WristWash: Towards Automatic Handwashing Assessment Using a Wrist-worn Device

Hong Li*, Shishir Chawla*, Richard Li, Sumeet Jain, Gregory D. Abowd, Thad Starner, Cheng Zhang, Thomas Plötz
Georgia Institute of Technology
Atlanta, USA
hong.li@gatech.edu

ABSTRACT
Washing hands is one of the easiest yet most effective ways to prevent spreading illnesses and diseases. However, not adhering to thorough handwashing routines is a substantial problem worldwide. For example, in hospital operations lack of hygiene leads to healthcare associated infections. We present WristWash, a wrist-worn sensing platform that integrates an inertial measurement unit and a Hidden Markov Model-based analysis method that enables automated assessments of handwashing routines according to recommendations provided by the World Health Organization (WHO). We evaluated WristWash in a case study with 12 participants. WristWash is able to successfully recognize the 13 steps of the WHO handwashing procedure with an average accuracy of 92% with user-dependent models, and with 85% for user-independent modeling. We further explored the system's robustness by conducting another case study with six participants, this time in an unconstrained environment, to test variations in the handwashing routine and to show the potential for real-world deployments.

ACM Classification Keywords
H.5.m Information Interfaces and Presentation: I.5 Pattern Recognition

Author Keywords
handwashing; Gesture Recognition; Hidden Markov Model.

INTRODUCTION
Handwashing is a standard procedure performed multiple times a day for keeping hands clean and preventing the spread of germs and diseases. Keeping the hands clean is particularly critical in clinics and hospitals for preventing healthcare-associated infections (HAIs) [1, 17]. However, an estimated 720,000 patients suffered from HAIs in the United States alone in 2011; nearly 10% of those patients died from the infections. Alarmingly, clinic personnel have reported not having enough knowledge about proper handwashing procedure or failing to adhere strictly to it due to a heavy workload or limited hand hygiene product accessibility [18, 16].

According to guidelines published by the World Health Organization (WHO) [21], proper handwashing consists of 13 steps, which are shown in Figure 1. The procedure ensures that every area of the hands is properly covered. Adherence to handwashing routines is typically assessed through questionnaires, self-reports, or third party observations. Such manual assessments require substantial effort and thus have low compliance rates, are unreliable due to inevitable memory bias if not provided in time, or are simply impracticable for logistical reasons. As such, there is a desire for automated assessments.

Approaches based on placing or attaching devices around the sink are technically feasible but often not scalable. For comprehensive assessments they require hardware deployment at every sink. For certain smaller-scale, residential scenarios camera-based systems have been developed and deployed [7]. However, privacy concerns often prohibit the installation of cameras in bathrooms. Alternatively, indirect observations like correlating the consumption of washing products to handwashing frequency have been proposed [2]. However, such approaches do not capture the quality of the actual handwashing process and are thus not suitable for effective hygiene assessments (for example, in hospital scenarios).

In this paper, we present WristWash, a wrist-worn device and analysis method for capturing and analyzing handwashing. The system limits instrumentation to an inexpensive wrist-worn device, requiring minimal effort by the wearer and increasing the scenarios where the system could be used. The watch-like device contains an inertial measurement unit (IMU) and onboard storage and is battery powered for autonomous operation. Movement data are analyzed using a Hidden Markov Model based assessment routine that detects the 13 steps of proper handwashing as recommended by the WHO standard procedures. We evaluated WristWash through a case study with twelve participants in a lab setting. Our system achieved approximately 92% average recognition accuracy in a user-dependent scenario and 85% recognition accuracy on average in user-independent tasks for continuous recognition. Furthermore, we explored the feasibility of WristWash in an out-of-lab deployment demonstrating the effectiveness for real-world applications.

*equal contribution

ISWC'18, October 8–12, 2018, Singapore, Singapore

ISBN 978-1-4503-5967-2/18/10…$15.00
DOI: https://doi.org/10.1145/3267242.3267247

Figure 1: Standard handwashing routine according to the WHO [21].

BACKGROUND

Existing automated handwashing assessments can be categorized into two major types: *i)* camera-based; and *ii)* based on body-worn sensors – as surveyed below.

Computer Vision-based Approaches Mihailidis *et al.* developed a sensing agent for an intelligent environment that assists older adults with dementia in their daily activities, including recognizing handwashing [13]. Maekawa *et al.* presented a solution which employs a camera on a wrist mounted device to recognize activities of daily living (ADL) including handwashing. They focus on object use that is indicative of activities; for example, manipulating soap serves as a proxy for handwashing [12]. Hoey *et al.* designed a real-time vision-based system to assist a person with dementia with washing their hands. Their solution collected video from a camera mounted above a sink to track objects of interest (e.g., hands and towels) [7]. They combined a Bayesian sequential estimation and a decision-theoretic framework for tracking handwashing procedures. Llorca *et al.* developed a solution based on a camera mounted above the sink to be able to monitor handwashing procedures. Their approaches measured the user's hand motions via implementing a multi-class classification using an ensemble support vector machine [11].

One of the major concerns for vision-based handwashing recognition solution is the unavoidable privacy issues as these systems need to install cameras to monitor the target activities.

Pervasive Sensing-based Approaches As an alternative to camera-based approaches, other pervasive sensing methods have been employed for the assessment of handwashing routines. Mihailidis *et al.* proposed a prototype that used switches and motion sensors integrated into the environment to infer handwashing activities [14]. This system, however, cannot recognize individual washing steps as it is required by the WHO. Kinsella *et al.* developed an automated dispenser monitoring system to count handwash episodes in hospitals [9]. The system is based on detecting interaction with wall-mounted soap and alcohol gel dispensers. Unfortunately, this solution is not very robust, resulting in many false positive predictions and a rather coarse analysis level. Edmond *et al.* presented a handwashing detection system which utilized an alcohol sensor to detect the vaporization from using the sanitizer [3]. This system also operates at a rather coarse level, not allowing for actual washing assessments. Uddin *et al.* presented a wearable sensing framework which employed a 9-axis wristband to monitor and recognize human activities. Their scheme used handwashing as one of its examples, but it does not provide details of the handwashing procedure [20].

We present a wrist-worn device that enables automated handwashing recognition with minimal effort, instrumentation, and fewer privacy concerns. WristWash provides detailed offline analysis based on sensor data collected during handwashing. WristWash application scenarios comprise monitoring and teaching individuals proper, that is hygienic, handwashing routines. Compared to previous solutions, such as Harmony [16], our approach has the following advantages:

- WristWash facilitates handwashing detection with continuous recognition. Our model automatically determines start and end points for each handwashing step, which is more challenging as well as more informative and practical than mere gesture classification.
- We explore the feasibility of our automated analysis approach in a real-world home study, demonstrating the effectiveness for realistic application scenarios.
- We assess the recognition capabilities of three different models (user-dependent, user-adapted and user-independent). Our results indicate that the developed system is well-suited for handwashing recognition.

HARDWARE DESIGN

WristWash is built on a wrist-mounted device, which comprises an Adafruit Feather M0 Adalogger board [4], a SparkFun six degrees of freedom IMU (including a three axis accelerometer and a three axis gyroscope) [19]. We chose these sensors as their sizes make the overall device suitable for participants to wear around their wrist. Furthermore, the chosen hardware is energy efficient, which allows for continuous operation as may be required in hospital settings. The IMU is connected to the Feather M0 board via the I^2C communication protocol. WristWash stores sensor data on a 4GB microSD card. The device is powered by a 3.7v 500mAh lithium-ion polymer battery. WristWash records sensor data with a sampling rate of 200Hz. Figure 2 shows the device.

HANDWASHING RECOGNITION

The analysis of handwashing procedures is based on a preprocessing and recognition pipeline. Preprocessing normalizes

Figure 2: Wrishwash device and its integrated sensing capabilities.

sensor data and translates the input data stream into a sequence of feature vectors, which is then analyzed through the integrated segmentation and classification stage. Figure 3 gives an overview of the analysis pipeline, which will be described in detail below.

Preprocessing and Feature Extraction Sensor data is recorded at 200Hz but downsampled to 50Hz for efficiency reasons. All sensor readings are normalized to zero mean and unit variance. A sliding window procedure (window length: 0.06s with 70% overlap between subsequent frames) then extracts analysis frames for which features are extracted. Frame-wise feature extraction is performed over all axes. We explored different window lengths ranging from 0.04s to 0.2s and empirically chose 0.06s as the final window size as it led to the best recognition accuracy. Selecting features to adequately model short frames of accelerometer data is a challenging task. Typically, hand picked statistical attributes are chosen as feature representation, but it has been shown that such heuristically picked features alone do not always lead to satisfying recognition results. We use the empirical cumulative distribution function representation (ECDF; 3 coefficients) [6] along with mean, standard deviation, kurtosis and skew as the feature representation of the accelerometer data.

Recognition We employ Hidden Markov Models (HMMs) for segmentation and classification of the sequences of feature vectors extracted from the accelerometer data. An HMM is a sequential model, which is both efficient and effective for decoding temporal information, and as such is well suited for the analysis problem at hand [5].

We employ a linear left-to-right model topology for our handwashing HMMs. Each HMM has 15 states including start and end states. In this left-to-right linear configuration each state is connected to itself and the immediate successor within the graph. Model training is performed through Baum Welch optimization.

The handwashing models are fully continuous HMMs with three Gaussians per state for emission modeling. In addition to the handwashing HMMs we also employ a NULL class model. The NULL model is a Gaussian Mixture Model with three mixtures (effectively, an HMM with a single state). All models are evaluated competitively (The NULL model is used only in the out-of-lab study).

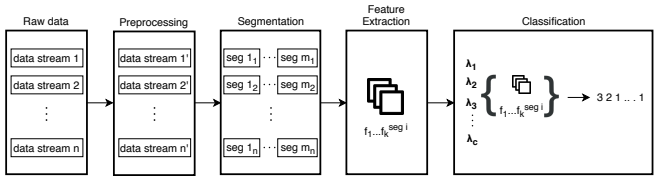


Figure 3: Activity recognition pipeline. Raw sensor data are normalized to zero mean and unit variance before sliding-window frame extraction. Feature extraction is performed for every frame. Resulting feature vectors are then fed into the HMM-based recognition backend.

For classification the sequence of feature vectors is aligned against all 13 step models separately, and the model that maximizes the log-likelihood for the input data determines the prediction. We use the Hidden Markov Model Toolkit (HTK Toolkit) [8] for the HMM training and classification tasks. Figure 3 summarizes the recognition process.

EXPERIMENTAL EVALUATION

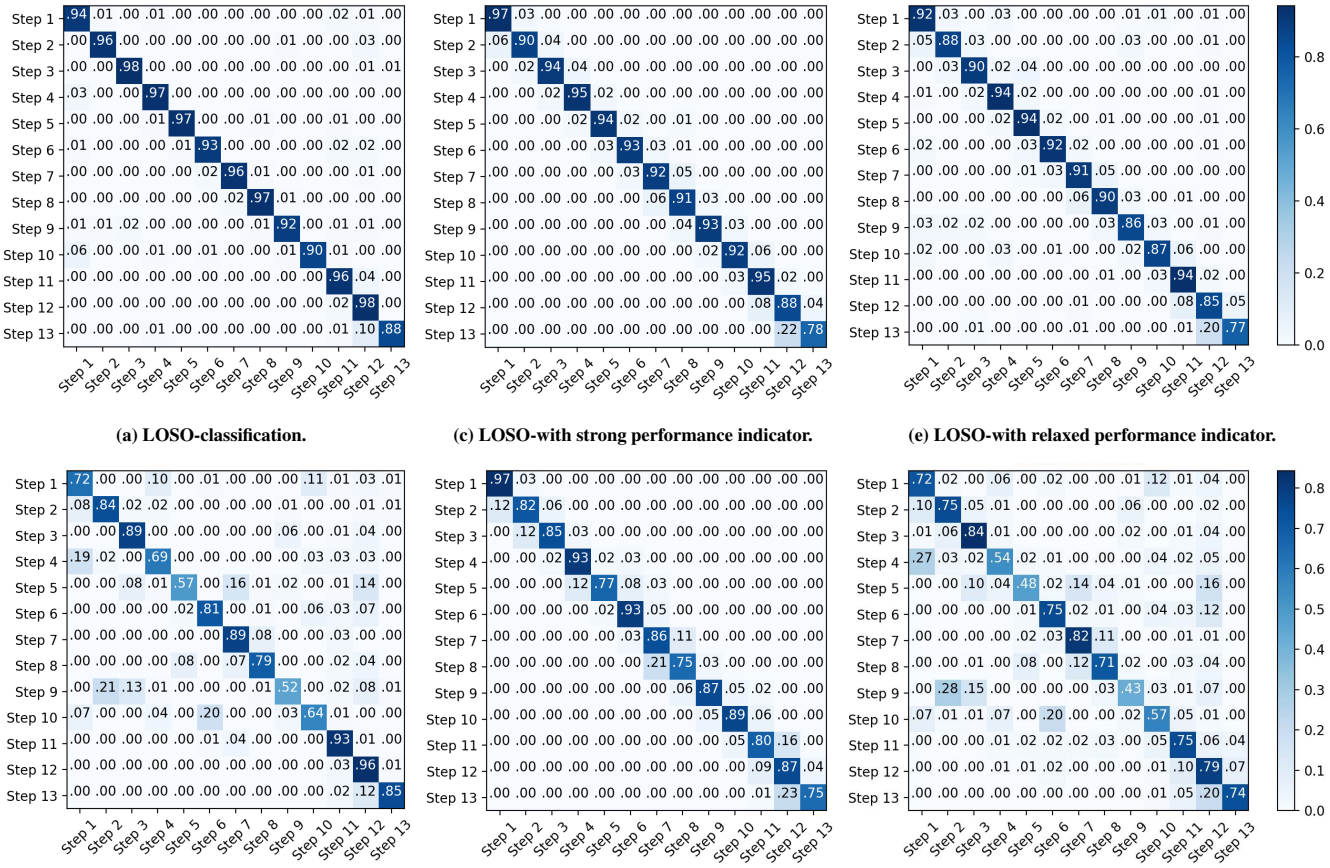
Data collection We collected data both in the lab and outside of the lab.

For the **lab study**, we recruited 12 participants (2 females, 10 males; age range: 23-28 years) to evaluate the developed system. All participants were right-handed and were asked to wear WristWash on their dominant hand. Participants were asked to watch an instructional WHO handwashing video so they could learn the procedure. After five practice sessions, we collected nine handwashing sessions from each participant. These sessions were videotaped for ground truth annotation. In the recorded sessions, all participants adhered to the WHO protocol except for participants 11 and 12, who accidentally missed some of the handwashing steps. The data for these participants was held out for validation (see below).

For the **out-of-lab study**, we collected data from six participants (2 females, 4 males; age range: 25-49 years). All participants were right handed and wore the device on the right hand. The participants were asked to first watch the instructional video and practice the procedure in order to learn the handwashing steps. For each participant, we collected nine sessions to train our models. Only one participant (participant 4) missed one step (step 4) in three of the nine lab sessions. After collecting nine training sessions, each participant was asked to wear the device for four hours with no restrictions regarding their activities to collect the out-of-lab data. During the four hour session, the participants were asked to wash their hands in a specified sink where we placed a camera so that we could obtain ground truth.

Lab Study

Recognition We divide the experimental evaluation into two parts: step classification and continuous recognition. The former aims at discriminating amongst the 13 manually presegmented handwashing steps of the WHO handwashing routine, and accuracy is defined as the percentage of steps correctly identified. The latter evaluates complete (or incomplete) handwashing procedures comprising up to the full 13 WHO steps. The continuous system must automatically segment the individual steps (i.e., accurately determine start and end times for every performed step) as well as classify these steps correctly. Accuracy is based on the percentage of data frames (0.06 sec each) that are correctly classified.



(a) LOSO-classification. (c) LOSO-with strong performance indicator. (e) LOSO-with relaxed performance indicator.

(b) LOPO-classification. (d) LOPO-with strong performance indicator. (f) LOPO-with relaxed performance indicator.

Figure 4: Confusion matrices for both classification and continuous recognition tasks. Plot (a) and (b) show the average step-based confusion matrices for classification task with LOSO and LOPO validation; Plot (c) and (d) are the average frame-based confusion matrices for continuous recognition using strong performance indicator with LOSO and LOPO models separately; Plot (e) and (f) are the average frame-based confusion matrices for continuous recognition using relaxed performance indicator with LOSO and LOPO models; The average recognition accuracies are always over 90% for LOSO validation in the three scenarios. The recognition accuracies for LOPO are 78%, 85% and 69% for classification, continuous recognition with strong performance indicator and continuous recognition with relaxed performance indicator, respectively.

We evaluate the performance with user-dependent and user-independent models. For the user-dependent model we perform leave-one-session-out (LOSO) validation. In LOSO we use eight sessions from a participant to train the HMMs and hold out one session to test the model. We do this for all nine sessions of a participant and compute the average recognition accuracy. We repeat the protocol for every participant and report the average accuracy across all participants.

For the user-independent model, we use leave-one-participant-out (LOPO) validation with data from nine participants for training, and test on the remaining participant. We do the same for every participant and report the average accuracy score.

Classification:

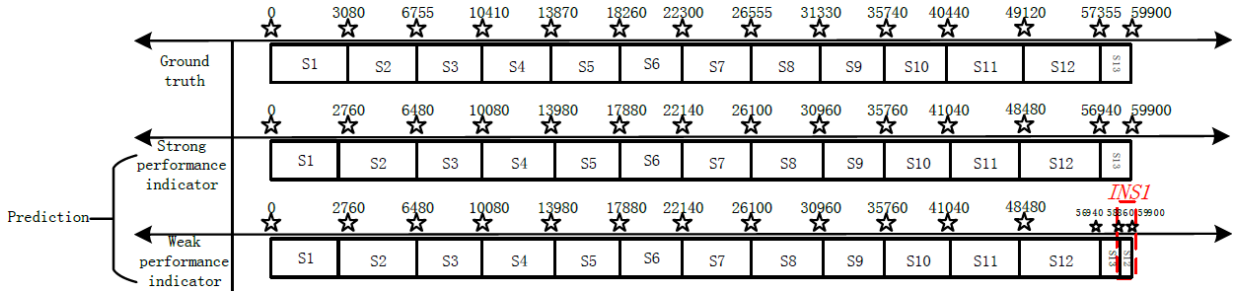
User-dependent model The LOSO accuracy score averaged across all participants is approximately 95%. Figure 4(a) shows the average confusion matrix for the LOSO classification. This result shows that the recognition problem is relatively easy when the model is designed for a specific user.

User-independent model The problem becomes more complex when the model is user independent, i.e., no data from the target user is used in training the model. User-independent modeling is a harder problem as the model is expected to learn with limited training data and has to generalize to unseen data.

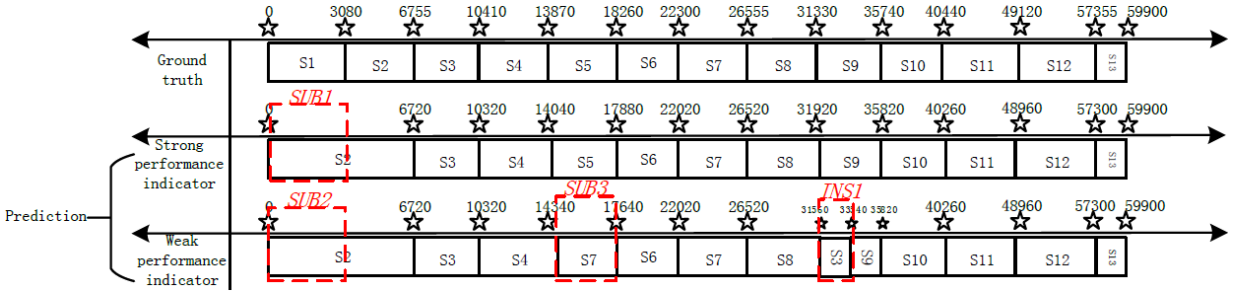
Average recognition accuracy for user-independent evaluation is 78%. As can be seen in the confusion matrix (Figure 4(b)), step 5 is mostly misclassified as step 7, because these two steps are very similar. In both steps the right hand is used to wash the left hand, and the movement of the wrist is very similar. A similar situation occurs with step 9 and step 2, where the right hand is cleaning the left hand, and again for steps 10 and 6. This result shows that the recognition is mostly accurate but similar steps are sometimes confused amongst each other.

Recognition results from user-dependent and user-independent scenarios demonstrate the model's capability in recognizing the different steps in the handwashing procedure. User-dependent models work very well but the more challenging problem is building effective user-independent models.

Continuous recognition: In our second set of experiments we evaluated WristWash in more dynamic scenarios with a focus on identifying skipped steps of the handwashing procedure, and on measuring durations of each performed handwashing step. In summary, this analysis resembles the quality assessment of handwashing as it is required for hospital routines



(a) LOSO (One insertion error is highlighted in red dash rectangle: INS_1.)



(b) LOPO (Three substitution errors and one insertion error are encircled in red dash rectangle.)

Figure 5: Continuous recognition results for LOSO and LOPO task for session 9 of participant 4. A sequence of ground truth handwashing steps are shown alongside outputs from strong performance indicator and relaxed performance indicator recognition. The stars indicate the boundaries of each gesture in ground truth and predictions. The numbers above the stars are millisecond time stamps for each gesture. The average recognition accuracy for strong performance indicator in LOSO is 92%. With relaxed performance indicator, the model achieves 90% recognition accuracy in LOSO on average. In LOPO, there are several types of errors: Substitution (SUB_1-SUB_3), Insertion (INS_1), Underfill and Overfill. The overall continuous recognition accuracy is 85% with strong performance indicator and 69% with relaxed performance indicator.

as recommended by the WHO. We evaluate the continuous recognition problem using two performance indicators:

Strong performance indicator: Recognizer's grammar assumes the participant follows the order of the 13 WHO handwashing steps (Figure 1), but may accidentally skip one or more steps.

Relaxed performance indicator: Recognizer's grammar assumes the participant may execute the steps in any order, any number of times. Note that in some tasks proper handwashing could allow different orders of some of the 13 steps.

We present performance for continuous recognition using user-dependent, user-independent, and user-adapted models for the two performance indicators described above. For the strong performance indicators, the model performs an enforced alignment of the data to the pre-defined sequence of handwashing steps. Misalignments and deletions are possible. For the more flexible task ("relaxed performance indicators") substitutions and insertions are also possible. Figure 4(c)-(f) show LOSO and LOPO results for 10 participants using both indicators.

User-dependent model As Figure 5(a) shows, our model predicts the handwashing steps with 92% frame accuracy when using the strong performance indicator and 90% frame accuracy when using the relaxed performance indicator. Errors are related to underfilling and overfilling [15], which are caused by incorrectly determining the exact start and end time of the particular handwashing steps. In the relaxed performance indicator case we observe insertion, underfilling and overfilling errors when classifying incorrectly, which is expected as the performance indicator does not impose any restrictions.

User-independent model As can be seen in Figure 5(b), there are now more errors in the recognition process. Even though we adopt the strong performance indicator, the prediction still completely misses step 1. The recognition results for the relaxed performance indicator are worse as more insertion and substitution errors are introduced. The recognition frame accuracy drops from 85% to 69%, a drop that was expected when changing to user-independent models.

User-adapted model To improve the accuracy of the user-independent model we explore a user-adapted model. In the user-adapted model we use the same protocol as the user-independent model but also add some (N) number of sessions from the target participant to train a personalized HMM. To evaluate this model we test the model on three sessions held out from the target user (sessions 7, 8 and 9). We increase N from 0 to 6 and report accuracies of the model as more number of target user sessions are added to the training set.

Figure 6 illustrates the results for continuous recognition using user-adapted models with the relaxed performance indicator. Each plot shows the accuracy changes for a participant with respect to the number of sessions added into the training set. For most participants, the accuracy scores reach over 80% on using 5 participant sessions, and the scores continue to improve on using 6 sessions. The accuracies improved by 17% absolute on average for these 10 participants. This result shows that using only a few sessions from a target user improves the accuracy of the user-independent system by a huge factor.

Overall the results show that it is feasible to build an effective handwashing solution with high recognition accuracies for a specific user (user-dependent models). However, the tasks

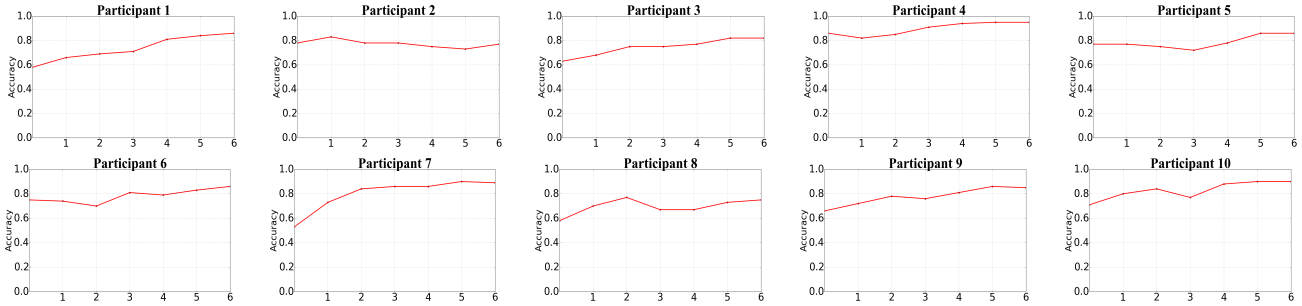


Figure 6: The continuous recognition results for user-adapted models. X-axis represents the number of sessions (*s*) added into the training data. Y-axis shows the accuracy scores with respect to the number of sessions added to the training data.

P	S	Missed steps	Strong performance indicator	
			ACC	Predicted steps
11	1	2 & 3	0.49	1, 4, 5, 6, 7, 8, 10, 11, 12, 13
	2	7 & 8	0.83	1, 2, 4, 5, 6, 9, 10, 11, 12, 13
	3		0.81	1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 13
	4	4	0.80	1, 2, 5, 6, 7, 9, 10, 11, 12, 13
	5		0.85	1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 13
	6		0.82	1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 13
	7		0.80	1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 13
	8	4	0.80	1, 2, 3, 5, 6, 7, 9, 10, 11, 12, 13
	9	9	0.86	1, 2, 4, 5, 6, 7, 8, 10, 11, 12, 13
12	1		0.79	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 13
	2	5 & 6	0.90	1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 13
	3	4	0.80	2, 5, 6, 7, 8, 9, 10, 11, 12, 13
	4	4	0.76	1, 2, 3, 5, 6, 7, 8, 9, 11, 13
	5		0.82	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
	6		0.65	1, 3, 4, 5, 6, 7, 9, 10, 11, 13
	7		0.72	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 13
	8	3	0.66	2, 4, 5, 6, 7, 9, 11, 13
	9		0.70	1, 3, 4, 5, 6, 7, 8, 9, 11, 12

Table 1: Continuous recognition results for participant 11 and 12 for the strong performance indicator case (see text for description). Column *P* and *S* represent Participant ID and Session ID; *Missed steps* indicate the handwashing steps missed per session; *ACC* is the frame recognition accuracy for each session; *Predicted steps* lists the sequence of recognized steps.

become harder when developing a user-independent model. Errors like deletions, substitution, underfill and overfill arise in both strong and relaxed performance indicator experiments. User-adapted models are an effective compromise.

Validation of generalization To further test the generalization capabilities of the system, we test the developed models on two more participants (participants 11 and 12). In our initial dataset of 10 participants, the participants did not miss any handwashing steps as outlined in the handwashing procedure. However, participants 11 and 12 missed some steps during their handwashing sessions (see Table 2 for more detail). With this experiment we aim to analyze the effectiveness of the system in detecting the missed steps, i.e., if the system does not make predictions for the missed steps, then we are successful in determining that the steps that were not recognized were not completed by the user.

For performance evaluation on these participants, we trained the model using all data collected from participants 1 to 10. All experimental results are reported for both the strong performance indicator and the relaxed performance indicator. To measure the quality of handwashing, we present the predicted duration versus the ground truth duration for each step, the sequence of handwashing procedures as well as the recognition accuracy in continuous handwashing gesture recognition.

Frame recognition accuracy for the relaxed performance indicator is around 58% on average per participant. This prediction result is similar to Figure 5(b). There are multiple errors in the predictions such as insertion, substitution, underfill and overfill. In light of the experiment performance for LOPO using relaxed performance indicator we argue that user-independent continuous recognition can be improved by adopting the strong performance indicator. Table 1 presents the continuous recognition results for participant 11 and 12 using the strong performance indicator. The average frame recognition accuracy improves from 58% to 77%. Although there are still errors happening in the prediction, the model can correctly predict the handwashing sequence with a high accuracy. Table 1 shows the predicted steps for each session.

As can be seen from the continuous recognition results for participant 11 and 12 (Table 1), the model never predicts a missing step, which shows that the system had no false positives for the missing handwashing steps. HMMs predicted the handwashing sequence correctly for participant 11 - session 1, and for participant 12 - session 2. For participant 11, there are a total of 110 handwashing steps in 9 sessions. Our model correctly predicted 96 steps and missed 14 steps (which means the step accuracy is 87%). For participant 12, the step accuracy is 82% after applying the same analysis.

Table 2 compares average step durations as predicted vs ground truth, illustrating the effectiveness and consistency of the automatically produced assessments. The values represent ratios of predicted length vs ground truth length. Values greater than 1 represent overfills, less than 1 are underfills, and equal to 1 represent exact predictions. For most steps the predictions are very close to ground truth (values close to 1), with a tendency to slightly overfill more often.

Out-of-lab Study

After evaluating the performance of WristWash in the lab study, we identified that the challenging part of the handwashing recognition task lies in building a more general user-independent model. To further investigate the performance of the system, we conducted an additional out-of-lab study with six participants. (Details of the data collection procedure have been described before.) In the out-of-lab study the data

Step	1	2	3	4	5	6	7	8	9	10	11	12	13
Participant 11	1.07	1.4	0.88	1.33	1.03	1.02	1.2	1.02	1.05	1.07	1.05	1.26	1.15
Participant 12	1.35	1.52	1.0	0.98	1.18	1.14	1.22	0.89	1.06	1.13	1.56	0.58	1.35

Table 2: Ratios of predicted duration versus ground truth duration for each handwashing step. Values greater than 1 represent overfilling, less than one represent underfilling cases, values that are exactly 1 represent perfect matches.

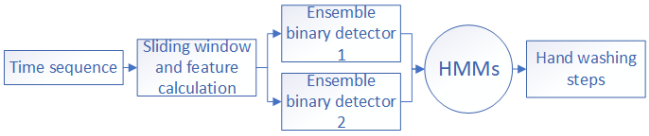


Figure 7: Modified recognition pipeline for the home study. Preprocessing and feature extraction are performed as before (Figure 3). Start and end steps for handwashing procedures are identified by two ensemble binary detectors before transferring to HMMs.

is collected in an unconstrained environment where the user is allowerd to do any activity they wish to perform.

In order to segment handwashing episodes from the four hour sessions we integrate an additional recognition component that focuses on automatic end-point detection by using two binary classifiers that detect the start and end steps of the WHO procedure (steps 1 & 13). Figure 7 illustrates the modified recognition pipeline, and Alg 1 describes the new procedure.

Algorithm 1 Ensemble binary detectors for step 1 and 13.

Input: Annotated dataset, $D_{train} + D_{home}$; Participant id i;
Output: identified time stamps for step 1 and 13;

```
Data preprocessing and feature calculation
Calculate descriptive statistics (e.g., mean handwashing duration T_i; mean durations α and β for step 1 and step 13 in training sessions) for participant i;
Train ensemble binary classifier (C_1) for step 1
Train ensemble binary classifier (C_2) for step 13
Employ trained model to predict D_home
Aggregate consecutive frames (threshold: α) and record time for first frame in candidate list Start.
Aggregate consecutive frames (threshold: β) and record time for last frame in candidate list End.
while j < len(Start) do                          ▷ Find all detected intervals
    while k < len(End) do
        if Start[j] + T_i > End[k] && Start[j] + T_i < End[k+1] then
            washDuration ∪ [Start[j], min(Start[j] + T_i, End[k+1])];
            j+=2; break;
        if Start[j] + T_i < End[k] && Start[j] + T_i < End[k+1] then
            washDuration ∪ [Start[j], min(Start[j] + T_i, End[k])];
            j+=2; break;
        j+=1;
    i+=1;
return washDuration
```

End-point detectors were trained using the data from the nine pre-collected sessions. We use ensemble-based end-point detectors that effectively cope with the imbalanced classification task (end-points are rare exceptions within the continuous sensor data streams) [10]. The ensembles comprise models that individually cover end-points and other data. We employ naive Bayes-based classifiers because they are straightforward to integrate into the overall architecture with modest requirements with respect to model training. Data pre-processing and feature extraction is identical to the initial procedure.

Participant ID	**1**	**2**	**3**	**4**	**5**	**6**
# Episodes (ground truth)	4	5	4	3	3	2
# Detected episodes	4	5	5	5	3	2
Mean accuracy Step 1	0.5s	1s	2s	2s	0.3s	0.1s
Mean accuracy Step 13	8s	4s	4s	10s	1s	5s
HMM accuracy	65%	67%	54%	48%	80%	74%

Table 3: The detected handwashing step 1 and 13 for each participant using user-dependent model.

User-dependent model For user-dependent modeling, we train the end-point detectors and the step HMMs using the nine sessions data from one participant P_i and apply the trained model to predict the four hours home study for P_i. Table 3 describes the number of handwashing episodes in the recorded dataset (row 1), the number of detected handwashing episodes (row 2), the mean accuracy for detecting steps 1 and 13 (rows 3 and 4), and the recognition accuracies of HMMs using the relaxed performance indicator (row 5). We use the mean time deviations (in seconds) to represent the accuracy for detecting steps 1 and 13. The time deviations are calculated basing on the difference between the start time from ground truth and the start time detected by the detectors. Lower time deviations indicate higher accuracies in identifying the end-points.

This procedure segmented all handwashing episodes with a low false positive rate. There was one erroneous detection for Participant 3, and two erroneous detections for participant 4. The detectors are very accurate in detecting step 1 (rub hands together) because the step is rather consistent. In contrast, detecting step 13 is more challenging: *i)* The duration of step 13 (turn off faucet) is very short. *ii)* The participants sometimes used their left hand to close the faucet; thus, the motion was not captured by the IMU sensor.

User-independent model: For user-independent modeling, we train the end-point detectors and step HMMs on data from the nine sessions from all other five participants (except participant *Pi*) and report predictions for the four hour session for *Pi*. Table 4 presents the results for the six participants using user-

Participant ID	1	2	3	4	5	6
# Episodes (ground truth)	4	5	4	3	3	2
# Detected episodes	4	5	7	5	4	3
Mean accuracy Step 1	0.8s	2s	0.25s	6s	5s	1s
Mean accuracy Step 13	9s	10s	6s	12s	15s	10s
HMM accuracy	40%	51%	38%	35%	36%	54%

Table 4: The detected handwashing step 1 and 13 for each participant using user-independent model.

independent models. Again, segmentation works well for step 1, whereas detecting step 13 is more challenging yet acceptable for practical applications.

DISCUSSION

WristWash is a handwashing analysis solution that can help assessing and thus, ultimately, implementing hygienic handwashing as recommended by the WHO. The results of our experimental evaluation have demonstrated the general effectiveness of WristWash. However, they have also unveiled some limitations that leave room for future developments.

User-dependent models are relatively straightforward to construct, which leads to excellent step classification performance of over 90%. A user-independent model would be more desirable but is much more challenging to achieve as classification accuracies drop to about 75%. Thus, generalization for new users *without retraining or adaptation* is limited. However, we have also demonstrated that moderate amounts of user specific training data are sufficient for effective model adaptation that substantially improve recognition accuracy over the user-independent case.

In our experiments, we first made the participants learn the standard steps following the instructions issued by the WHO [21]. We then asked the participants to practice for a few sessions before they join our user study. Therefore, they are rather well-trained with regards to standard handwashing steps. Thus, our system is primarily suitable for continuous assessment of adherence to the already learned routine – rather than assessing handwashing of those without training. Such a system may be practical – hospital staff routinely undergo similar training, and WristWash could be of value for maintaining hygienic handwashing standards.

In this paper, we present a system that can automatically detect handwashing procedures, which was evaluated in both an in-the-lab as well as an out-of-lab study. We discover opportunities and challenges towards a fully autonomous and self-sustained device in the future. We plan to explore more algorithms (e.g., RNNs and motif discovery) as well as new hardware for further improvement.

REFERENCES

1. B. Allegranzi and D. Pittet. 2009. Role of hand hygiene in healthcare-associated infection prevention. *Journal of Hospital Infection* 73, 4 (2009), 305–315.
2. W. Bischoff, T. Reynolds, C. Sessler, M. Edmond, and R. Wenzel. 2000. Handwashing compliance by health care workers: the impact of introducing an accessible, alcohol-based hand antiseptic. *Arch. of internal medicine* 160, 7 (2000), 1017–1021.
3. K. Edmond and A. Zaidi. 2010. New approaches to preventing, diagnosing, and treating neonatal sepsis. *PLoS medicine* 7, 3 (2010), e1000213.
4. Feather 2018. Adafruit Feather Mo Adalogger. `https://www.adafruit.com/product/2796`. (2018).
5. G. Fink. 2014. *Markov models for pattern recognition: from theory to applications*. Springer Science & Business Media.
6. N. Hammerla, R. Kirkham, P. Andras, and T. Ploetz. 2013. On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution. In *Proc. of the ACM ISWC*. 65–68.
7. J. Hoey, P. Poupart, von B., T. Craig, C. Boutilier, and A.x Mihailidis. 2010. Automated handwashing assistance for persons with dementia using video and a partially observable Markov decision process. *CVIU* 114, 5 (2010), 503–519.
8. HTK 2018. HTK Speech Recognition Toolkit. `http://htk.eng.cam.ac.uk`. (2018).
9. G. Kinsella, A. Thomas, and R. Taylor. 2007. Electronic surveillance of wall-mounted soap and alcohol gel dispensers in an intensive care unit. *Journal of Hospital Infection* 66, 1 (2007), 34–39.
10. Yang Liu, Aijun An, and Xiangji Huang. 2006. Boosting prediction accuracy on imbalanced datasets with SVM ensembles. In *PAKDD*. Springer, 107–118.
11. D. Llorca, I. Parra, M. Sotelo, and G. Lacey. 2011. A vision-based system for automatic hand washing quality assessment. *Machine Vision and Applications* 22, 2 (2011), 219–234.
12. T. Maekawa, Y. Yanagisawa, Y. Kishino, K.o Ishiguro, K. Kamei, Y. Sakurai, and T. Okadome. 2010. Object-based activity recognition with heterogeneous sensors on wrist. In *Percom*. Springer, 246–264.
13. A. Mihailidis, J. Boger, T. Craig, and J. Hoey. 2008. The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study. *BMC geriatrics* 8, 1 (2008), 28.
14. A. Mihailidis, G. Fernie, and W. Cleghorn. 2000. The development of a computerized cueing device to help people with dementia to be more independent. *Tech and Disability* 13, 1 (2000), 23–40.
15. D. Minnen, T. Westeyn, T. Starner, J. Ward, and P. Lukowicz. 2006. Performance metrics and evaluation issues for continuous activity recognition. *Performance Metrics for Intelligent Systems* 4 (2006).
16. M. Mondol and J. Stankovic. 2015. Harmony: A hand wash monitoring and reminder system using smart watches. In *Proc. of the 12th EAI*. ICST, 11–20.
17. D. Pittet, B. Allegranzi, H. Sax, S. Dharan, C. Pessoa-Silva, L. Donaldson, and J. Boyce. 2006. Evidence-based model for hand transmission during patient care and the role of improved practices. *The Lancet infectious diseases* 6, 10 (2006), 641–652.
18. D. Pittet, A. Simon, S. Hugonnet, Carmen L. Pessoa-Silva, V. Sauvan, and T. Perneger. 2004. Hand hygiene among physicians: performance, beliefs, and perceptions. *Annals of internal medicine* 141, 1 (2004), 1–8.
19. SparkFun 2018. SparkFun 6 Degrees of Freedom IMU. `https://www.sparkfun.com/products/10121`. (2018).
20. M. Uddin, A. Salem, I. Nam, and T. Nadeem. 2015. Wearable sensing framework for human activity monitoring. In *Proc. of the ACM WSA*. 21–26.
21. WHO 2018. WHO: How to handwash? With soap and water. `https://www.youtube.com/watch?v=3PmVJQUCm4E`. (2018). Accessed: 2018-2-5.