Improved Strategies for Multi-modal Atmospheric Sensing to Augment Wearable IMU-Based Hand Washing Detection

Robin Burchard(✉), Hurriat Ali, and Kristof Van Laerhoven

University of Siegen, 57076 Siegen, Germany
robin.burchard@uni-siegen.de

Abstract. Previous work in the area of hand washing detection has hinted at the usefulness of atmospheric sensors for hand washing detection. Specifically, a humidity sensor can be used to measure nearby tap water flow using wearable devices. For this work, we expand on previous findings and a pre-existing dataset by recording 10 additional participants with a self-made open-source prototype recording device. We introduce an updated dataset with 20 participants instead of 10 participants, for which we make available IMU, humidity, temperature, and pressure measurements. The newly recorded participants conducted more complex background activities, which increased our dataset's real-world relevance. Additionally, we show how to train an optimized deep-learning-based classifier on different parts of the combined dataset, improving on the previous study's results, achieving significantly better F1 scores (82% instead of 70%) on the pre-existing dataset. Furthermore, by leveraging a BIO-BANK semi-supervised pretrained model, we show that, unlike in previous work, the addition of humidity sensors to IMU data has a positive impact on the classification performance on the old and the new dataset, improving the F1-score on the combined dataset from 60% to 68%. All code and data are publicly available on GitHub.

Keywords: Multi-Modal · Hand Washing Detection · Atmospheric Sensing · Human Activity Recognition · Humidity Sensor · Data Recording · Open Source

1 Introduction

The detection of hand washing is related to multiple applications in our everyday lives. A system that can automatically detect and analyze hand washing frequency, duration, and performance would be useful and could act as a personal hygiene assistant. Additionally, a similar system could be employed in professional environments, especially in the food industry and in the medical

R. Burchard and H. Ali—Contributed equally to this research.

Ö. Durmaz Incel et al. (Eds.): iWOAR 2025, LNCS 16292, pp. 308–323, 2026.
https://doi.org/10.1007/978-3-032-13312-0_18

domain. Properly washing one's hands has been shown to dramatically reduce the spread of pathogens in the population [7]. Therefore, a system that detects hand washing over the day could help its user to maintain a high level of hand hygiene, protecting the user and their surroundings.

Hand washing detection can also be employed in the context of obsessive-compulsive disorder, where overly frequent hand washing has a negative impact [5,21,22]. There, a system could help by logging hand washing occurrences over the day, by helping the user understand how often and how long they wash their hands or by providing valuable insights to treatment experts.

Both RGB-camera-based and wearable sensor-based methods exist to detect hand washing. However, the use of cameras comes paired with privacy concerns, and IMU sensor-based detection suffers from the ambiguity of rapid movement patterns [5]. Some devices, like the Apple Watch [12] have built-in proprietary hand washing detectors, but there exists no data on their reliability. While traditional sensor-based human activity recognition systems often rely on IMU data only, this work focuses on making use of the contextual information, which can be provided by additional modalities. One example would be Bluetooth beacons placed in a users home, which provide a clue about the location, but come with the downside of being constrained to environments where beacons can be placed [19]. We therefore opt for a modality that we can easily measure on-device, everywhere, with affordable, highly accurate sensors: Humidity. Our previous, preliminary study, for which we investigated multiple environmental sensors (humidity, pressure, temperature), had shown promising results, as we could find a clear pattern in the humidity signal when a participant was washing their hands [6]. Thus, for this work, we recorded 10 additional participants while they were conducting different everyday activities and while they were washing their hands. Whereas our previous study was solely a feasibility study, in which the additional modalities could not be shown to improve the classification performance, we were able to outperform the previous study's preliminary results substantially by including additional participants, more diverse background activities, and applying a more sophisticated deep learning pipeline.

1.1 Goals and Contributions

The goal of this work was to improve on the previous study's preliminary results and further evaluate the usefulness of the addition of a humidity sensor.

Our contributions are threefold:

1. Recording, labeling, comparing, and making available data from 10 more participants with more diverse background activities.
2. In-depth evaluation of the usefulness of the humidity sensor in hand washing detection, on previously available and newly recorded data, significantly improving on the previous study's results
3. Employing and evaluating more complex network architectures, including the application of a pre-trained model, to achieve significantly improved classification results.

2 Related Work

While some works on hand washing detection only using IMUs exist, they are mostly in-lab studies or constrain the user to hand washing patterns recommended by the WHO, which does not cohere well with real-world hand washing without artificial constraints [19,23]. For other studies that also focus on unstructured hand washing recognition and achieve high classification accuracies [13], other limitations apply, such as small sample sizes or the absence of leave-one-out validation patterns. Thus, we will focus on multi-modal approaches for unconstrained hand washing detection in this work.

2.1 Sensors for Multi-modal Activity Recognition

While human activity recognition (HAR) can be approached with a single sensor modality, such as the commonly used RGB(D)-cameras or IMUs, previous work in multi-modal HAR exists and employs a multitude of sensor modalities. Combining multiple sensing modalities leads to higher classification performance due to the usually provided additional context, but introduces additional complexity [10]. The most commonly fused sensors include IMU and RGB(D)-cameras, or other visual systems in various positions (e.g., body worn or stationary in a task-specific location). Additional modalities include audio, environmental sensors such as temperature, humidity, barometer, or light, and physiological sensors such as measuring oxygen saturation, heart rate, or electrocardiography [9,10,16]. However, not all sensing modalities can be applied for all applications and in all environments. Especially, cameras and microphones are ethically difficult, as they record data of their surroundings, including, e.g., private conversations. For our research interest in general-purpose omnicontextual hand washing detection, cameras are hardly feasible, due to their inappropriateness in bathrooms and many public spaces. We thus conclude that in hand washing detection, the needed additional context should be provided by privacy-preserving modalities. E.g., microphone data, as used in a preliminary lab study by Zhuang et al. [26], would need to be processed on-device and then discarded. A good basis for microphone-aided hand washing detection could be offline tap water audio detection [4]. Unlike cameras and microphones, the atmospheric data we utilized for this work is anonymous by default, and therefore does not pose a challenge to the users' privacy.

2.2 Humidity Sensing for Activity Recognition

Any activity detection problem related to changes in ambient humidity could likely profit from humidity sensors. Oftentimes, humidity sensors are paired with other atmospheric sensors such as barometers and temperature sensors to enhance the classification of activities of daily living [2,8,20]. In these works, the atmospheric sensors are applied together with other modalities such as IMU recordings. The atmospheric sensors aid the classification by providing additional context to the otherwise ungrounded IMU recordings.

Picking up on this idea, in a recent work, we proposed WearPuck, a wearable sensing platform that synchronously records accelerometer, gyroscope, humidity, temperature, and barometric pressure [6]. We applied the novel fully open-source data collection device to the task of hand washing detection, in an experiment with 10 participants and a total of 40 hand washes. Although it was shown that especially the recorded humidity changes measurably during hand washing instances, the prediction performance of simple machine learning classifiers did not improve consistently with added humidity features. Therefore, we concluded that additional efforts in data recording, data processing, and machine learning were the logical next step. To the best of the authors' knowledge, no other work utilizing humidity sensors and IMUs for hand washing detection has been published.

2.3 Machine Learning for Multi-modal Activity Recognition

Fusion Strategies. In multi-modal activity recognition, one of the main problems is the question of when and how to fuse the different modalities' signals. In early fusion, the input modalities are fused before passing them into the machine learning models. Early fusion enables the models to learn the different sensors' co-dependencies jointly. In late fusion, the modalities are processed separately and are only fused at the decision stage [10], so that the models initially learn independent features for all modalities, before fusing these representations and classification. Münzner et al. showed that for the PAMAP2 dataset, late fusion performs better than early fusion [15], but the performance difference was small, and the best method must likely be determined for each dataset and modality combination separately.

Learning Strategies. In HAR tasks, labeled data has to be obtained with great effort and is therefore scarce, leading to small labeled datasets. As deep-learning models require extensive amounts of data to be trained, these small datasets are suboptimal. One solution can be sought in semi-supervisedly pre-trained models, which were trained on large amounts of unlabeled data and which are able to either provide good embeddings of the modality for downstream tasks or can be fine-tuned for a specific task and dataset. Their performance is usually significantly higher than for models trained only on the downstream dataset [25]. Thanks to the scientific community, many such models are freely available. One example of such a self-supervised pre-trained model is the HARNet model by Yuan et al. [24], which we also employed for this publication. The HARNet model is based on ResNet and was trained on 700,000 person-days of accelerometer data taken from the UK Biobank accelerometer dataset.

3 Dataset Expansion: Recording and Validation

3.1 Collection of New Data

The previously collected data from 10 participants served as a good baseline for the initial proof-of-concept. While the inclusion of additional modalities did not improve the preliminary machine learning performance, we were able to highlight a distinct response pattern of the humidity sensor to hand washing.

To create a better model and train more efficiently, additional data was required. Hence, we collected data from 10 additional participants, who had not partaken in the first experiment. All participants were volunteers and signed an informed consent form. The previous data collection included a lot of sitting and desk work in its recording procedure, interrupted by short walks, hand washing, and stair walking, as well as a few other activities. To increase the variability of the dataset, we also enforced additional background activities during the recording. By doing so, we increased the difficulty of classification, as there are more movement patterns, and the new dataset contains more active behaviour. The background activities include:

- going up and down the stairs (as in previous work)
- playing the guitar (new)
- playing with a ball (new)
- playing video games (new)
- washing dishes (new)
- other natural movements that involve active hand use (new)

As a result of this increased diversity, the classification task became more challenging. However, this also enhanced the ecological validity of the dataset, as the recorded movements better reflect real-world behavior.

The data was collected using our own, open-source wearable device, WearPuck[1]. The technical aspect of the data recording and labeling procedure was identical to the one described in our previous publication [6]. Thus, we refer the interested reader to this publication for a more detailed description of the recording, labeling, and post-processing steps.

3.2 Validation and Comparison to Existing Dataset

The newly recorded n = 10 participants (7f, 3 m, aged between 20 and 30 years) performed a total of 39 hand washing instances. Thus, the newly recorded dataset contains roughly the same amount of data as we recorded for the same duration with the same number of participants. However, as explained in Sect. 3.1, the background activities were much more diverse. Additionally, the durations of the hand washes were different, with a mean duration of 45 s.

Figure 1 shows the response of the recorded humidity values to the hand washing. Both the pre-existing dataset and the newly collected data show a similar behaviour, with a peak of around +8%-points between 20 to 50 s after the beginning of a hand washing instance. As previously reported, the humidity starts to increase immediately when hand washing starts. After peaking, the humidity signal starts to decline again. We expect that the hand washing mostly ends a short time ahead of the peak, which is better visible in the subplot (a) of Fig. 1. For the newly collected data, the trend is still visible, but slightly less

[1] https://github.com/kristofvl/WearPuck.

pronounced, due to the higher diversity in hand washing duration, which spreads out the end of hand washing instances more.

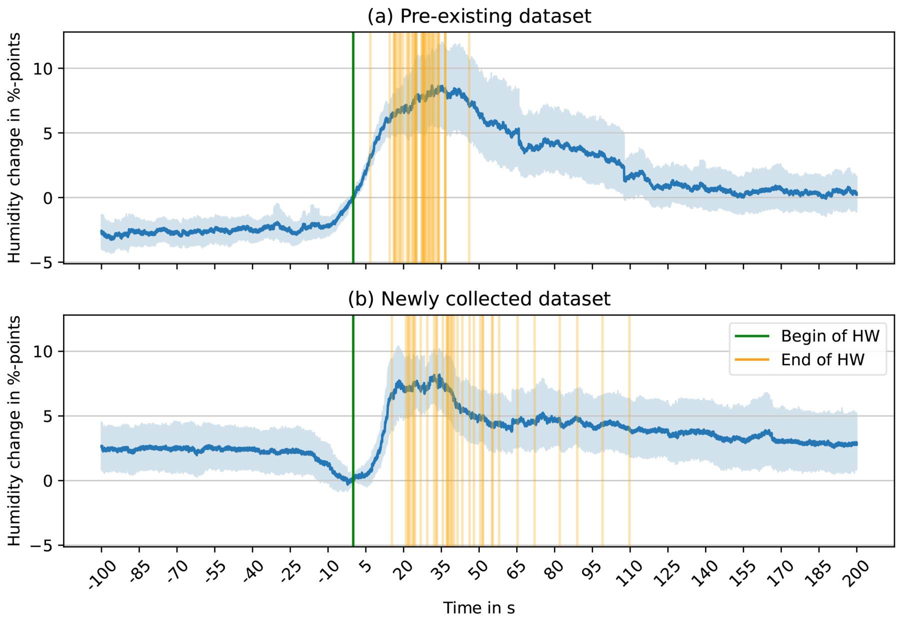

Fig. 1. Response of the humidity sensors to hand washing, averaged (in dark blue) over all recorded hand washes of (a) the pre-existing dataset and (b) the newly collected dataset, with bootstrapped 95% confidence interval (in light blue). The start of the hand washing is marked with a green vertical line. The yellow vertical lines mark the respective ends of all the handwashing instances. (Color figure online)

Figure 2 displays the distribution of hand washing durations across the pre-existing, the newly recorded, and the combined datasets. The hand washing durations differ, with the newly collected participants washing for 45 s on average, compared to 25 s in the previous data collection. The new data collection also includes some outlier durations of up to 109 s. This comparatively long duration was not enforced or encouraged by the conducting experiment supervisors, as participants were asked to wash their hands as they normally would, if they felt "dirty". After combining the datasets, we ended up with a mean hand washing duration of 35 s (median: 30 s). The minimum hand washing duration of 6.5 s was not undercut by the newly recorded participants. The maximum hand washing duration in the combined dataset is 109 s.

As shown in Table 1, the newly recorded dataset contains more hand washing than the pre-existing dataset, thanks to the participants washing for longer durations. The number of instances of hand washing is similar, but the newly recorded participants contribute more washing data to the combined dataset.

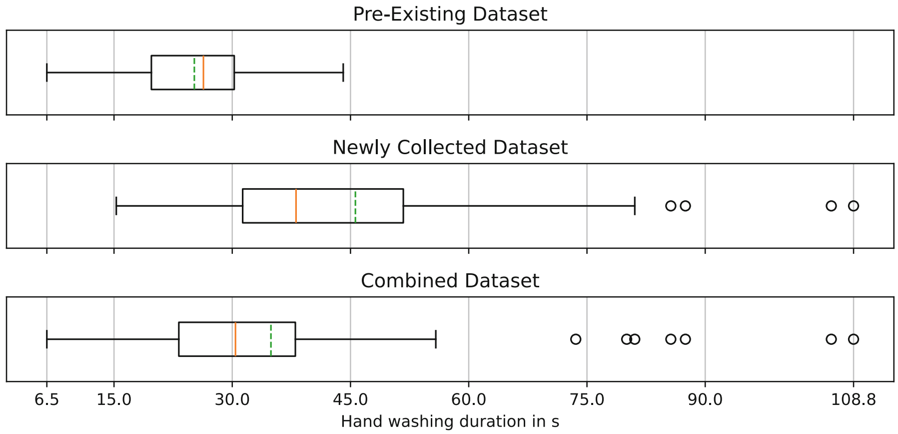

Fig. 2. Statistics for the duration in seconds of all existing and newly recorded hand washes. The box plot shows the median (orange solid line), mean (green dashed line), quartiles (box extents), outliers (circles), and minimum and maximum durations inside 1.5 times the inter-quartile range (whiskers). (Color figure online)

Table 1. Recorded minutes of activities and ratio of hand washing for the pre-existing and the newly recorded dataset. The new dataset contains 1.64 times as many minutes of hand washing as the pre-existing dataset. The combined dataset contains 47.7 min of hand washing, which makes up for around 4% of the recorded 1240 min.

Dataset	Recording duration (min)	Hand washing duration (min)	Ratio (%)
Pre-existing	612.18	18.05	2.95
Newly recorded	627.99	29.65	4.72
Combined	1240.17	47.71	3.85

The combined dataset now has a total length of 1240 min (almost 21 h), and contains 48 min (4%) of hand washing. Altogether, this means that the total hand washing duration in the combined dataset has now become 2.64 times as large as in the pre-existing dataset only. The dataset remains highly imbalanced, with 96% of it belonging to the class of background activities. However, hand washing is still over-represented in comparison to what could be expected in-the-wild.

Overall, the newly recorded data significantly enlarges our database and contains more variety in the durations of handwashing. Meanwhile, the effect of the humidity rising while washing the hands is equally present in the new data as in the previously collected. Thus, we conclude that the measured effect seems to generalize well to diverse other environments and participants.

4 Methods

For this publication, we revisited the previously recorded dataset and applied a better deep learning model to the hand washing classification with and without the added humidity modality. Afterward, we combined the data of the newly collected 10 additional participants with the existing dataset and developed an additional optimal model for this combined dataset. All code and data are publicly available in our GitHub repository[2].

4.1 CNN-GRU Model and Training Procedure

Due to the usually high performance of previous combinations of CNNs with recurrent network parts like DeepConvLSTM [3,18], we opted for a CNN model with a gated recurrent unit (GRU). The model with the GRU slightly outperformed DeepConvLSTM in preliminary testing. The full network consisted of two one-dimensional convolutional layers, followed by max-pooling, the GRU, a 4-headed attention mechanism, layer norm, global average pooling, and two dense layers with dropout ($p = 0.5$) as the classification head. We trained it on the IMU data (accelerometer, gyroscope) with and without the added modality of humidity, on sliding windows (5 s, 50% overlap).

To make use of the humidity data, we extracted six types of features from the sliding windows of sensor readings that capture both the statistical structure and temporal dynamics of the signal. The mean and standard deviation quantify the overall humidity level and local variability, to capture and identify the characteristic rise and fluctuation patterns we associate with hand washing. We use frequency-domain features derived from the Fourier transform to encode periodic components introduced by repetitive patterns, which the humidity sensor might pick up. Higher-order statistics such as skewness and kurtosis capture asymmetric distributions and heavy tails, as likely caused by sharp humidity spikes. We also calculated the mean of first-order differences, i.e., the mean steepness between two consecutive sensor values, which reflects the rate of change over time, to detect the rapid increase and the decrease while and after washing the hands (c.f. Fig. 1). While this worked well on the pre-existing dataset, we developed a more elaborate set of features for the combined dataset, which we describe below.

The previous dataset contained 10 controlled office-setting recordings where the participants performed handwashing mostly followed by returning to a seated position. This means that the motion patterns that were outside the hand washing segments had limited variation. The sensor modalities recorded in the previous dataset were the same as for our new recording, including a 3-axis accelerometer and gyroscope sampled at 52 Hz, as well as a humidity, temperature, and atmospheric pressure sensor sampled at 1 Hz.

While the previous publication relied on random forest classifiers (RF) and hand-crafted features, we used a deep neural network trained with focal loss and

[2] https://github.com/AliHurriat/HandwashingDetection-WearPuck.

SMOTETomek [1] resampling to handle class imbalance. We focused on the previously best-performing 5 s sliding window approach to segment the time-series data. The classification was performed using the same leave-one-participant-out method to ensure full comparability and the most realistic approximation of real-world performance on unseen users.

To assess the generalizability of the model and to validate the newly recorded data, we also evaluated the handwashing detection on a combined dataset of the new and the previously collected dataset. As opposed to the previous dataset, the newly collected dataset contains a wider range of real-world activities, so we expected the performance to differ.

4.2 Applying Semi-supervised Embeddings and Better Modality Fusion

To improve generalization across subjects and address the limitations of handcrafted features, as well as to tackle the problem of the still small dataset size, we employed a semi-supervised learning approach using a pretrained model to generate IMU embeddings. Specifically, we leveraged HARNet, a ResNet-based model from the BioBank SSL repository [24]. This model was pre-trained on large-scale (>700.000 person days) wearable sensor data in a self-supervised manner. For each 5-second window of accelerometer data, a 1024-dimensional embedding can be extracted from HARNet to capture rich motion representations.

In parallel, we included 11 statistical and temporal features from the humidity signal, including the measures mean, standard deviation, minimum, median, maximum, count of high values, range between 10%-, and 90%-percentile, peak count, mean and sd of first order derivative, and difference between last and first value of each window. We redesigned the humidity features compared to previous work, in order to make better use of the humidity response discussed in Sect. 3.2.

The humidity features were concatenated with the 1024-dimensional IMU embeddings to form a unified input vector for classification.

The concatenation of humidity features and the IMU embeddings was then jointly used to train a classification head. We trained a three-layer fully connected neural network (layers: 128->64->1 neuron(s)) with dropout, batch normalization, and Gaussian noise layers on this feature set. We applied a focal loss to handle class imbalance and further used SMOTE-Tomek resampling to improve minority class representation. Additionally, data augmentation was performed on samples belonging to the positive class using Gaussian noise injection and amplitude scaling. All features were standardized using z-score normalization. The predictions were smoothed using a median filter to reduce jitter. All evaluations in this publication followed a leave-one-participant-out (LOSO) protocol, allowing us to evaluate performance on completely unseen subjects, thus ensuring the best approximation of real-world performance. The results for each subject were then averaged to form the final F1 scores and accuracies.

As shown in Sect. 3.2, the class distribution is imbalanced and therefore high F1 scores are difficult to achieve, while extremely high accuracy values are easier to achieve but less meaningful.

5 Machine Learning Results

Although we performed multiple separate experiments on different subsets of the available data, the main results are jointly shown in Table 2. This table shows the mean results, averaged over all participants.

Table 2. F1 score and accuracy score result comparison of different methods (newly contributed methods in bold font). While previous work could not make use of the humidity features well, our CNN-GRU model was able to outperform the reported F1-score by 13% points. The accuracy was not reported for the Random Forest baseline. On the combined dataset, including humidity values boosts the performance for the HARNet-embedding based network (SSL Emb.). This model, based on the semi-supervised BioBank IMU embeddings, performs best on the larger and more complicated dataset (F1 = 0.68).

Dataset & Model	F1 Score		Accuracy	
	IMU+H	IMU	IMU+H	IMU
Pre-ex. Dataset (Baseline, RF, [6])	0.69	0.70	-	-
Pre-ex. Dataset (CNN-GRU)	**0.82**	0.74	**0.99**	0.98
Combined Dataset (CNN-GRU)	0.54	0.58	0.93	0.94
Combined Dataset (SSL Emb.)	**0.68**	0.60	**0.96**	0.94

We evaluated the CNN-GRU model's performance with and without humidity features on the pre-existing dataset using LOSO cross-validation. In the previous study, including humidity as a sensing modality did not have a positive impact on the classification performance (0.69 with humidity features vs. 0.70 without humidity features). However, as also shown in Table 2, when we applied our CNN-GRU model to the pre-existing dataset, the models that included humidity data achieved an average F1 score of 0.82, compared to 0.74 when humidity characteristics were excluded. Only a single percentage-point of accuracy could be gained when including humidity features (99% vs 98%). However, as explained above, accuracy is a less meaningful metric on highly imbalanced data. In general, adding humidity now significantly contributed to the higher performance of the model, in combination with the deep learning architecture outperforming the RF classifier.

Building on the analysis of the previous dataset, we next examined the model's performance on the combined dataset. As shown in Table 2, when applying the same CNN-GRU model, only relatively low F1-scores (0.54 with humidity features, 0.58 without humidity features) were reached, and including humidity values slightly worsened the performance. This led us to the conclusion, that the newly recorded data's additional background activities made it harder to train a general model, as certain background activities (such as playing the guitar) were only conducted by some subjects, and never in the pre-existing dataset.

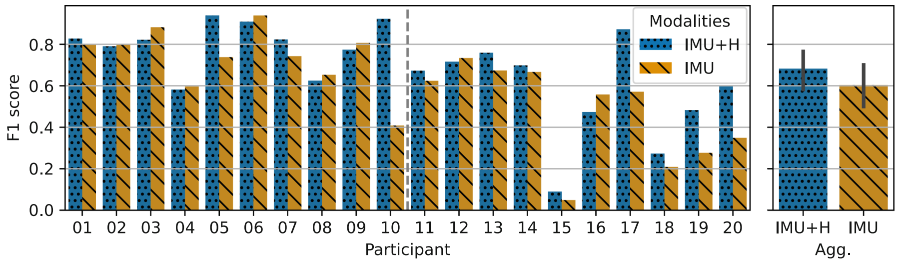

Fig. 3. Combined dataset results for the embedding-based model. Left: Per participant F1 scores of the LOSO-evaluation, with and without added humidity (IMU+H/IMU). Participants 1 to 10 belong to the pre-existing dataset, and 11 to 20 belong to the newly recorded dataset (separated by the dashed gray vertical line). Right: Aggregated F1 scores for both modality sets.

Our solution of applying the pre-trained BioBank "harnet" model, together with better manually engineered humidity features yielded much stronger performance on the combined dataset. When including humidity, an F1 score of 0.68 was reached (+0.14 compared to CNN-GRU), while without humidity features, the F1 score was lower with only 0.60 (+0.02 to CNN-GRU). The accuracy values of 96% vs 94% followed suit. Per participant results for this best model are shown in Fig. 3. Generally, the performance for participants belonging to the pre-existing dataset is stronger, with a mean F1 score of 0.8 for this subset (IMU+H). Notably, the results with and without the added humidity sensor are similar for most subjects. For some distinct participants, namely 05, 10, 17, 19, and 20, adding the humidity sensor boosts performance more significantly.

Another special participant is participant 15, for whom the system failed to detect hand washing reliably, with an F1 score below 0.1. We investigated this failure by visualizing hand washing procedures of participant 15 and comparing them to other participants. One such example is displayed in Fig. 4, where we compare the accelerometer data and humidity values for one entire hand washing procedure. We found the usual high frequency pattern (visible for participant 13 in subplot (b) of Fig. 4, 8–22 s) to be completely missing from participant 15's hand washing procedures, which explains the model's difficulties in detecting the washes. This finding highlights the uniqueness in hand washing patterns, which can differ strongly from person to person. Added to that, it shows that humidity changes alone are also not suitable for reliable detection, as humidity changes can occur for different reasons and during different activities, such as "washing dishes", which is included in the newly recorded dataset. Excluding participant 15 from only the evaluation step or from both the training and evaluation step

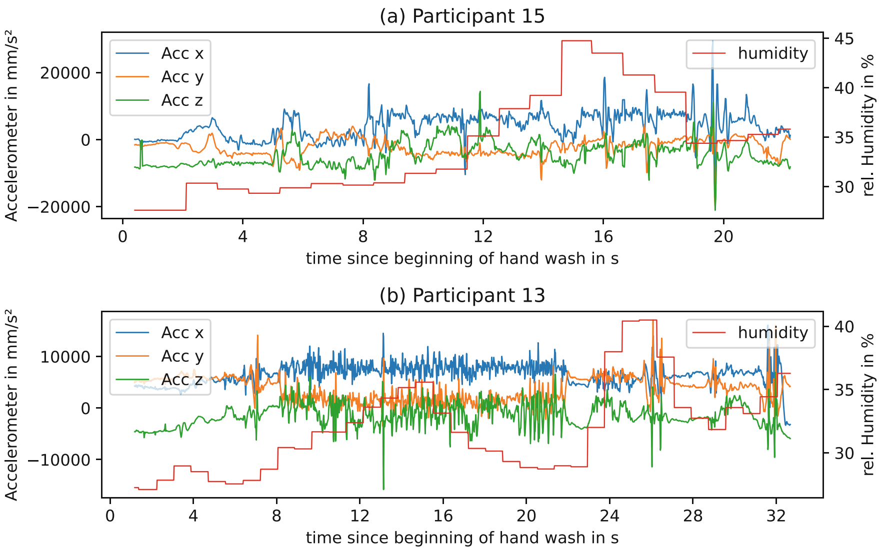

Fig. 4. Two hand washing procedures, with accelerometer and humidity values plotted. (a) For participant 15, (b) for participant 13. While the humidity rises for both subjects with a small delay, the IMU shows that, in comparison, participant 15 (a) moves more slowly, as the distinctive high-frequency hand washing pattern exists only for participant 13 (b).

increased the mean LOSO F1 score to 0.71 and 0.72, respectively, showing that this participant's peculiar hand washing patterns slightly confounded the model.

6 Discussion and Future Work

From the machine learning results on all collected data, we can derive that humidity is a valuable modality for hand washing detection. While previous work failed to reliably make use of the clearly visible humidity pattern around hand washes, our new results show that adding humidity to the IMU data provides context and thus improves the performance. This statement does not hold true, when the complexity of the activities and the small amount of available training data diminish the model's generalization capabilities, as was the case with our CNN-GRU model on the combined dataset. Humidity alone can also not solve the problem, as the results for participant 15 highlighted, for whom the movement patterns were too unique to be recognized well.

The variance of achieved F1 scores between different subjects is high, due to the highly personal hand washing styles. Especially participant 15 displayed a unique hand washing pattern. Although the model performed better when it could make use of humidity, the performance remained weak for this specific participant. In a previous study [6], we showed that personalized re-training with only a small amount of data can boost the performance further, even for participants for which a generalized model struggles. A "user calibration step", i.e., personalized retraining, therefore makes sense for the real-world deployment of such hand washing detection systems. This finding is fully in line with the literature, where real-world hand washing is also described as "unstructured" [13]. Even with the differences between subjects, the comparison of modalities showed that the model with humidity as an additional modality was never significantly outperformed. However, for some participants, adding humidity significantly improved the performance. Therefore, we conclude that its inclusion is justified, and we urge researchers to continue to include it in future data recordings.

As in many other domains in which labeled training data is scarce, utilizing a model that was pre-trained semi-supervisedly on a large unlabeled dataset yielded the best performance. We did not try finetuning the parameters of the pre-trained model, and only trained the classification head. Finetuning all layers could lead to even better performance in the future, as it has been shown to sometimes perform better than just training a fully-connected classification head [14,17]. With the significantly higher F1 scores achieved on both the previous dataset and the combined dataset, we came one step closer to a reliable, environment-independent, in-the-wild hand washing detection model, thanks to the inclusion of a pre-trained model in the pipeline and to adding more sophisticated, yet handcrafted, humidity features.

The newly recorded data is more diverse in its activities, and thus it is harder to predict the included hand washing instances. However, a large portion of the performance that was lost when compared to the simpler pre-existing dataset could be made up using the semi-supervised embedding model. The dataset is now twice as large, which, in theory, makes training models easier and enables better generalization capabilities. We freely offer the combined dataset for downloading in our GitHub repository[3].

The WearPuck recording device was also re-validated with this research, as we used it to record another 10 participants. It worked reliably and made it fairly easy to collect data from the participants, as all sensors are attached to one device and no further synchronization is needed.

In the future, even more complex models can be tested for offline classification. Our collected data can serve as a starting point for model training, likely of a pre-trained established model and architecture. For online classification, lightweight models, which can be run directly on wearable devices, could be trained with it.

An additional modality, which could probably be employed well to detect hand washing events would be sound. Microphone recordings could help, as the sound of tap water is usually audible, and tap water datasets already exist [11], just not in conjunction with hand washing [4]. A combined wearable system of IMU, humidity sensor, and microphone could be the solution to the most reliable system for the environment-agnostic detection of hand washing.

[3] https://github.com/AliHurriat/HandwashingDetection-WearPuck.

7 Conclusion

In this work, we extended our previous publication on hand washing detection using multi-modal sensing. We deepened our understanding of the usefulness of the inclusion of humidity sensors by analyzing it on a per-participant level, and, unlike related work, we were able to show that the humidity sensor boosts performance when added to IMU recordings for the hand washing detection task. To do so, we leveraged a model that was pre-trained on large amounts of unlabeled IMU data. We expect that humidity sensors can be useful in handwashing-related tasks and other water- or humidity-related HAR tasks as well, which remains underexplored.

We analyzed the weak outlier performance of one participant due to their unique hand washing style and propose to use small amounts of data for personalization in future work.

Additionally, we presented and evaluated the recordings of 10 new participants, increasing the size of the dataset to 20 participants and more than 20 h. The newer recordings contain significantly more unlabeled background activities, less idleness, and are fully compatible with the previous recordings. Our dataset can be used to train even more sophisticated models for the task of hand washing detection.

Possible future work includes combining even more wearable sensors into one multi-modal detection system, which in turn also requires additional data collection and method development.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Batista, G.E.A.P.A., Prati, R.C., Monard, M.C.: A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explor. Newsl. **6**(1), 20–29 (2004). https://doi.org/10.1145/1007730.1007735
2. Bharti, P., De, D., Chellappan, S., Das, S.K.: HuMAn: complex activity recognition with multi-modal multi-positional body sensing. IEEE Trans. Mob. Comput. **18**(4), 857–870 (2019). https://doi.org/10.1109/TMC.2018.2841905
3. Bock, M., Hölzemann, A., Moeller, M., Van Laerhoven, K.: Improving deep learning for HAR with shallow LSTMs. In: Proceedings of the 2021 ACM International Symposium on Wearable Computers, ISWC 2021, pp. 7–12. Association for Computing Machinery, New York (2021). https://doi.org/10.1145/3460421.3480419
4. Burchard, R., Laerhoven, K.V.: Enhancing Wearable Tap Water Audio Detection through Subclass Annotation in the HD-Epic Dataset (2025). https://doi.org/10.48550/arXiv.2505.20788
5. Burchard, R., Scholl, P.M., Lieb, R., Van Laerhoven, K., Wahl, K.: WashSpot: real-time spotting and detection of enacted compulsive hand washing with wearable devices. In: Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 483–487. ACM, Cambridge United Kingdom (2022). https://doi.org/10.1145/3544793.3563428

6. Burchard, R., Van Laerhoven, K.: Multi-modal atmospheric sensing to augment wearable IMU-based hand washing detection. In: Konak, O., Arnrich, B., Bieber, G., Kuijper, A., Fudickar, S. (eds.) Sensor-Based Activity Recognition and Artificial Intelligence, pp. 55–68. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-80856-2_4
7. Burton, M., Cobb, E., Donachie, P., Judah, G., Curtis, V., Schmidt, W.P.: The effect of handwashing with water or soap on bacterial contamination of hands. Int. J. Environ. Res. Public Health **8**(1), 97–104 (2011). https://doi.org/10.3390/ijerph8010097
8. De, D., Bharti, P., Das, S.K., Chellappan, S.: Multimodal wearable sensing for fine-grained activity recognition in healthcare. IEEE Internet Comput. **19**(5), 26–35 (2015). https://doi.org/10.1109/MIC.2015.72
9. Demrozi, F., Pravadelli, G., Bihorac, A., Rashidi, P.: Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey. IEEE Access **8**, 210816–210836 (2020). https://doi.org/10.1109/ACCESS.2020.3037715
10. Ezzeldin, M., Ghoneim, A., Abdelhamid, L., Atia, A.: Survey on multimodal complex human activity recognition. FCI-H Inform. Bull. **7**(1) (2025)
11. Gemmeke, J.F., et al.: Audio set: an ontology and human-labeled dataset for audio events. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 776–780 (2017). https://doi.org/10.1109/ICASSP.2017.7952261
12. Set up Handwashing on Apple Watch (2024). https://support.apple.com/en-my/guide/watch/apdc9b9f04a8/watchos
13. Lattanzi, E., Calisti, L., Freschi, V.: Unstructured handwashing recognition using smartwatch to reduce contact transmission of pathogens. IEEE Access **10**, 83111–83124 (2022). https://doi.org/10.1109/ACCESS.2022.3197279
14. Logacjov, A., Herland, S., Ustad, A., Bach, K.: SelfPAB: large-scale pre-training on accelerometer data for human activity recognition. Appl. Intell. **54**(6), 4545–4563 (2024). https://doi.org/10.1007/s10489-024-05322-3
15. Münzner, S., Schmidt, P., Reiss, A., Hanselmann, M., Stiefelhagen, R., Dürichen, R.: CNN-based sensor fusion techniques for multimodal human activity recognition. In: Proceedings of the 2017 ACM International Symposium on Wearable Computers, pp. 158–165. ACM, Maui Hawaii (2017). https://doi.org/10.1145/3123021.3123046
16. Ni, J., Tang, H., Haque, S.T., Yan, Y., Ngu, A.H.H.: A Survey on Multimodal Wearable Sensor-based Human Action Recognition (2024). https://doi.org/10.48550/arXiv.2404.15349
17. Nshimyimana, D., Rey, V.F., Suh, S., Zhou, B., Lukowicz, P.: PIM: Physics-Informed Multi-task Pre-training for Improving Inertial Sensor-Based Human Activity Recognition (2025). https://doi.org/10.48550/arXiv.2503.17978
18. Ordóñez, F., Roggen, D.: Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors **16**(1), 115 (2016). https://doi.org/10.3390/s16010115
19. Samyoun, S., Shubha, S.S., Sayeed Mondol, M.A., Stankovic, J.A.: iWash: a smartwatch handwashing quality assessment and reminder system with real-time feedback in the context of infectious disease. Smart Health **19**, 100171 (2021). https://doi.org/10.1016/j.smhl.2020.100171
20. Vepakomma, P., De, D., Das, S.K., Bhansali, S.: A-Wristocracy: deep learning on wrist-worn sensing for recognition of user complex activities. In: 2015 IEEE

12th International Conference on Wearable and Implantable Body Sensor Networks (BSN), pp. 1–6. IEEE, Cambridge, MA, USA (2015). https://doi.org/10.1109/BSN.2015.7299406

21. Wahl, K., Scholl, P.M., Miché, M., Wirth, S., Burchard, R., Lieb, R.: Real-time detection of obsessive-compulsive hand washing with wearables: research procedure, usefulness and discriminative performance. J. Obsessive-Compulsive Related Disord. **39**, 100845 (2023). https://doi.org/10.1016/j.jocrd.2023.100845
22. Wahl, K., et al.: On the automatic detection of enacted compulsive hand washing using commercially available wearable devices. Comput. Biol. Med. **143**, 105280 (2022). https://doi.org/10.1016/j.compbiomed.2022.105280
23. Wang, C., Sarsenbayeva, Z., Chen, X., Dingler, T., Goncalves, J., Kostakos, V.: Accurate measurement of handwash quality using sensor armbands: instrument validation study. JMIR Mhealth Uhealth **8**(3), e17001 (2020). https://doi.org/10.2196/17001
24. Yuan, H., et al.: Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. NPJ Digit. Med. **7**(1), 91 (2024). https://doi.org/10.1038/s41746-024-01062-3
25. Zeng, M., Yu, T., Wang, X., Nguyen, L.T., Mengshoel, O.J., Lane, I.: Semi-supervised convolutional neural networks for human activity recognition. In: 2017 IEEE International Conference on Big Data (Big Data), pp. 522–529 (2017). https://doi.org/10.1109/BigData.2017.8257967
26. Zhuang, H., Xu, L., Nishiyama, Y., Sezaki, K.: Detecting hand hygienic behaviors in-the-wild using a microphone and motion sensor on a smartwatch. In: Streitz, N.A., Konomi, S. (eds.) Distributed, Ambient and Pervasive Interactions. LNCS, pp. 470–483. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-34609-5_34