Approaching Fall Classification Using the UP-Fall Detection Dataset: Analysis and Results from an International Competition

Hiram Ponce and Lourdes Martínez-Villaseñor

Abstract This chapter presents the results of the Challenge UP – Multimodal Fall Detection competition that was held during the 2019 International Joint Conference on Neural Networks (IJCNN 2019). This competition lies on the fall classification problem, and it aims to classify eleven human activities (i.e. five types of falls and six simple daily activities) using the joint information from different wearables, ambient sensors and video recordings, stored in a given dataset. After five months of competition, three winners and one honorific mention were awarded during the conference event. The machine learning model from the first place scored 82.47% in F_1-score, outperforming the baseline of 70.44%. After analyzing the implementations from the participants, we summarized the insights and trends of fall classification.

Keywords Ambient assisted living · Machine learning · Competition · Human fall detection · Abnormal behavioral analysis

1 Introduction

Falls are frequent especially among old people and it is a major health problem according to World Health Organization [2]. Fall detectors can alleviate this problem and can reduce the time in which a person who suffered a fall receives assistance. Recently, there has been an increase in fall detection system development based mainly in sensor and/or context approaches. An important challenge reported in literature [3] is the lack of publicly available datasets that enable comparison between techniques. In that sense, we provide this dataset in the benefit of researchers in the

H. Ponce (✉) · L. Martínez-Villaseñor (✉)
Facultad de Ingeniería, Universidad Panamericana,
Augusto Rodin 498, 03920 Ciudad de México, Mexico
e-mail: hponce@up.edu.mx

L. Martínez-Villaseñor
e-mail: lmartine@up.edu.mx

H. Ponce et al. (eds.), *Challenges and Trends in Multimodal Fall Detection for Healthcare*, Studies in Systems, Decision and Control 273,
https://doi.org/10.1007/978-3-030-38748-8_6

fields of wearable computing, ambient intelligence, and vision. In addition, new machine learning algorithms can be proven with this dataset.

In this competition, participants can to do experiments considering different combination of multimodal sensors in order to determine the best combination of sensors with the aim of improving the reliability and precision of fall detection systems. It is also important for the human activity recognition and machine learning research communities to be able to fairly compare their fall detection solutions.

This competition can be interesting in particular to the growing research community of human activity recognition and fall detection. Moreover, it is also attractive to any person interested in solving signal recognition, vision, and machine learning challenging problems given that the multimodal dataset provided opens many experimental possibilities.

2 Description of the Competition

The *Challenge UP – Multimodal Fall Detection* competition, or simply the competition, was co-located during the 2019 International Joint Conference on Neural Networks (IJCNN 2019). The awarding ceremony of the competition was held on July 15th, 2019 in Budapest, Hungary. However, it was opened from December 3rd, 2018 to April 26th, 2019. The details about this competition are described following.

2.1 Aims and Scope

The competition aimed to classify eleven human activities (i.e. 5 types of falls and 6 simple daily activities) using the joint information from different wearables, ambient sensors and video recordings, stored in a given dataset. This classification was restricted to be done by any, possibly hybrid, machine learning models.

To do so, the competition was scheduled in several steps, mainly for training the model with labeled data and for testing the model with unlabeled data. For convenience, participants were able to use as much information as they wanted. In that sense, the competition dealt with different engineering and computational skills from the participants, through the sensor to image signal processing, the fusion of them, and the abilities to design and deploy different intelligent systems to reach the goal.

2.2 Data

For this competition, we used the UP-Fall Detection dataset [1]. This is a public and large dataset mainly for fall detection and classification that includes 12 activities

Table 1 Statistics of the subjects, adopted from [1]

Subject ID	Age	Height (m)	Weight (kg)	Gender
1	18	1.70	99	Male
2	20	1.70	58	Male
3	19	1.57	54	Female
4	20	1.62	71	Female
5	21	1.71	69	Male
6	22	1.62	68	Male
7	24	1.74	70	Male
8	23	1.75	88	Male
9	23	1.68	70	Female
10	19	1.69	63	Male
11	20	1.65	73	Female
12	19	1.60	53	Female
13	20	1.64	55	Male
14	19	1.70	73	Female
15	21	1.57	56	Female
16	20	1.70	62	Male
17	20	1.66	54	Female

and three trials per activity. Subjects performed 6 simple human daily activities as well as 5 different types of human falls. These data were collected over 17 subjects (see Table 1) using a multimodal approach, i.e. wearable sensors, ambient sensors and vision devices. The consolidated dataset (812 GB), as well as, the feature dataset (171 GB) is publicly available in: http://sites.google.com/up.edu.mx/har-up/. At the time of the competition, the dataset remained private and until April 27th, 2019.

The data was collected over a period of four weeks, in the Faculty of Engineering, Universidad Panamericana in Mexico City, Mexico. During data collection, 17 subjects (9 males and 8 females) of 18–24 years old (i.e. mean height of 1.66 m and mean weight of 66.8 kg), were invited to perform 11 different activities, as shown in Table 2. Falls and daily activities are not overlapped. So, each trial contains information of one of these activities. All the sequences of data was labeled manually. In addition, an *unknown/other* activity was labeled for other unrecognizable activities different from the previous ones [1].

This dataset comprises five Mbientlab MetaSensor wearable sensors collecting raw data from the 3-axis accelerometer, the 3-axis gyroscope and the ambient light value. These wearables were placed in the left wrist, under the neck, at right pocket of the pants, at the middle of waist (in the belt), and in the left ankle. Also, one electroencephalograph (EEG) NeuroSky MindWave helmet was included to measure the raw brainwave signal from one EEG channel sensor located at the forehead. For ambient sensors, the dataset retrieved information from six infrared sensors placed, as a grid, 0.40 m above the floor of the room, to measure the changes in the interruption

Table 2 Types of activities and falls in the dataset

Type	Description	Activity ID
Fall	Forward using hands	1
	Forward using knees	2
	Backward	3
	Sideward	4
	Attempting to sit in an empty chair	5
Daily activity	Walking	6
	Standing	7
	Sitting	8
	Picking up an object	9
	Jumping	10
	Laying	11
Other	Unknown	20

of these devices. Lastly, two Microsoft LifeCam Cinema cameras were located at 1.82 m above the floor, one for lateral view and the other for frontal view, related to the motion of the activities. Table 3 summarizes all the sensors installed for data collection. The dataset was down-sampled to 18 Hz for data synchronization and coherence purposes [1]. Lastly, Fig. 1 shows the placements of wearables, ambient sensors and cameras while collecting the dataset [1]. For further details about the UP-Fall Detection dataset, see [1].

2.2.1 Training Data

For the training data, we exposed the raw dataset from 9 subjects with IDs: 1, 3, 4, 7, 10–14; with all three trials per activity. These data also contained all class labels (activity IDs). The training data set represented 70% of all data considered for this competition. No missing values were presented in the training set.

2.2.2 Testing Data

For the testing data, we exposed the raw dataset from 3 subjects with IDs: 15–17; with all three trials per activity. In this case, data did not contained the class labels. This obeys to the goal of the competition, and the labels of this portion of data remained privately for the participants. In the evaluation step, these labels were used for evaluating the performance of the classification models developed by the participants. No missing values were presented in the testing set.

Table 3 List of devices for measurements, adopted from [1]

Device ID	Device name	Channel name	Units	Signal ID
1	Wearable ankle	X-axis accelerometer	g	1
		Y-axis accelerometer	g	2
		Z-axis accelerometer	g	3
		Roll gyroscope	deg/s	4
		Pitch gyroscope	deg/s	5
		Yaw gyroscope	deg/s	6
		Luminosity	Lux	7
2	Wearable pocket	X-axis accelerometer	g	8
		Y-axis accelerometer	g	9
		Z-axis accelerometer	g	10
		Roll gyroscope	deg/s	11
		Pitch gyroscope	deg/s	12
		Yaw gyroscope	deg/s	13
		Luminosity	Lux	14
3	Wearable waist	X-axis accelerometer	g	15
		Y-axis accelerometer	g	16
		Z-axis accelerometer	g	17
		Roll gyroscope	deg/s	18
		Pitch gyroscope	deg/s	19
		Yaw gyroscope	deg/s	20
		Luminosity	Lux	21
4	Wearable neck	X-axis accelerometer	g	22
		Y-axis accelerometer	g	23
		Z-axis accelerometer	g	24
		Roll gyroscope	deg/s	25
		Pitch gyroscope	deg/s	26
		Yaw gyroscope	deg/s	27
		Luminosity	Lux	28
5	Wearable wrist	X-axis accelerometer	g	29
		Y-axis accelerometer	g	30
		Z-axis accelerometer	g	31
		Roll gyroscope	deg/s	32
		Pitch gyroscope	deg/s	33
		Yaw gyroscope	deg/s	34
		Luminosity	Lux	35
6	EEG headset	Raw brainwave signal	μV	36
7	Infrared 1	No interruption	False(0)/true(1)	37
8	Infrared 2	No interruption	False(0)/true(1)	38
9	Infrared 3	No interruption	False(0)/true(1)	39
10	Infrared 4	No interruption	False(0)/true(1)	40
11	Infrared 5	No interruption	False(0)/true(1)	41
12	Infrared 6	No interruption	False(0)/true(1)	42
13	Camera 1	Lateral view	640×480 px	43
14	Camera 2	Frontal view	640×480 px	44

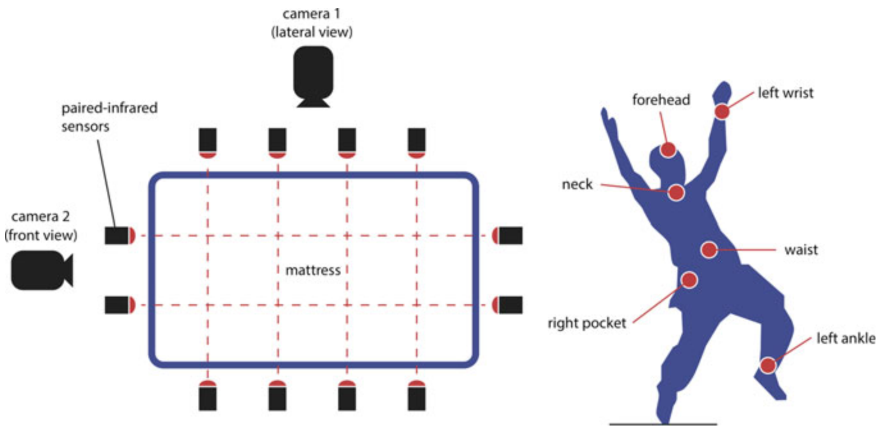

Fig. 1 Layout of the sensors and cameras used in the UP-Fall detection dataset, adopted from [1]

2.3 Classification Task

The main task of the competition is to classify the falls and activities of 3 subjects (testing data set). This is a challenging task since there are diverse of subjects (see Table 1) and they performed activities in different ways. Moreover, the best combination of sensors, feature selection and feature extraction procedures is a challenging task in human activity recognition.

2.4 Metrics and Evaluation

The F_1-score metric was used in the evaluation of the competition. F_1-score considers the average $precision_\mu$ and average $recall_\mu$ of the test as shown in (1), where average $precision_\mu$ computes in average, of all activities and falls, of the number of true positives over the sum of true and false positives; and average $recall_\mu$ computes in average, of all activities and falls, of the number of true positives over the sum of true positives and false negatives. The greater and close to 1, the better the metric.

$$F_1 score = 2 \times \frac{precision_\mu \times recall_\mu}{precision_\mu + recall_\mu} \tag{1}$$

For evaluation, we asked the participants to send the class estimations of the 3 subjects of the testing set. However, these estimations are done in 1-second time window frames. In that sense, the estimated classes were calculated as the most frequent class in 1-second. Similarly, the labels that we retained were also condensed in the most frequent class per 1-second windows without overlapping.

2.5 *Competition Policies*

The following conditions of participation were required during the competition. These policies applied for winning the competition, and the event was divided into several steps through five months of competition; as described below. Participation required complying with the rules of the challenge, published in the official website of the competition (https://sites.google.com/up.edu.mx/challenge-up-2019/).

2.5.1 Conditions of Participation

Prize eligibility was restricted by US government export regulations and the host country laws (Budapest, Hungary). The organizers, sponsors, their students, close family members (parents, sibling, spouse or children) and household members, as well as any person having had access to the truth values or to any information about the data or the challenge design giving him (or her) an unfair advantage, were excluded from participation. However, a disqualified person might submit one or several entries in the challenge and request to have them evaluated, provided that they notify the organizers of their conflict of interest. If a disqualified person submitted an entry, this entry was not be part of the final ranking and did not qualify for prizes.

The participants were aware that organizers reserve the right to evaluate for scientific purposes any entry made in the challenge, whether or not it qualifies for prizes. For participation, the participants registered through the Registration Form displayed in the official website. Teams or solo participants were allowed for entering to the competition.

2.5.2 Awards

The three top ranking participants qualified for awards (travel award, prize and award certificate). To compete for awards, the participants were asked for sending a short paper briefly describing their methods and the codes used for getting the results. There was no other publication requirement. However, this edited book intended to publish the main results of the competition, from the point of view of the participants and the organizers.

2.5.3 Timeline

The competition opened from December 3rd, 2018 until April 26th, 2019. During the five months period, the competition was divided into several steps as shown in Fig. 2. These dates comprised the registration opening (December 3rd, 2018); the training set release (January 14th, 2019) for analyzing and training models by participants; the testing set release (March 25th, 2019) for testing the trained models;

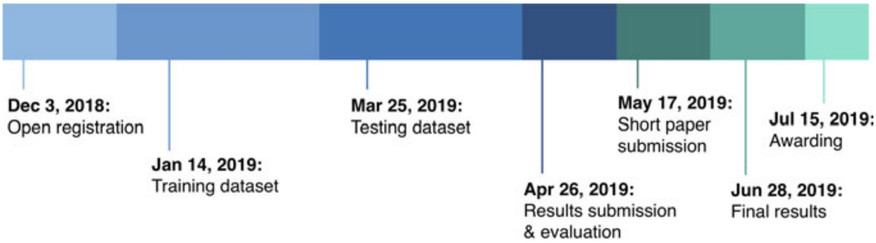

Fig. 2 Timeline of the competition

the submission deadline (April 26th, 2019) for submitting the testing results; the short paper submission deadline (May 17th, 2019) for submitting the complimentary paper describing the way to achieve the challenge; the final decision (June 28th, 2019) for presenting the shortlisted participants; and lastly, the awarding ceremony (July 15th, 2019) for presenting the winners of the competition during the conference IJCNN 2019.

3 Results from the Competition

For this competition, 22 registrations were done (11 as individuals and 11 as teams). Participants were from 14 different countries: Australia, Brazil, China, Estonia, France, Germany, India, Iran, Ireland, Macedonia, Saudi Arabia, Taiwan, Togo and United States of America.

After the results and short paper submission, we announced the three winners of the competition based on the F_1-score metric:

- *First place*: Hristijan Gjoreski (and team) [82.47%]
- *Second place*: Egemen Sahin [34.04%]
- *Third place*: Patricia Endo (and team) [31.37%]
- *Honorific mention*: Vuko Jovicic [60.40%].

The *First place* team used the sensor signals from the wearables. They firstly corrected the orientation of the sensor signals due to the fact that wearables were placed without any particular orientation. After that, they trained three machine learning models, but random forest was the best model that performed 82.47% in F_1-score. Figure 3 shows the confusion matrix of the testing results.

The *Second place* individual tackled the challenge using firstly a standardization of the sensors data (i.e. wearables, ambient sensors and brainwave helmet). Then, he trained 1-dimensional convolutional neural network. After this process, the model performed 34.04% in F_1-score. Figure 4 shows the confusion matrix of the testing results.

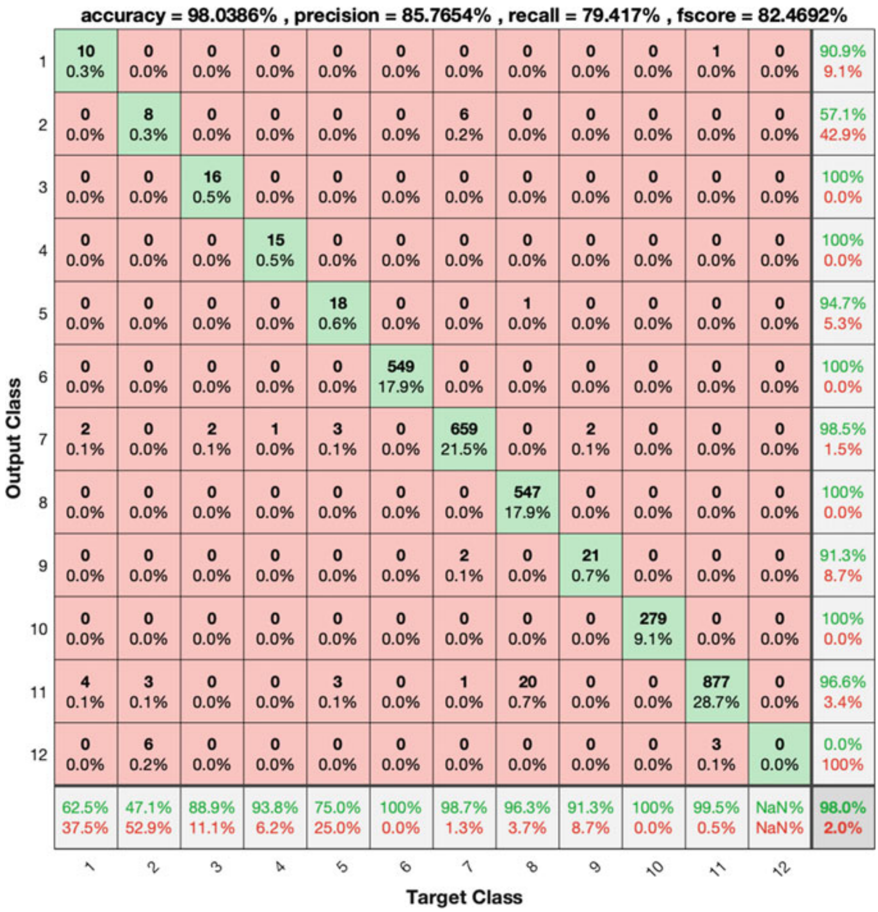

Fig. 3 Confusion matrix of the testing results from *First place*

The *Third place* team employed a bidirectional long short-term memory networks model to achieve the fall classification problem. In this regard, they performed 31.37% in F_1-score. Figure 5 shows the confusion matrix of the testing results.

Lastly, the *Honorific mention* individual obtained a great result in terms of the F_1-score; but, he did not submit the short paper. In this regard, we did not know how he achieved the performance of his model. For that reason, this individual could not be one of the winners. Figure 6 shows the confusion matrix of the testing results.

Although we did not provided a baseline for the participants, we tested four conventional machine learning models: support vector machines (SVM), random forest (RF), multilayer percepron (MLP) and k-nearest neighbors (KNN). This benchmark was published in [1]. We reproduce the baseline in Table 4. As shown, the result from the *First place* is the only one that outperforms the baseline, while the result from *Honorifc mention* is equivalent to the KNN performance.

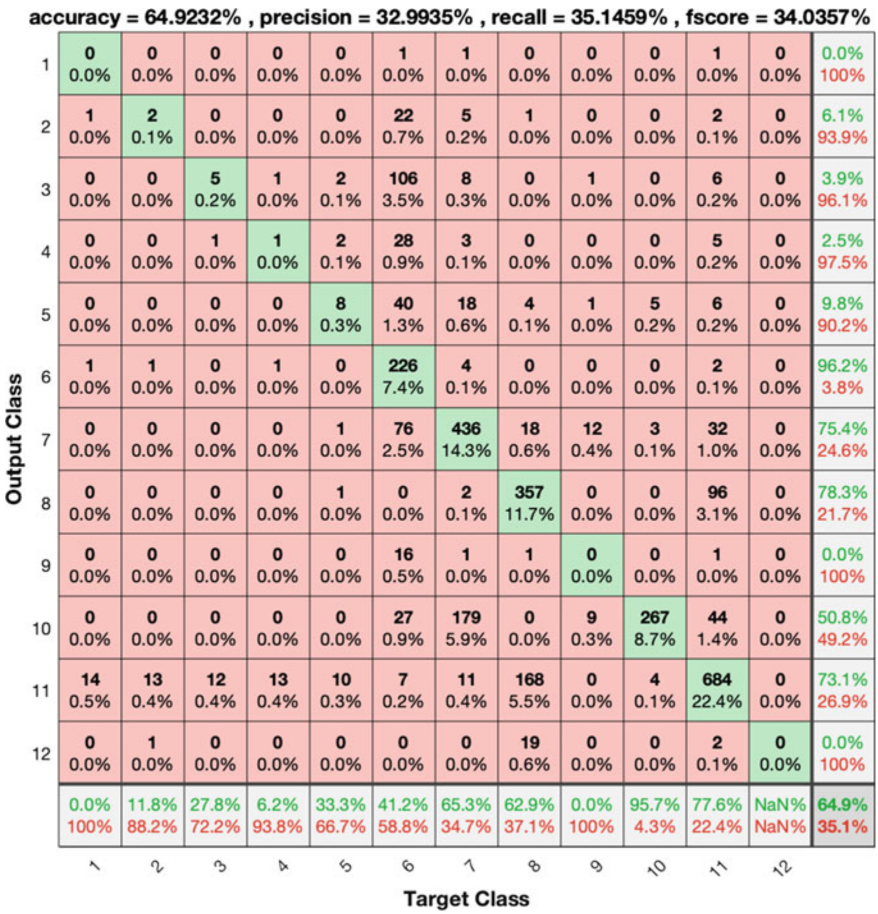

Fig. 4 Confusion matrix of the testing results from *Second place*

Table 4 Baseline using four conventional machine learning models. Values reported are the corresponding F_1-score evaluation, in terms of mean and standard deviation

Model	F_1-score (%)
RF	69.36 ± 1.35
SVM	55.82 ± 0.77
MLP	70.44 ± 1.25
KNN	60.51 ± 0.85

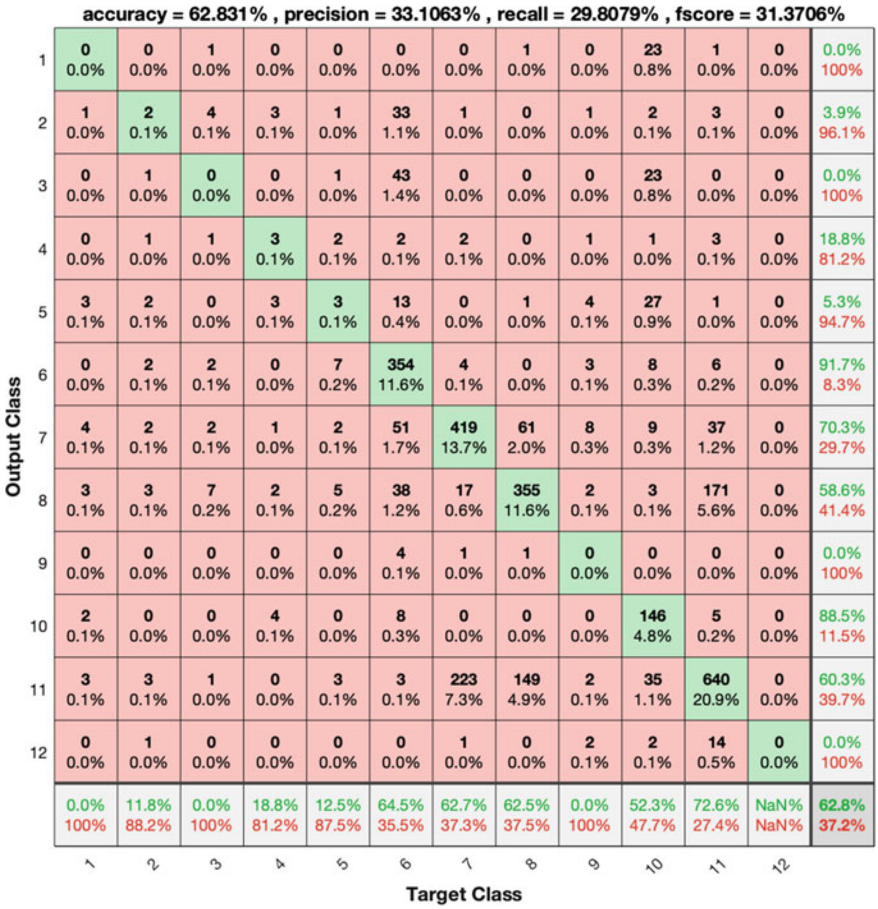

Fig. 5 Confusion matrix of the testing results from *Third place*

4 Concluding Remarks

This competition aimed to propose a multi-class classification model for the problem of human fall classification. In addition, the competition was proposed for challenging participants to apply their computational and machine learning skills in a public, large and multimodal dataset. After the competition ends, we can conclude the following remarks.

In terms of the machine learning models used, it can be seen that conventional machine learning models were employed (e.g. RF, decision trees and KNN). But also, more recent models like convolutional neural networks or bidirectional long short-term memory networks were implemented. Moreover, in terms of the data modality, wearable-based approaches are the most frequent used (i.e. in this competition in

Fig. 6 Confusion matrix of the testing results from *Honorific mention*

all the cases). Ambient sensors were selected in just one attempt. But, cameras were not used by any of the participants. The latter can be associated to the fact that video processing considers complexity and different skills that many of the practitioners do not have. Also, a multimodal approach was not done by any of the participants. It is worth noting that multimodal offers better performance, but it is complex to approach and computationally expensive. In terms of the workflow in data manipulation, participants considers a similar pipeline mainly consisting on: data pre-processing, (temporal) segmentation, feature engineering and training machine learning models. To this end, selection of the best machine learning models and pipelines have to be studied further. Right now, quantitative metrics leads the decision-making process; but this should not be the only criteria for selecting machine learning models and/or strategies to approach fall classification.

On the other hand, the UP-Fall Detection dataset fulfilled the expectations of practitioners in the field of human activity recognition and fall classification. In this regard, this dataset masks the data acquisition problem by giving clean and coherent sensor and camera signals. It can also be used for benchmark machine learning models, as well as different modalities approaches. It is important to highlight that this dataset is publicly available, so practitioners in the field can access and use it as required. Lastly, this dataset provides an important test-bed of machine learning models that can improve the skills of users to develop other applications like in robotics, human-machine interaction, ambient assisted living, among many others.

Finally, fall classification is still an open problem in computer sciences and healthcare, and different open issues have to be faced. For instance, subjects do not perform actions in the same way; but, underlying patterns can be extracted for further analysis. There is some limitation in data since target population is difficult to recruit (e.g. population size, age, type of impairments, etc.). Also, there is highly unbalanced data sets (*falls* vs. *no-falls*). In terms of the sources of information, detection of the best placement of sensors/cameras (and features) is still an issue. Moreover, limitations in resources like computations, memory or budget are constant obstacles in the deployment of these fall classification systems. Of course, there is a need for real-time implementations that should be studied and enhanced. Furthermore, data privacy is still an open concern of fall classification mainly because sensors and cameras are intrusive in daily lives.

Acknowledgements This research has been funded by Universidad Panamericana through the grant "Fomento a la Investigación UP 2018," under project code UP-CI-2018-ING-MX-04.

References

1. Martínez-Villaseñor, L., Ponce, H., Brieva, J., Moya-Albor, E., Núñez-Martínez, J., Peñafort-Asturiano, C.: UP-fall detection dataset: a multimodal approach. Sensors **19**(9), 1988 (2019)
2. Organization, W.H., et al.: Who Global Report on Falls Prevention in Older Age 2007. World Health Organization, pp. 1–7 (2015)
3. Zhang, Z., Conly, C., Athitsos, V.: A survey on vision-based fall detection. In: Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, p. 46. ACM (2015)