Multi-class Multi-label Classification for Cooking Activity Recognition

Shkurta Gashi, Elena Di Lascio, Silvia Santini

Abstract In this paper we present an automatic approach to recognize cooking activities from accelerometer data. We rely on a dataset that contains three-axis acceleration data collected with three devices, including two wristbands, two smartphones and a motion capture system. The data is collected from three participants while preparing *sandwich*, *fruit salad* and *cereal* recipies. The participants performed several fine-grained activities while preparing each recipe such as, e.g., *cut*, *peel*. We propose to use `multi-class` classification approach to distinguish between cooking recipes and a `multi-label classification` approach to identify the fine-grained activities. Our approach achieves 81% accuracy to recognize fine-grained activities and 66% accuracy to distinguish between different recipes using leave-one-subject-out cross-validation. The multi-class and multi-label classification results are 27 and 50 percentage points higher than the baseline. We further show the effect on classification performance of different strategies to cope with missing data and show that imputing missing data with an iterative approach provides 3 percentage points increment to identify fine-grained activities. We confirm findings from the literature that extracting features from multi-sensors achieves higher performance in comparison to using single sensor features.

Shkurta Gashi
Università della Svizzera italiana (USI), Switzerland, e-mail: shkurta.gashi@usi.ch

Elena Di Lascio
Università della Svizzera italiana (USI), Switzerland, e-mail: elena.di.lascio@usi.ch

Silvia Santini
Università della Svizzera italiana (USI), Switzerland, e-mail: silvia.santini@usi.ch

1 Introduction

Automatic detection of physical activities – known as Human Activity Recognition (HAR) [1] – is a key research area in mobile and ubiquitous computing. One of the main aims of HAR is to detect user's behavior with the goal of allowing mobile systems to proactively assist users with their tasks [1]. The wide range of HAR applications include also remote monitoring of daily activities of elderly or cognitively impaired people with the ultimate goal of developing technologies to assist and promote a healthy lifestyle.

Quantity and quality of food intake are particularly crucial factors contributing to a healthy lifestyle [23]. An unhealthy diet may lead to nutrition-related diseases, which in turn can reduce the quality of life [21]. A system able to monitor people's cooking and, thus, eating behavior could provide insightful information to the user towards the improvement of their health. For instance, it could remind elderly people living alone of a missing cooking step or help monitoring of a healthy diet.

While impressive progress has been made in cooking activity recognition and HAR in general, the recognition of complex and fine-grained human activities is still an open research problem [18]. This is first due to the fact that the same activity can be performed in different ways, both by the same person or by different persons [18]. For instance, people perform different arm or wrist postures for holding or picking an object. Another challenge stems from the fact that there is no clear separation between activities but rather continuous motion and repetitive movements [18], which does not allow to segment activities precisely to develop ground-truth for each activity separately [1]. Further, a complex activity might be composed of multiple fine-grained activities, which are characterized by low interclass variability and fine-grained body motions [24, 16]. The majority of work in this area either focuses only on complex or fine-grained activities recognition and do not detect multiple fine-grained activities that occur simultaneously.

In this paper we address the problem of identifying both complex coarse-grained and fine-grained cooking activities. In particular, we propose automatic approaches to distinguish actions performed while preparing three different meals, namely, *sandwich*, *fruit salad* and *cereal*. We refer to these three activities as `macro-activities`. Additionally, we identify several fine-grained activities that occur while preparing each meal such as, e.g., *cut*, *peel*, *take*, *pour*, *put*. We refer to these fine-grained activities as `micro-activities`. To distinguish one macro-activitiy from the others, we develop a multi-class classification pipeline. Since a macro-activity can be composed of multiple micro-activities, we propose to address the problem of micro-activity recognition using a multi-label classification approach. Multi-label learning consists in predicting more than one output category for each input sample [8] and seems appropriate to identify micro-activities occurring while preparing a macro cooking activity. To train and validate our approach, we use an existing data set, presented in [5, 6, 7], which contains acceleration data collected with wristbands, smartphones and a motion capture system.

Our results show that macro-activities can be distinguished with an accuracy of 66% using k-nearest neighbor and leave-one-user-out cross-validation, which is

27 percentage points increment from the baseline classifier that always predicts the majority class. Our multi-label classification approach identifies micro-activities with an accuracy of 81%, which represents a 50 percentage points increment from a baseline that always predicts the most frequent micro-activity.

2 Related Work

In the ubiquitous and wearable research community multiple methods have been proposed for the automatic recognition of human activities [1, 2, 12, 3]. The activities explored in these approaches are rather coarse-grained that include full-body movements such as, e.g., walking, waving or jumping. These activities may not be very relevant for application domains that aim to distinguish between more fine-grained activities such as, e.g., cut and peel and other complex activities such as, e.g., prepare sandwich or salad, as we do in this work.

Several researchers have addressed the specific problem of recognizing cooking activities [4, 13, 16, 23, 18, 19, 21, 22]. Pham et al. [4] propose a real-time approach to classify fine-grained cooking activities such as, e.g., peeling, slicing and dicing, using accelerometer data. Their method achieves an accuracy of 82% using a leave-one-subject-out (LOSO) cross validation approach. Lago et al. [13] investigate the use of single and multiple sensors to distinguish between macro-activities such as, e.g., *setting a table*, *eating a meal*. Their approach achieves an F1 score of 51% using data from multiple sensors and LOSO validation procedure. In contrast to these approaches, we investigate the problem of recognizing both macro and micro-activities performed while cooking different recipes. We use a data set, presented in [5, 6, 7], collected with multiple sensors available in wristbands, smartphones and a motion capture system.

Authors in [18, 19, 20] provide multi-modal sensor data sets of humans performing multiple activities in a kitchen environment, including cooking and food preparation. Tenorth et al. [18], for instance, provide the TUM kitchen data set which includes data such as video sequences, full body motion capture data recorded by a markerless motion tracker, RFID tag readings and more. While these data sets are very diverse and include similar activities and data as the ones used in this work, the authors in [19, 20] do not present any automatic approach to distinguish between cooking activities and the data set used in [18] does not contain data from on-body sensors such as wristband and smartphone, which are less intrusive and more likely to be used by humans while they are cooking their meals. Further these data sets include video or audio data, which are often considered as privacy invasive.

In contrast to the work presented above, we investigate the use of multi-class and multi-label classification for micro and macro-activities recognition. Multi-label classification has been used for cooking ingredients and recipe recognition from images in [25] and for physical activity recognition from accelerometer sensor [26]. To the best of our knowledge, this approach has not been previously applied to cooking activity recognition from accelerometer sensor data.

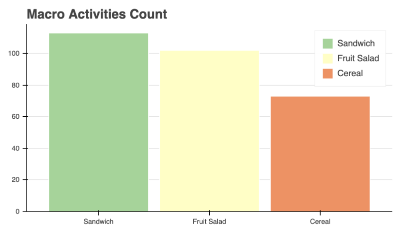

Fig. 1: Overview of the number of samples for each type of macro-activity.

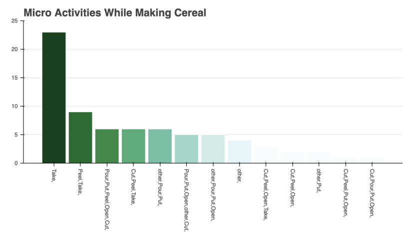

Fig. 2: Number of micro-activities while preparing a cereal.

3 Dataset

We describe in this section the data set used to train and validate the activity classifiers described in the next section.

Procedure and Participants. Four participants performed three macro-activities (meals), five times each. The data has been collected in a controlled setting and participants followed a script for each macro-activity and were asked to act as naturally as possible. The macro-activities were: sandwich, fruit salad and cereal, which included the micro-activities: cut, peel, take, put, pour, wash, mix, and open.

Devices and Collected Data. Movement data during cooking activities is collected using two wristbands – one for each wrist –, two smartphones – for right arm and left hip – and a motion capture system (mocap) with 29 markers. All the sensors collected three-axis acceleration data. The mocap, for instance, represents three-axis acceleration data for markers of different parts of the body such as, e.g., top/front/rear head, left/right shoulder/elbow/wrist/knee and more. The sampling frequency for the sensors in the wristband, smartphone and mocap were 100Hz, 50Hz and 100Hz, respectively. The sensor data is then segmented into 30-second windows and annotated with one of the macro-activities and multiple micro-activities. Figure 1 shows the total amount of macro-activities collected from three participants. In particular there are 113 samples of preparing a sandwich, 102 of fruit salad and 73 of cereal, in total there are 288 macro-activities. Figures 2 to 4 show the micro-activities that occur while preparing each macro-activities. We use the data from three participants to train and test the classifiers.

4 Cooking Activity Recognition Pipeline

The macro and micro-activity recognition pipelines are composed of: data preprocessing, synchronization, imputation strategy, feature extraction, classification,

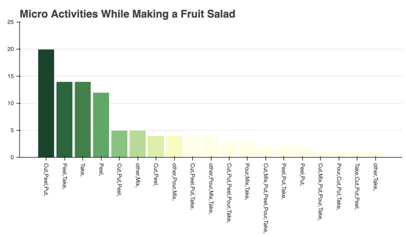

Fig. 3: Number of micro-activities while preparing a fruit salad.

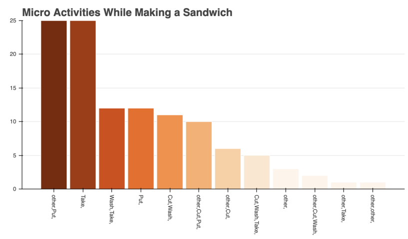

Fig. 4: Number of micro-activities while preparing a sandwich.

and evaluation procedure steps, which are common in HAR [1].

Pre-processing. We first pre-process the raw data acquired by different sensors to be able to analyze them simultaneously. The wristband sensor data has been transmitted to the smartphone via a blacktooth connection. This may cause some sensor data to arrive with delay to the smartphone or even duplicated. To account for this, we first sort the data by the time of measurement and drop duplicates. Given that the wristband and mocap sensors captured data with a sampling frequency of 100Hz and smartphone data with 50Hz, we re-sample wristband and mocap data to 50Hz. To synchronize the sensor readings, we first get the first and last timestamp when the data was collected from all the sensors. We then generate new timestamps with 50Hz sampling frequency and add data from each sensor whenever available.

Imputation Strategy. A basic strategy to handle missing data is to discard all the data where readings from at least one sensor are missing. However this implies loosing valuable and vast amounts of data especially when multiple sensors are used. To account for this issue, we explore different imputation strategies, namely, `mean`, `constant`, `most frequent` and `iterative` imputation explained in [8, 28]. For the first three strategies, we impute missing sensor data using the mean or the most frequent value of the sensor data or a constant value (e.g., 0), the latter has also been explored in [11]. For the iterative imputation strategy, we model each sensor data with missing values as a function of other sensors. In particular the missing sensor data is considered as output y and the other sensor data as input X, then a regressor is fit on (X, y) for all available data and is used to predict the missing values of y. Authors in [10] have also explored this imputation strategy in a different context. We estimate the missing data from one sensor using the data from other available sensors first because the same macro and micro-activity has been performed multiple times by a user. Thereby, the available data when the activity is performed once could be indicative of the missing data when the same activity is performed again because the same user might perform the activity in a similar way. For instance, when cutting food one hand is usually static and the other hand moves (operate the knife). Additionally, multiple sensors have been used to measure the movement in the same part of the body. For instance, the missing values from the left wristband data can be estimated using the mocap marker for the left hand.

Feature Extraction. We reduce the 30-second sensor segments into features that might help discriminate cooking activities recognition. We extract features from each sensor separately as well as combine different sensors. We group the features into two categories: *single-sensor* and *multi-sensors* features. In this way we investigate which features play an discriminative role in distinguishing between cooking activities. Table 1 shows a detailed list of features and modalities we use in this work.

Table 1: Summary of the features extracted from the five sensor modalities used in this paper. L/R stand for left and right, respectively. H stands for hip and A for arm. ACC Magn refers to acceleration magnitude calculated from the X, Y and Z axis. * features extracted from this signal have been used only for micro-activity recognition. ** This feature has been extracted from all the signals except x, y, and z-axis.

<table>
<tr><th>Feature group</th><th>Signal</th><th>Features</th></tr>
<tr><td>Single-sensor</td><td>Wristband (L/R): ACC Magn*

Smartphone (H/A): ACC Magn

Mocap:

X, Y, Z-axis for the following markers
Head (Top/Front/Rear)
Shoulder (L/R)
Elbow (L/R)
Wrist (L/R)
Offset (R)

ACC Magn of the following markers
Shoulder (L/R)
Elbow (L/R)
Wrist (L/R)*</td><td rowspan="2">Statistical features
mean
median
skewness**
minimum
maximum
standard deviation</td></tr>
<tr><td>Multi-sensor</td><td>Direction unit vector between
each axis of the following sensors:
Elbow (L/R)
Wrist (L/R)
Elbow (R) with Wrist (R)
Elbow (R) with Wrist (L)
Elbow (L) with Wrist (R)
Elbow (L) with Wrist (L)

Distance between the following vectors:
Wristbands (L/R)
Wristband and Smartphone: (R/A) and (L/H)
Mocap: Wrist (L) and Elbow (L),
Wrist (R) and Elbow (R), Knee (L/R)</td></tr>
<tr><td>Total features</td><td colspan="2">312 for macro-activity and 342 for micro-activity recognition</td></tr>
</table>

`Single-sensor features.` We first compute the magnitude of acceleration for left and right wrist, left hip, right arm and the markers from the camera as in [27]. We then extract statistical features from each individual signal, i.e., x-axis from the left hand or acceleration magnitude from the right hand. The statistical features we extract are *minimum*, *maximum*, *mean*, *median*, *standard deviation* and *skewness* as suggested in [1, 12]. We hypothesize that the upper body position and movement plays a more significant role to distinguish between different macro and micro-activities because the activities explored in this work include more movements of the upper part of the body e.g., cut, peel, slice. Therefore, from the mocap we expect to see a difference in the features mainly from the first 10 markers (e.g., top front or rear head, right or left shoulder, right offset, right or left elbow, right or left wrist). These features can be further grouped by device in *wristband*, *mocap* and *smartphone* features.

`Multi-sensor features.` We then combine the signals from two sensors available in the same or two different devices. In particular, we compute the direction unit vector between each pair of the following markers from the mocap: top front or rear head, right or left shoulder, right offset, right or left elbow, right or left wrist. We then compute the direction unit vector from signals collected with two devices such as right wrist data collected from the wristband with left arm data collected from the smartphone, similar to [9]. We then extract the statistical features, similar to the single-sensor features, from the directional unit vectors. From the data exploration shown in Figures 2-4, we observe that some activities such as, e.g., take, put, cut are very common for all macro-activities and others such as, e.g., mix, wash and open are unique for each macro-activity. Thereby, we aim at extracting features that could help to better characterize these activities, which would then in turn help to distinguish between macro and micro-activities. We expect the statistical features extracted from the combined multi-sensors could help to distinguish between the classes. For instance, the distance between the left and right hand could be lower when we wash some food than when we mix the salad. Similarly, when we open the milk for the cereal the distance between the elbow and wrist might be different from when we wash the food when making a sandwich.

We then concatenate all the single and multi-sensor features in a single feature vector, known as feature-level fusion approach in HAR [1]. We scale each feature before providing as input to the classifiers, using the standard scaler[1], as a common pre-processing procedure in [8].

Multi-class Macro-Activity Recognition. To distinguish between the *sandwich*, *fruit salad* and *cereal* macro-activities, we set up a multi-class classification problem [8]. We experiment with a range of supervised classifiers including support vector machine, k-nearest neighbours, random forest, decision trees, gradient boosting and multi-layer perceptron. K-nearest neighbors (kNN) achieved the best results, therefore, we report results using only kNN. As a baseline classifier we use a random

[1] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

classifier that always predicts the majority class, used also in similar problems in [14].

Multi-label Micro-activity Recognition. From the Figures 2 to 4 we can observe that each 30-second segment may contain a unique micro-activity such as, e.g., take, peel, or put, or multiple micro-activities such as, e.g., cut/peel/take. For this reason, to recognize the micro-activities in a window we set up a multi-label classification problem. In the multi-label classification the classifier learns from a set of instances, where each instance can belong to one or multiple classes, and is able to predict a set of classes for each new instance [8, 15, 26, 25]. We experiment with a range of supervised classifiers that support multi-label classification as suggested in [8], including k-nearest neighbours, random forest, decision trees, multi-layer perceptron and extra trees classifier. We obtain the best results using k-nearest neighbours (kNN) and we report only those. We consider as a baseline a classifier that always predicts the most frequent micro-activity.

Evaluation Procedure and Metric. We evaluate generalization of our models to new users, by measuring its performance on a subject whose data has not been seen before, known as leave-one-subject-out (LOSO) or person-independent approach [1]. In this approach the classifiers are trained with data of all subjects except one, which is used as test set. This procedure is repeated for all the subjects and the performance of the model is reported as average score across all the iterations. To evaluate the performance of the macro and micro-activity recognition pipelines, we consider the accuracy metric [8]. Accuracy quantifies the fraction of samples correctly classified by the model. For the multi-label classification problem, we first compute the accuracy for each test sample and average the results to obtain an overall metric.

5 Results and Discussion

In what follows we report the classification results obtained applying the analysis described in the previous section. In particular, we report the macro and micro-activity recognition results, the difference of features between different activities, the performance obtained with different imputation strategies and feature groups.

5.1 Macro and Micro-activity Recognition Results

Figure 5 shows the multi-class classification results for the baseline classifier and kNN classifier trained using all the features described in Section 4 and applying the LOSO validation procedure. The accuracy for random forest is 66%, which is 27 percentage points increment from the baseline. This implies that we are able to correctly identify 66% of the macro-activities in the data set. Figure 6 shows the

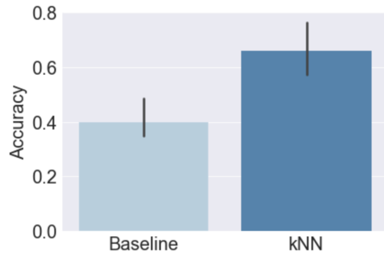

Fig. 5: Accuracy of the multi-class kNN classifier and baseline for distinguishing between macro-activities.

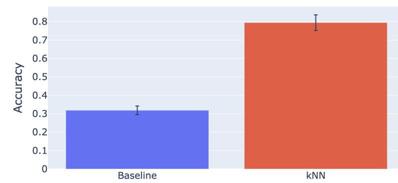

Fig. 6: Accuracy of the kNN classifier and baseline for identifying micro-activities using multi-label classification approach.

multi-label classification results for the baseline classifier and kNN trained using all the features described in Section 4 and applying the LOSO validation procedure. The accuracy for kNN is 81%, which is 50 percentage points increment from the baseline classifier that always predicts the most frequent micro-activity. These results imply that we can recognize 81% of the micro-activities that occur while preparing a macro-activity.

5.2 Interpretation of Cooking Activities Features

We then investigate the difference between the features for different macro and micro-activities. In this section we present some exemplary features and their difference among different classes.

Macro-activity features. Figure 7 shows the difference of the mean acceleration magnitude of the right hand between the three classes. In particular we can observe that while preparing a salad there are less right hand movement rather than when preparing a sandwich. This is because when we prepare the salad we perform more wrist movements such as, e.g., cut, peel, put – as also shown in Figure 3 –, whereas when preparing a sandwich and cereal we perform more full body and intense movements such as, e.g., take and wash, as shown in Figures 2 and 4.

Micro-activity features. Figure 8 shows the difference of the mean acceleration magnitude for the right wrist for different micro-activities. We observe that the average movement magnitude is higher for take and wash. This is expected as these two activities require full body movement or movement of both hands in comparison to the others such as, e.g., open, when we mainly use the wrist. It is also interesting to see that the activity 'other' contains many outliers. This might be due to the nature of this activity, which may include several micro-activities ranging from light to vigorous movement.

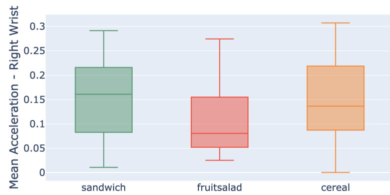

Fig. 7: Distribution of the mean acceleration magnitude from the right wrist for *sandwich*, *fruit salad*, and *cereal* macro-activities.

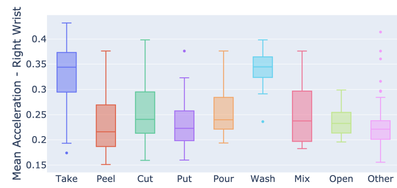

Fig. 8: Distribution of the mean acceleration magnitude from the right wrist for different micro-activities explored in this work.

Imputation strategy	Accuracy
None	0.78
Most frequent	0.79
Constant	0.80
Mean value	0.80
Iterative	**0.81**

Table 2: Multi-label classification results for iterative, mean, constant, most frequent and no imputation strategies.

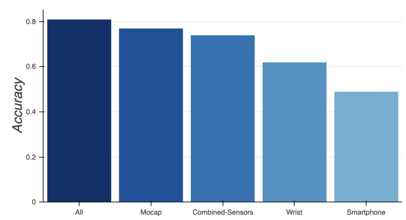

Fig. 9: Multi-label classification results using features extracted from all modalities (single and multi-sensor), from wristband, mocap and smartphone only and for using only multi-sensor features.

5.3 Performance with Different Imputation Strategies

We also investigate the impact of the imputation strategies adopted in the cooking activities recognition. Table 2 shows the multi-label classification results using the imputation strategies explored in this work. We achieve 81% accuracy when using the iterative imputation strategy, which is 3 percentage points higher than when not using an imputation strategy. The increment by 3 percentage points in performance hints at the importance of using data imputation and the suitability of iterative imputation for improving recognition performance. The performance when using other imputation strategies is 80% for the mean and constant, and 79% for most frequent, which is still slightly higher than no data imputation. We observe similar results for the macro-activity recognition but for simplicity we decide to report only the multi-label classification performance.

5.4 Single-sensor and Multi-sensor Features Performance

We then evaluate the performance of micro-activity recognition while cooking using a single and multi-sensor features. Figure 9 shows the multi-label classification results using features extracted from one sensor modality alone or combining multiple sensor modalities. We obtain the best results using the features extracted from all sensors with an accuracy of 81%, which is in line with findings from literature [13]. This confirms the necessity to have multiple sources of data to capture the characteristics of different and more complex activities, as also discussed in [17]. The performance using only the data from the motion capture system is comparable to using the data from all the sensors. This implies that in case of missing sensors the motion capture system could be used alone.

6 Limitations and Future Work

While this work shows promising results in the identification of macro and micro cooking activities, future research is needed to overcome the limitations of our approach. A limitation stems from relying on the usage of sensors from multiple devices. While using multiple devices enhances the recognition performance and data quality, in a real system all the devices might not be available all the time. Future work should focus on optimizing the performance with the fewest number of devices. Additionally, our macro and micro-activity identification pipelines rely on hand-crafted features and do not capture the temporal nature of the sensor data. Future work should focus on exploration of deep learning, as in [2, 3] methods to automatically extract features (i.e., using convolutional neural networks) and to consider the sequential nature of the data (i.e., using long short-term memory neural networks), which might more effectively identify micro-activities that occur in a sequential manner.

7 Conclusions

In this paper we present our approach on automatic recognition of macro and micro-activities while cooking using acceleration data from multiple sensors. We show that it is feasible to distinguish between macro-activities with 66% accuracy using a multi-class random forest classifier. We further show that our multi-label classification approach can recognize micro-activities while cooking with an accuracy of 81%. We then show that using data from other sensors to predict missing sensor data increases the performance by 3 percentage points and could be an interesting direction for future research. We confirm findings from related work that using data from multiple sensor modalities performs significantly higher than using some sensors alone.

Overall, our findings enable new possibilities in the design and development of automatic systems for supporting people in their daily activities.

References

1. Bulling, Andreas, Ulf Blanke, and Bernt Schiele. "A Tutorial on Human Activity Recognition using Body-worn Inertial Sensors." ACM Computing Surveys (CSUR) 46, no. 3 (2014): 1-33.
2. Radu, Valentin, Catherine Tong, Sourav Bhattacharya, Nicholas D. Lane, Cecilia Mascolo, Mahesh K. Marina, and Fahim Kawsar. "Multimodal Deep Learning for Activity and Context Recognition." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, no. 4 (2018): 1-27.
3. Guan, Yu, and Thomas Plötz. "Ensembles of Deep LSTM Learners for Activity Recognition Using Wearables." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, no. 2 (2017): 1-28.
4. Pham, Cuong, Thomas Plötz, and P. Oliver. "Real-time Activity Recognition for Food Preparation." In Proceedings of the IEEE International Conference on Computing and Communication Technologies, Nagercoil, Tamil Nadu, India. 2010.
5. Paula Lago, Shingo Takeda, Kohei Adachi, Sayeda Shamma Alia, Moe Matsuki, Brahim Benai, Sozo Inoue, Francois Charpillet, "Cooking Activity Dataset with Macro and Micro Activities", IEEE Dataport, doi: 10.21227/hyzg-9m49, 2020.
6. Paula Lago, Shingo Takeda, Sayeda Shamma Alia, Kohei Adachi, Brahim Benaissa, François Charpillet and Sozo Inoue, 'A Dataset for Complex Activity Recognition with Micro and Macro Activities in a Cooking Scenario', preprint, 2020.
7. Sayeda Shamma Alia, Paula Lago, Shingo Takeda, Kohei Adachi, Brahim Benaissa, Md Atiqur Rahman Ahad and Sozo Inoue, 'Summary of the Cooking Activity Recognition Challenge', Human Activity Recognition Challenge, Smart Innovation, Systems and Technologies, Springer Nature, 2020.
8. Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2019.
9. Ahuja, Karan, Dohyun Kim, Franceska Xhakaj, Virag Varga, Anne Xie, Stanley Zhang, Jay Eric Townsend, Chris Harrison, Amy Ogan, and Yuvraj Agarwal. "EduSense: Practical Classroom Sensing at Scale." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3, no. 3 (2019): 1-26.
10. Saha, Koustuv, Manikanta D. Reddy, Vedant das Swain, Julie M. Gregg, Ted Grover, Suwen Lin, Gonzalo J. Martinez et al. "Imputing Missing Social Media Data Stream in Multisensor Studies of Human Behavior." In 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 178-184. IEEE, 2019.
11. Jaques, Natasha, Sara Taylor, Akane Sano, and Rosalind Picard. "Multimodal Autoencoder: A Deep Learning Approach to Filling in Missing Sensor Data and Enabling Better Mood Prediction." In 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 202-208. IEEE, 2017.
12. Janko, Vito, Nina Rešçiç, Miha Mlakar, Vid Drobnič, Matjaž Gams, Gašper Slapničar, Martin Gjoreski, Jani Bizjak, Matej Marinko, and Mitja Luštrek. "A New Frontier for Activity Recognition: The Sussex-Huawei Locomotion Challenge." In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, pp. 1511-1520. 2018.
13. Lago, Paula, Moe Matsuki, and Sozo Inoue. "Achieving Single-Sensor Complex Activity Recognition from Multi-Sensor Training Data." arXiv preprint arXiv:2002.11284 (2020).
14. Meurisch, Christian, Artur Gogel, Benedikt Schmidt, Timo Nolle, Frederik Janssen, Immanuel Schweizer, and Max Mühlhäuser. "Capturing Daily Student Life by Recognizing Complex Activities Using Smartphones." In Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, pp. 156-165. 2017.
15. Sorower, Mohammad S. "A Literature Survey on Algorithms for Multi-label Learning." Oregon State University, Corvallis 18 (2010): 1-25.
16. Rohrbach, Marcus, Sikandar Amin, Mykhaylo Andriluka, and Bernt Schiele. "A Database for Fine Grained Activity Detection of Cooking Activities." In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1194-1201. IEEE, 2012.

17. Zinnen, Andreas, Ulf Blanke, and Bernt Schiele. "An Analysis of Sensor-oriented vs. Model-based Activity Recognition." In 2009 International Symposium on Wearable Computers, pp. 93-100. IEEE, 2009.
18. Tenorth, Moritz, Jan Bandouch, and Michael Beetz. "The TUM Kitchen Data Set of Everyday Manipulation Activities for Motion Tracking and Action Recognition." In 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, pp. 1089-1096. IEEE, 2009.
19. De la Torre, Fernando, Jessica Hodgins, Adam Bargteil, Xavier Martin, Justin Macey, Alex Collado, and Pep Beltran. "Guide to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database." (2009).
20. Roggen, Daniel, Alberto Calatroni, Mirco Rossi, Thomas Holleczek, Kilian Förster, Gerhard Tröster, Paul Lukowicz et al. "Collecting Complex Activity Datasets in Highly Rich Networked Sensor Environments." In 2010 Seventh International Conference on Networked Sensing Systems (INSS), pp. 233-240. IEEE, 2010.
21. Whitehouse, Samuel, Kristina Yordanova, Adeline Paiement, and Majid Mirmehdi. "Recognition of Unscripted Kitchen Activities and Eating Behaviour for Health Monitoring." (2016): 1-6.
22. Yordanova, Kristina, Samuel Whitehouse, Adeline Paiement, Majid Mirmehdi, Thomas Kirste, and Ian Craddock. "What's Cooking and Why? Behaviour Recognition During Unscripted Cooking Tasks for Health Monitoring." In 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 18-21. IEEE, 2017.
23. Yordanova, Kristina, Stefan Lüdtke, Samuel Whitehouse, Frank Krüger, Adeline Paiement, Majid Mirmehdi, Ian Craddock, and Thomas Kirste. "Analysing Cooking Behaviour in Home Settings: Towards Health Monitoring." Sensors 19, no. 3 (2019): 646.
24. Rohrbach, Marcus, Anna Rohrbach, Michaela Regneri, Sikandar Amin, Mykhaylo Andriluka, Manfred Pinkal, and Bernt Schiele. "Recognizing Fine-grained and Composite Activities Using Hand-centric Features and Script data." International Journal of Computer Vision 119, no. 3 (2016): 346-373.
25. Bolaños, Marc, Aina Ferrà, and Petia Radeva. "Food Ingredients Recognition Through Multi-label Learning." In International Conference on Image Analysis and Processing, pp. 394-402. Springer, Cham, 2017.
26. Mohamed, Raihani. "Multi-label Classification for Physical Activity Recognition from Various Accelerometer Sensor Positions." Journal of Information and Communication Technology 17, no. 2 (2020): 209-231.
27. Leeger-Aschmann, Claudia S., Einat A. Schmutz, Annina E. Zysset, Tanja H. Kakebeeke, Nadine Messerli-Bürgy, Kerstin Stülb, Amar Arhab et al. "Accelerometer-derived Physical Activity Estimation in Preschoolers–comparison of Cut-point Sets Incorporating the Vector Magnitude vs the Vertical Axis." BMC public health 19, no. 1 (2019): 513.
28. Burkov, Andriy. The Hundred-page Machine Learning Book. Quebec City, Can.: Andriy Burkov, 2019.