	.61	1.00	1.00	.62
10	50	.83	.83	.36	.87	.87	.47	.95	.95	.57	1.00	1.00	.60	1.00	1.00	.60
30	10	.80	.80	.49	.92	.92	.58	.94	.94	.60	.99	.99	.62	1.00	1.00	.64
30	30	.83	.83	.37	.90	.90	.51	.95	.95	.56	.98	.98	.59	1.00	1.00	.60
30	50	.85	.85	.30	.91	.91	.43	.96	.96	.52	.99	.99	.55	1.00	1.00	.57
50	10	.82	.82	.47	.92	.92	.57	.96	.96	.59	.99	.99	.61	1.00	1.00	.62
50	30	.86	.86	.34	.91	.91	.45	.96	.96	.52	.99	.99	.55	1.00	1.00	.55
50	50	.87	.87	.24	**.92**	**.92**	**.36**	.96	.96	.45	1.00	1.00	.50	1.00	1.00	.51

TABLE IV: A grid search was conducted over the hyperparameters k_{SP}, k_{SW}, and θ, evaluating performance in terms of Accuracy (Acc), Macro-F1 (M-F1), and Error Rate (N_D). The selected configuration is indicated in bold.

	Walking	Running	Stopped	Go-Upstairs	Go-Downstairs
Walking	.89	.01	0	.06	.04
Running	.01	.99	0	0	0
Stopped	0	0	1	0	0
Go-Upstairs	.09	0	0	.84	.07
Go-Downstairs	.03	.01	0	.06	.90

(a)

	Walking	Running	Stopped	Go-Upstairs	Go-Downstairs	Driving	At-PC
Walking	.92	0	0	.06	.02	0	0
Running	0	.99	0	0	.01	0	0
Stopped	0	0	1	0	0	0	0
Go-Upstairs	.1	0	0	.83	.05	.02	0
Go-Downstairs	.06	0	0	.08	.83	.03	0
Driving	.01	0	0	0	0	.99	0
At-PC	0	0	.01	0	0	0	.99

(b)

Fig. 4: Accuracy achieved by the data fusion system for (a) simple and (b) complex activities.

the classification accuracy, as shown in the second row of Fig. 2. In this case, the **SP** classifier frequently misclassified *At-PC* as *Stopped*, and often confused *Driving* with activities involving leg movement such as *Running*, *Walking* and *Go-Upstairs*. A significant number of false positives highlight the classifier's difficulty in distinguishing between similar motion patterns. Similarly, the **SW** classifier misidentified *At-PC* as *Go-Downstairs*, likely influenced by hand movements on the keyboard and mouse, and misidentified *Driving* as other dynamic, hand-intensive activities. These results confirmed that single classifiers are ineffective in contexts where tasks present different motion patterns on multiple devices.

The evaluation of the data fusion model began by closely examining the behavior of the Bayesian network at different thresholds $\theta_i \in [0.6, 1.0]$, given its role as a statistical thresholding mechanism within the system. For each value of θ, we further explored different combinations of k_{SP} and k_{SW}, in order to assess the overall performance of the system in terms of accuracy, Macro-F1, and the previously defined error rate N_D (Eq. 1).

The results (see Table IV) of the 10-fold stratified cross-valued grid search performed on k_{SP}, k_{SW}, and θ revealed that the most effective configuration - achieving the lowest error rate while maintaining strong overall performance - was obtained with a threshold of $\theta = 0.7$ and cluster parameters of $k_{SP} = k_{SW} = 50$, resulting in an average accuracy of about 92.2% and an error rate N_D of 36%. The chosen values of k, despite the potential risk of overfitting, highlight the effectiveness of using a larger number of clusters to successfully capture the complex representation of activity patterns. It is worth noticing that the chosen configuration is preferable to alternatives with slightly higher accuracy but significantly higher error rates. A lower N_D indicates that a greater proportion of predictions are generated with high confidence (i.e., above the θ threshold), thereby increasing both the reliability and interpretability of the model's output. In contrast, configurations with higher N_D typically reflect models that rarely make confident predictions, thereby limiting their practical effectiveness. Given this trade-off, our chosen configuration represents a balanced and robust choice.

Then, the performance of the complete HAR system - including the data fusion module - was evaluated on the two sets of activities discussed so far. The results shown in Fig. 4. The advantage of the Bayesian data-fusion module becomes particularly apparent when compared to the unimodal baselines that rely solely on data from either the smartphone (SP) or the smartwatch (SW); please refer to Fig. 2 for those results. For simple activities, the proposed approach consistently achieves an higher classification accuracy, outperforming both SP and SW individually. *Running* reaches 99% accuracy with the fusion model compared to 90%-94% (SP and SW only, respectively), while *Go-Upstairs* improves from an average 85% (on both SP and SW) to 90% with the fusion module. For the remaining simple activities, a slight average increase of 8% is observed, with the exception of *Stopped*, which was already correctly classified. These results highlight the model's ability to effectively integrate complementary signals from both devices, thereby improving reliability even in scenarios where individual systems already perform relatively well. The improvements are even more remarkable when semantically complex activities are considered (Fig. 4 (b) vs. Fig. 2 (c)–(d)). In unimodal settings, activities such as Driving and At-PC are frequently misclassified, with the smartphone-only system achieving just 6% and 0% accuracy, respectively, and the smartwatch-only system performing only marginally better (6% and 18%). Conversely, the adoption of the data-fusion module leads to a dramatic increase in recognition accuracy, reaching 99% for both classes. Surprisingly, although the overall performance across activities has been improved, we acknowledge a slight decrease in accuracy for *Go-Downstairs* (-2% on average); however, this comes with a consistent reduction in false positives for the *Go-Upstairs* class, whose accuracy increases from an average of 70% on both devices, to a solid 83%. Furthermore, activities such as *Walking* and *Running* show a notable improvement in accuracy (approximately 8.5% on average). This indicates the system ability in disambiguating complex, context-dependent activities that pose significant challenges for simpler models; integrating data from both SP and SW effectively overcomes the inherent limitations of each device, providing an enhanced and refined understanding of the activities being performed.

IV. Conclusion

In this work, we presented a Human Activity Recognition (HAR) system that adopts a data fusion approach to integrate sensory information from multiple smart devices, e.g., smartphones and smartwatches. The system is supported by a dedicated preprocessing module that enhances the quality of sensor data, thereby increasing the processing accuracy and classification reliability. HAR is performed through a Bayesian network trained on the predictions of individual classifiers so as to make the recognition more robust. Moreover, activities are categorized into semantically simple and complex groups, which helps the system to extract more precise extraction and analysis of activity-specific features. The fusion of data from both devices effectively mitigates the uncertainties inherent in individual sensor measurements, resulting in superior performance compared to single-device models. In particular, as demonstrated by the extensive experimental analysis, the system overcomes the limitations of isolated atomic classifiers by achieving high accuracy in recognizing both simple and complex activities, even when they share similar motion characteristics. Future work will explore the scalability of the system through extensive testing with additional sensor devices and a wider range of activities. Potential enhancements include integrating new sensing modalities, evaluating performance in different environments, and incorporating advanced machine learning techniques to further refine activity recognition under varying users' conditions.

Acknowledgment

This work was partially supported by the "S3 Campus" Project (POC Sicilia 2014/2020).

References

[1] S. Li, L. D. Xu, and S. Zhao, "The internet of things: a survey," *Information systems frontiers*, vol. 17, pp. 243–259, 2015.

[2] I. A. Essa, "Ubiquitous sensing for smart and aware environments," *IEEE personal communications*, vol. 7, no. 5, pp. 47–49, 2000.

[3] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, no. 1, pp. 381–386, 2020.

[4] S. Gaglio, G. L. Re, and M. Morana, "Human activity recognition process using 3-d posture data," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 586–597, 2014.

[5] 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.

[6] F. Concone, S. Gaglio, G. Lo Re, and M. Morana, "Smartphone data analysis for human activity recognition," in *AI*IA 2017 Advances in Artificial Intelligence*. Cham: Springer International Publishing, 2017, pp. 58–71.

[7] F. Concone, G. Lo Re, and M. Morana, "A fog-based application for human activity recognition using personal smart devices," *ACM Trans. Internet Technol.*, vol. 19, no. 2, Mar. 2019.

[8] F. Concone, C. Ferdico, G. Lo Re, and M. Morana, "A federated learning approach for distributed human activity recognition," in *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2022, pp. 269–274.

[9] J. Bleiholder and F. Naumann, "Data fusion," *ACM computing surveys (CSUR)*, vol. 41, no. 1, pp. 1–41, 2009.

[10] S. Qiu, H. Zhao, N. Jiang, Z. Wang, L. Liu, Y. An, H. Zhao, X. Miao, R. Liu, and G. Fortino, "Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges," *Information Fusion*, vol. 80, pp. 241–265, 2022.

[11] M. Ehatisham-Ul-Haq, A. Javed, M. A. Azam, H. M. Malik, A. Irtaza, I. H. Lee, and M. T. Mahmood, "Robust human activity recognition using multimodal feature-level fusion," *IEEE Access*, vol. 7, pp. 60 736–60 751, 2019.

[12] B. Qi, W. Zhao, X. Wang, S. Li, and T. Runge, "A low-cost driver and passenger activity detection system based on deep learning and multiple sensor fusion," in *2019 5th International Conference on Transportation Information and Safety (ICTIS)*. IEEE, 2019, pp. 170–176.

[13] M. Muzammal, R. Talat, A. H. Sodhro, and S. Pirbhulal, "A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks," *Information Fusion*, vol. 53, pp. 155–164, 2020.

[14] N. Singh and P. Singh, "Exploring the effect of normalization on medical data classification," in *2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)*. IEEE, 2021, pp. 1–5.

[15] H. Hosseinianfar, A. Chizari, and J. A. Salehi, "Gopa: Geometrical optics positioning algorithm using spatial color coded leds (extended version)," *arXiv preprint arXiv:1807.06931*, 2018.

[16] A. Dehghani, O. Sarbishei, T. Glatard, and E. Shihab, "A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors," *Sensors*, vol. 19, no. 22, p. 5026, 2019.

[17] M. McTear, Z. Callejas, and D. Griol, *The Conversational Interface: Talking to Smart Devices*, 1st ed., ser. Engineering, Engineering (R0). Springer Cham, May 2016.

[18] A. Madabhushi and J. Aggarwal, "A bayesian approach to human activity recognition," in *Proceedings Second IEEE Workshop on Visual Surveillance (VS'99)(Cat. No. 98-89223)*. IEEE, 1999, pp. 25–32.

[19] J.-X. Pan, K.-T. Fang, J.-X. Pan, and K.-T. Fang, "Maximum likelihood estimation," *Growth curve models and statistical diagnostics*, pp. 77–158, 2002.

[20] S. Purushotham and B. Tripathy, "Evaluation of classifier models using stratified tenfold cross validation techniques," in *International conference on computing and communication systems*. Springer, 2011, pp. 680–690.

[21] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020.