HAWAD: Hand Washing Detection using Wrist Wearable Inertial Sensors

Md Abu Sayeed Mondol, John A. Stankovic
Department of Computer Science
University of Virginia
Charlottesville, USA
{mondol, stankovic}@virginia.edu

***Abstract*—Hand hygiene is crucial in preventing the spread of infections and diseases. Lack of hand hygiene is one of the major reasons for healthcare associated infections (HAIs) in hospitals. Adherence to hand hygiene compliance by the workers in the food business is very important for preventing food-borne illness. In addition to healthcare settings and food businesses, hand washing is also vital for personal well-being. Despite the importance of hand hygiene, people often do not wash hands when necessary. Automatic detection of hand washing activity can facilitate just-in-time alerts when a person forgets to wash hands. Monitoring hand washing practices is also essential in ensuring accountability and providing personalized feedback, particularly in hospitals and food businesses. Inertial sensors available in smart wrist devices can capture hand movements, and so it is feasible to detect hand washing using these devices. However, it is challenging to detect hand washing using wrist wearable sensors since hand movements are associated with a wide range of activities. In this paper, we present HAWAD, a robust solution for hand washing detection using wrist wearable inertial sensors. We leverage the distribution of penultimate layer output of a neural network to detect hand washing from a wide range of activities. Our method reduces false positives by 77% and improves F1-score by 30% compared to the baseline method.**

***Index Terms*—Hand washing, inertial sensor, smartwatch, accelerometer, smartwatch**

I. Introduction

Hand hygiene is extremely important for personal well-being as well as in healthcare settings and food businesses. Lack of hand hygiene is one of the major reasons for health-care associated infections (HAIs) in hospitals [1], [2]. HAIs result in deaths of patients and high costs to the hospitals [3]. Adherence to hand hygiene compliance by workers in the food business is essential to avoid food contamination and thus the outbreaks of food-borne illness [4]. Hand hygiene is also crucial for personal well-being. It plays a vital role in avoiding sickness due to germs and preventing the spread of contagious diseases. Despite the importance of hand hygiene, people often forget to wash their hands when necessary. Automatic detection of hand washing activity can facilitate just-in-time alerts when a person forgets to wash hands. Monitoring hand washing practices is also essential in ensuring accountability and providing personalized feedback, particularly in hospitals and food businesses. In this paper, we present HAWAD, a robust solution for hand washing detection using wrist wearable inertial sensors. Wrist wearable devices like smartwatches and fitness trackers are used widely. In contrast to in-situ sensors or cameras, wearable devices are ubiquitous and not limited in a specific location or context. In addition to hand washing detection, a wrist device can be used for context recognition and providing alerts to the user when necessary.

Inertial sensors available in smart wrist devices can capture hand movements, and so it is feasible to detect hand washing using these devices. However, it is challenging to detect hand washing using wrist wearable sensors since hand movements are associated with a wide range of activities. We refer any activity other than hand washing a NULL activity. It is nearly impossible to enumerate, let alone collect data for, all possible human activities. State-of-the-art solutions for activity recognition are mostly data-driven, and so the performance of the solutions largely depends on the data used to develop the models. The activity recognition models are usually developed and evaluated using data from a limited number of NULL activities, often in lab settings. Consequently, the models might perform poorly in the free-living context where many other NULL activities could be present. We illustrate the problem in Figure 1 using an example where data are available for hand washing (H) and three other NULL activities (A, B and C). There are no data available for activities D and E that are also NULL activities. If we train a classification model using data for hand washing (H) and available NULL activities (A, B, C), the decision boundary of the model would be based on these data, and consequently, activity E might be detected with very high probability as hand washing. Such solutions perform poorly in real-world context where users can perform many activities for which data are not available to train the classification model.

Most of the state of the art solutions for activity recognition, including those for hand washing detection [5]–[7], do not address the problem of NULL activities. These solutions mainly focus on feature engineering, parameter tuning, and/or the classification methods. Recently, neural networks are being used widely for activity recognition [8]–[11] due to their effectiveness over classical machine learning techniques like Random Forest and Support Vector Machine. These solutions also suffer from the problem of NULL activities. In this paper, we present a novel solution for addressing the problem of NULL activities and have evaluated the solution for hand washing detection. We mitigate the problem of NULL activities by leveraging the distribution of the penultimate layer of a neural

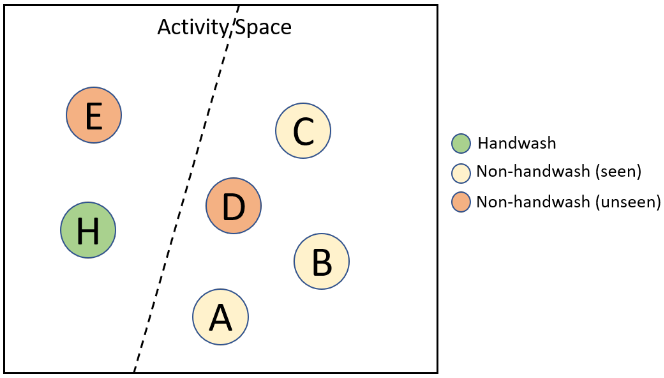

Fig. 1. Activity space

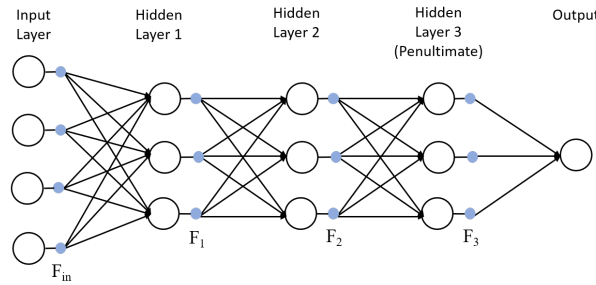

Fig. 2. Example of a Feedforward Neural Network. Instead of connecting a node directly with the nodes in the following layer, we put a connector in between to better show the input and embedded features.

network. Our solution detects out-of-distribution samples that mostly come from unseen activities. We collected a dataset that contains data from hand washing as well as several other activities. We trained a neural network model using our dataset and then tested the robustness and effectiveness of our solution using WISDM [12], a publicly available dataset that does not contain any data for hand washing. This dataset has data for 18 different activities, many of which are not present in our dataset. Our method reduces the false positives from the WISDM dataset by 77% and improves F1-score by 30% than the baseline method [6].

The major contribution of this paper are:

- We present a novel solution for hand washing detection that addresses the problem of NULL activities.
- We developed a dataset for hand washing detection using wrist wearable sensor and evaluated our solution using the dataset along with a public dataset.
- Our method reduces the false positives by about 77% and improves overall F1-score by 30% compared to the baseline method.

II. Method

As mentioned earlier, neural network based methods are usually more effective than classical machine learning techniques, like Random Forest and Support Vector Machine, in recognizing human activities. We also use a neural network for hand washing detection but additionally leverage the distributions of the penultimate layer outputs of the network to detect NULL activities. Each layer of a neural network transforms its input features to another feature space. The outputs at the penultimate layer of the network represent the final features that are usually classified by a Sigmoid or Softmax function. Figure 2 shows an example of a feedforward neural network with three hidden layers and one output node. The input features (F_i) are sequentially transformed to final features (F_3) where instances from the same classes come closer and from different classes moves further compared to the features from earlier layers. Figure 3 shows some instances from hand washing and other activities at different layers of the feedforward neural network. There are 64 nodes in each of the hidden layers of the network, and we have used the t-SNE [13] method to embed the outputs of each layer into 2-dimension for visualization. As illustrated in the figure, the instances are better separable over the layers. It should be noted that the decision boundary of the network is computed using the output of the penultimate layer (Figure 2). We have drawn a hypothetical decision boundary in Figure 3(d) just for illustration purposes. The instances from hand washing are clustered in a limited area, but any instances left to the boundary will be classified as hand washing. Let assume that instances in the area X and Y are from unseen NULL activities, i.e., from activities for which there was no data available during training the classification model. Though the instances from Y are correctly classified as negative by the model, instances from X are falsely classified as hand washing. Here, the instances of X are out of the distribution of the hand washing instances. We leverage the distribution of the penultimate layer to detect unseen NULL activities even though the instances are classified as hand washing by the neural network.

We use class conditional Gaussian distribution of the penultimate features to detect the out-of-distribution (OOD) instances. Similar to the method proposed by Kimin et al. [14], our method is not confined to networks with a specific architecture. We use a pre-trained network and get the penultimate features for a set of hand washing instances, referred to as Representative Set. For a set of representative instances $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, the mean and covariance are computed as:

$$\hat{\mu} = \frac{1}{N} \sum_i \mathbf{x}_i \tag{1}$$

$$\hat{\Sigma} = \frac{1}{N} \sum_i (\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^T \tag{2}$$

Here, $\mathbf{x}_i$ represents the penultimate feature i.e. the output of the penultimate layer for the i^{th} instance, and so $\mathbf{x}_i$ is a N dimensional vector where N is the number of nodes in the penultimate layer. It should be noted that all the instances of the representative set are from hand washing. We compute the distance of an instance from the mean ($\hat{\mu}$) using Mahanabolis distance [15] as:

$$d(\mathbf{x}) = (\mathbf{x} - \hat{\mu})^T \ \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu}) \tag{3}$$

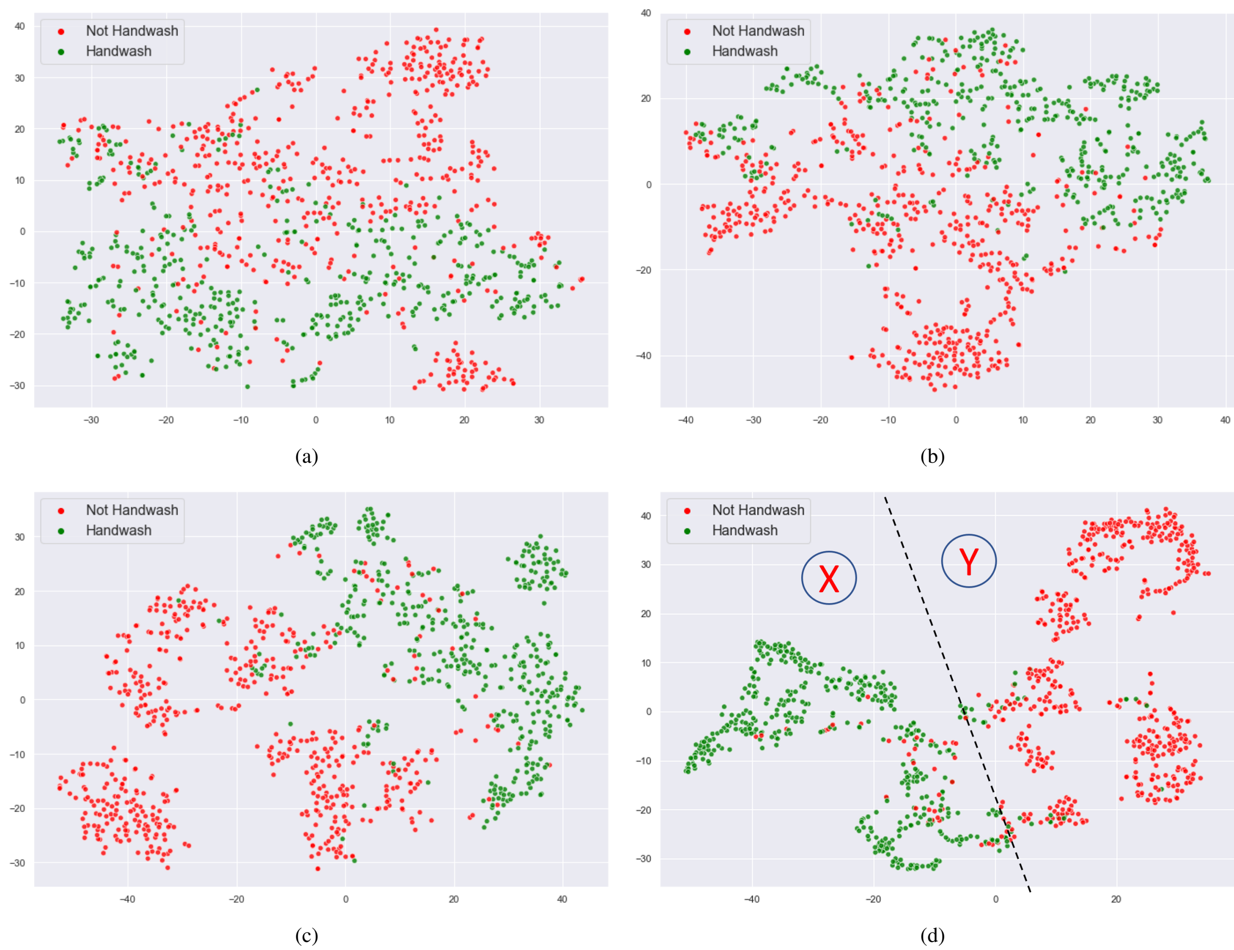

Fig. 3. t-SNE representation of some hand washing and other instances for (a) Input features, (b) output of layer 1, (c) output of layer 2 and (d) output of layer 3 of a feedforward neural network with 3 hidden layers. This figure is better visualized in color.

The Mahalanobis distance is usually more effective than Euclidean distance for detecting OOD samples [14]. Using the representative instances, we calculate a distance threshold (d_{th}) that covers most of the hand-washing instances. The distance threshold creates a boundary around the center of the positive (hand washing) instances, and any instance outside that boundary (OOD sample) is considered as a NULL activity (negative class). Consequently, activities from X (Figure 3(d)) is correctly detected as negative.

Our method is applied to the output of the penultimate layer, and so we use a pre-trained network. Training the neural network is not part of our solution. We need to estimate three parameters: $\hat{\mu}$, $\hat{\Sigma}$ and d_{th} that are used to infer the class of test instances. Figure 4 shows the steps for parameter estimation and inference. We describe them in more detail below.

A. Parameter Estimation

To estimate the parameters, we use a set of instances from hand washing, called Parameter Estimation Instances. It can be the hand washing instances from the dataset used to train and evaluate the network or any set of similar instances. We get the penultimate features of the instances and their probability of being hand washing using the pre-trained network. The penultimate features of the instances that are detected as positive (hand washing) by the network are used to construct the representative set. Our goal is to further test only the instances that are detected as positive by the network. So, we do not include the instances detected as negative into the representative set. We use Equation 1 and 2 to estimate the mean and covariance of the selected penultimate features. We calculate the Mahanabolis distance of each of the penultimate features using Equation 3 and then select some percentile (P) of the distances as the distance threshold. With a P percentile distance threshold, $(100 - P)\%$ of the instances truly detected as positive by the network are discarded as negative or OOD instances by our method. However, it would discard many negative instances that are falsely detected as positive (False Positives) by the network.

B. Inference

To infer the class of a test instance, we find its probability using the pre-trained model. If the instance is detected as hand washing, we use the penultimate features of the instance to find its distance. If the distance is greater than the threshold,

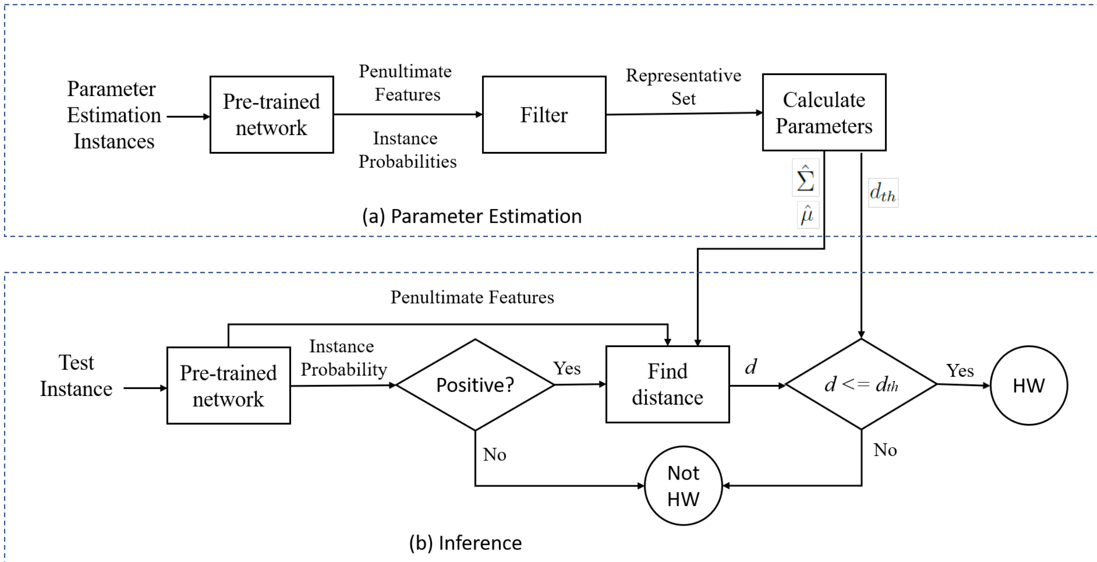

Fig. 4. Steps in (a) Parameter Estimation (b) Inference

the instance is detected as a negative or an *OOD* sample; otherwise, it is considered a positive instance. We do not process any instance that is detected as negative by the network. Our solution discards many false positives from the network with a small compromise on true positives, resulting in significant improvement in different performance metrics.

III. Experiments

A. Data Description

We have developed a dataset, called HAWAD dataset, by collecting data from 16 participants (9 males, 7 females) with age range between 17 to 36 years. The participants washed hands following the guideline by the World Health Organization (WHO) [16], as shown in Figure 5, as well as by rubbing hands in different other ways usually present in hand washing. It is extremely important for health and food workers to follow the guideline by WHO. Though others usually do not need to follow the guideline, several gestures from the guideline are often present in their hand washing. We also collected data for other activities including wiping water from hands with a towel or napkin, walking, opening/closing doors, using computers/phones, eating, and drinking. The data were collected using Samsung Gear Live, an Android-powered smartwatch. The dataset contains about 5 hours of data from each of the hands where nearly half of the data are from hand washing. We have collected acceleration, rotation rate, linear acceleration, and gravitational acceleration from the smartwatches. More details of the data and a preliminary study using a decision tree method are described in Harmony [5].

We have also used a public dataset, named WISDM [12], that contains accelerometer and gyroscope data for 18 activities from 51 subjects. The activities are walking, jogging, climbing stairs, sitting, standing, typing, brushing teeth, kicking a soccer ball, playing catch with a tennis ball, dribbling a basketball, writing, clapping, folding clothes, drinking from a cup and eating soup, chips, pasta, and a sandwich. This dataset does not have any data for hand washing, and so we use this dataset to detect out-of-distribution patterns. In the dataset, there are data available from both a smartphone and a smartwatch. We have used data from the smartwatch only.

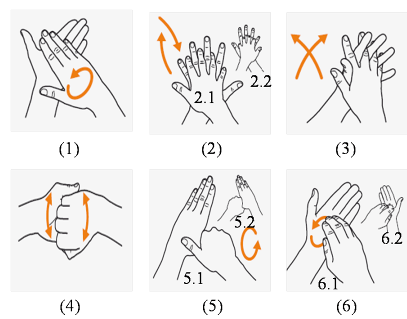

Fig. 5. Hand washing Guideline by World Health Organization (WHO). Image from Harmony [5].

B. Network Training

The WISDM dataset has both accelerometer and gyroscope data from the smartwatch. However, the data from these two sensors are not synchronized in the dataset. Also, using a gyroscope in addition to an accelerometer doesn't improve the performance significantly for hand washing detection, but consumes a significant amount of battery life from the watch [5]. So, we use data only from the accelerometer of the watch in our experiment. The accelerometer data from the watch are time series in nature. We segment the data into 1 second long frames with 0.5-second window sliding, and extract a set of features including mean, variance, root mean square, median, first quartile, third quartile, minimum, maximum, skewness, kurtosis from each of the axes of the sensors. We also use

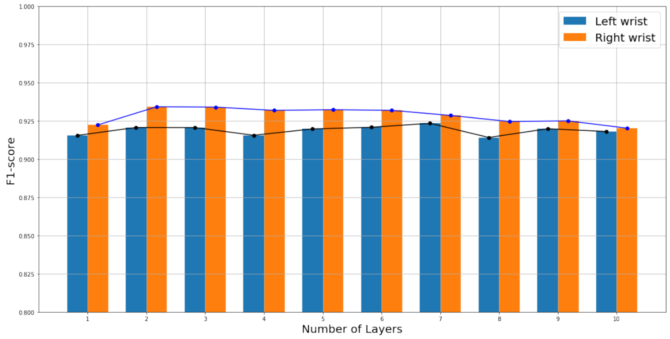

Fig. 6. F1-score for networks with different number of layers.

the covariance among the axes resulting in a total of 33 features. There is no pre-trained neural network available for hand washing detection from accelerometer data. So, we used our HAWAD dataset to train a feedforward neural network. There are 64 nodes in each of the layers of the network and we evaluated the models for different numbers of layers. We split the data into training (80%) and testing (20%) sets with random sampling. A dropout rate of 0.25 and a validation set (10% of the training data) along with an early stopping mechanism are used to reduce the problem of over-fitting. We developed and evaluated models for the left and the right hand separately. Figure 6 shows the F1-scores on the test data for a different numbers of layers. The performance doesn't differ significantly, but it reduces as the number of layer increases, particularly for the right hand. This is because the network is over-fitted as more layers are added. We have used the network with three hidden layers for the remaining experiments.

C. Out of Distribution

We used the pre-trained model to predict an instance. Since there is no hand washing data in the WISDM dataset, any instance from this dataset that is detected as hand washing is a false positive. About 6% and 5% of the instances from the WISDM dataset are detected as false positives by the neural network for the left and the right wrist, respectively. Figure 7 shows the output of each of the layers of the network for the hand washing and NULL instances from the validation dataset of HAWAD as well as some false positives from the WISDM dataset. Similar to Figure 3, instances from hand washing are better separable over the layers. In the penultimate layer (Figure 7(d)), many of the false positives are out of the distribution of the hand washing instances. However, they are detected falsely as hand washing due to their closeness to the hand washing instances than to the NULL instances. Our method detects such false positives using the distance threshold.

We used the hand-washing instances from our dataset to estimate the parameters ($\hat{\mu}$, $\hat{\Sigma}$, d_{th}). As mentioned earlier, a percentile is used to determine the distance threshold. Any instance detected as positive by the network is detected as negative by our method if the distance of the instance is greater than the threshold. The more we reduce the percentile the more instances are detected as negative by our solution and vice versa. Though a number of true positive instances may be detected as negative by our solution, a large number of false positives are correctly detected as negative, ultimately improving the overall performance. We define TPDNR (rate of the True Positives Detected as Negative) and FPDNR (rate of the False Positives Detected as negative) as:

$$TPDNR = \frac{True\ Positives\ Detected\ as\ Negative}{Total\ True\ Positives} \tag{4}$$

$$FPDNR = \frac{False\ Positives\ Detected\ as\ Negative}{Total\ False\ Positives} \tag{5}$$

Figure 8 shows the TPDNR and FPDNR for the left and the right wrist, respectively. The WISDM dataset does not have any data for hand washing, and so there is no true positive instances from this dataset. The results show that our solution corrects a significant portion of the false positives from both the HAWAD and the WISDM dataset. The less the percentile (and so the distance threshold), the more false positives are corrected. However, a small portion of the true positives is also detected as negative. For example, with 95 percentile threshold, we detect 5% of the true positives as negative but reduce the false positive rate from the WISDM dataset by 48% and 45% for the left and right wrist, respectively. It also reduces the false positives for our dataset by 27% and 25%, respectively. The results show that our method is very effective in detecting instances from unseen data and activities. The more we reduce the percentile, the more false positives are corrected, but it also results in more mistakes for the true positives. The percentile should be set according to the requirements of the applications on some metrics like precision, recall or F1-score. At the 100 percentile threshold, we select the distance using the maximum distance of all the true positives. There are some true positives that lie far away (outliers) from the mean of the Gaussian distribution. Consequently, the distance at 100 percentile is very high, and very few (nearly zero) false positives are discarded, even from the WISDM dataset. Setting the percentile to 99 corrects about 22% false positives from the WISDM dataset. This is because the distance threshold is reduced significantly compared to 100 percentile due to the removal of the outliers.

Figure 9 shows the precision, recall and F1-score for the left and right wrist, respectively. As expected, when the percentile is increased the precision decreases and the recall increases, and vice versa. This is because when percentile is increased, there are more true positives that increase the recall, but the number of false positives also increases, that results in reduction of the precision. The F1-score is the harmonic mean of precision and recall and widely used to balance between them. We see that our solution gives the best F1-score around 80 percentile threshold. The F1-scores at this percentile are 0.72 and 0.74 for the left and the right wrist, respectively. The F1-scores of the baseline method, the pre-trained network without using out-of-distribution as proposed by Galluzzi et

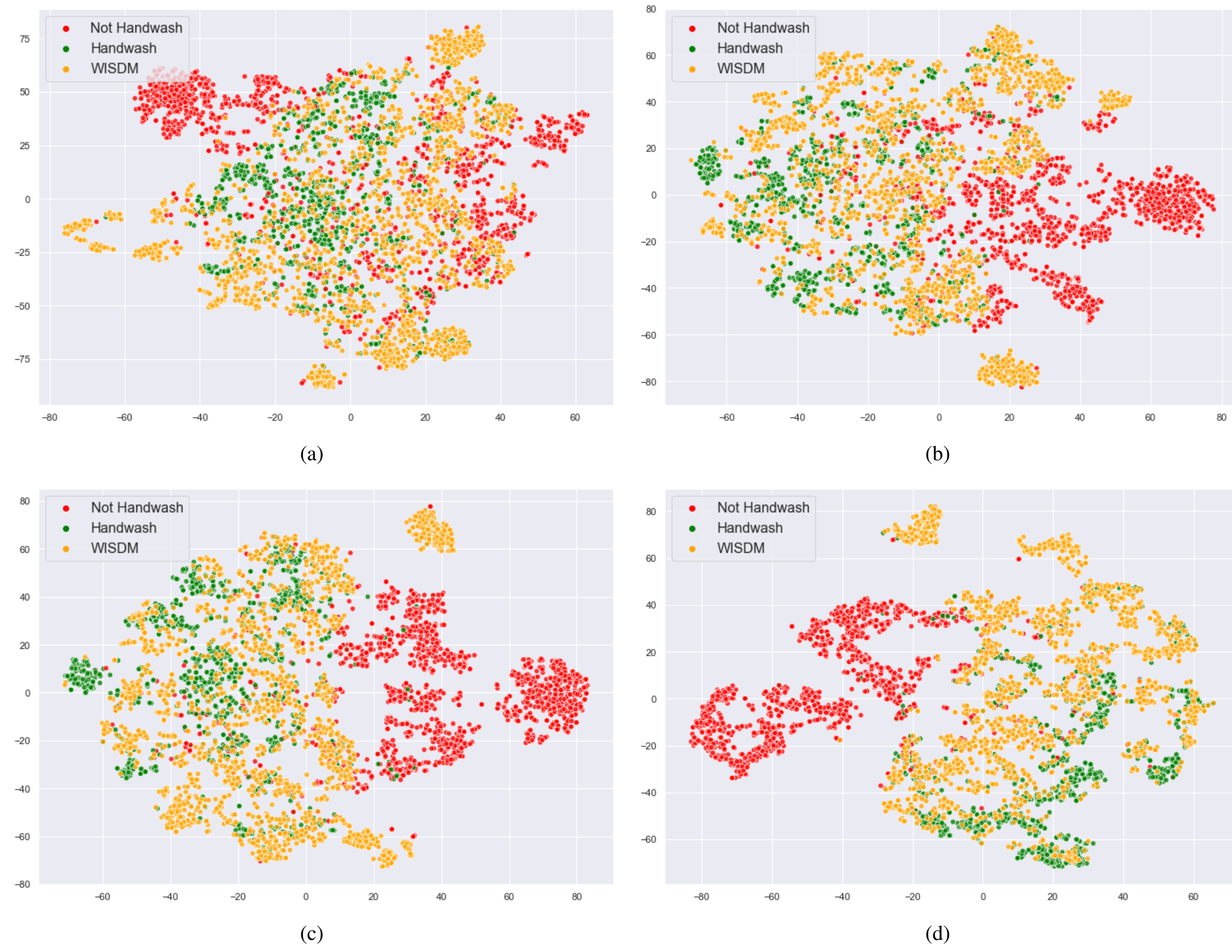

Fig. 7. t-SNE representation of the validation instances as well as some False positives from the WISDM dataset for (a) Input features, (b) output of layer 1, (c) output of layer 2 and (d) output of layer 3 of the network. This figure is better visualized in color.

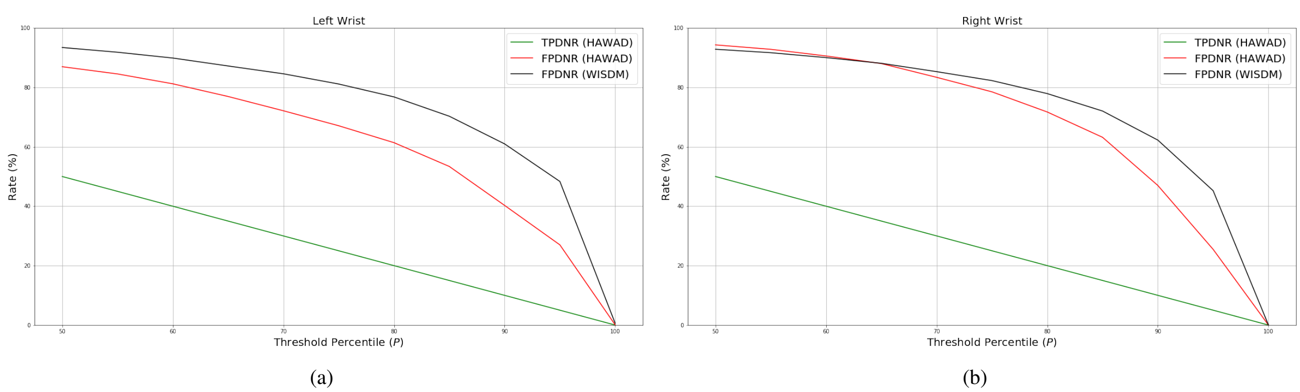

Fig. 8. TPDNR (HAWAD), FPDNR (HAWAD), FPDNR(WISDM) for different percentiles for (a) left wrist, (b) right wrist.

al. [6], are 0.55 and 0.57, respectively. So the F1-score is improved by 0.17, about 30% more compared to the baseline. At this percentile, the false positive rate from the WISDM dataset is reduced by 76.8% and 77% for the left and right wrist, respectively, and the reductions for our dataset are 61.4% and 71.7%, respectively. The result indicates we avoid a significant amount of false positives by using our out-of-distribution based method on top of existing neural networks. With a small compromise on recall, our method gains large precision, resulting a significant increase in the F1-score.

IV. Related Work

Human activity recognition using wearable sensors is an active research area with significant involvement of researchers from different domains, including computational science, healthcare, and engineering. Recent developments in wearable technology, particularly availability and widespread use of sensor-enabled tiny devices like fitness trackers and smartwatches, have thrust research in this direction. Most of the state-of-the-art solutions for activity recognition use neural networks. Hammerla et al. [11] explore different neural network architectures, including convolutional and recurrent neural networks for activity recognition using wearable sensors. They propose a regularization technique to improve the performance of the networks. DeepConvLSTM [9] segments the time series sensor data from the wearables and applies both convolution and recurrent neural networks on each of the segments independently. Guan et al. [8] ensembles a set of Long Short Term Memory (LSTM) networks to improve the performance of activity recognition tasks. They save the LSTM model after each of the epochs where the data used to train the network during an epoch is randomly selected from the training data and ensemble the top-performing LSTM learners for activity recognition. These works do not address the issue of NULL activities, particularly unseen NULL activities.

Galuzzi et al. [6] use sensors from wrist devices to detect hand washing. In addition to hand washing, they collect a limited number of NULL activities that include opening a jar, opening and eating the candy from the jar, tying shoes and applying bandages. They evaluated the performance of different machine learning techniques, namely K-Nearest Neighbors, Decision Tree, Neural Network, and Naive Bayes, for hand washing detection where the neural network outperforms the other methods. WristWash [7] uses Hidden Markov Model to detect different hand rubbing approaches suggested by the World Health Organization (WHO). These works do not address the problem of separating hand-washing activities from unseen NULL activities.

Detecting out-of-distribution is an active area of research. Hendrycks et al. [17] uses probabilities from softmax distributions to detect out-of-distribution samples. They evaluated their method using datasets from computer vision, natural language processing, and speech recognition. Lee et al. [14] propose a method that can be used with any pre-trained softmax neural classifier to detect abnormal samples. They use class conditional Gaussian distributions of the outputs of different layers of a neural network along with Mahalanobis distance to find confidence. The method works for both out-of-distribution and adversarial samples. They also demonstrated the use of their method in learning new classes. They use pre-trained convolutional neural networks on some vision datasets including CIFAR [18] and ImageNet [19]. These works on out-of-distribution detection are focused on computer vision, natural language processing, or speech recognition. We developed a solution for hand washing detection using wearable sensors.

V. Discussion

State of the art solutions for activity recognition focuses on the architecture or parameter tuning of the neural networks. We also use a neural network for hand washing detection. However, it is not focused on the neural network architecture or parameter tuning; rather, it works on top of a pre-trained neural network. So, our solution can be used with the existing neural network based solutions to detect hand-washing with more robustness and accuracy. We have used a feedforward network for evaluation, but our method is not network specific. We use the output of the penultimate layer only, and so it can be used for other types of neural networks like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Though the method presented here has been evaluated for hand washing, it is generic in nature, and so it can be used for other types of activities. Future work includes exploring the effectiveness of the method for different activity recognition tasks with different types of neural networks.

As our method uses the output of a neural network model, the additional computational cost for inference is very low. In addition to the computation required by the underlying neural network, our method only computes the Mahanabolis distance and compare the distance with the pre-defined threshold. Since we use a pre-trained network, there is no need to train a network for our solution. However, we trained a network for evaluation purposes as there is no pre-trained network available for hand-washing detection using accelerometer data.

Compared to other solutions that require a lot of parameter tuning, there is only one parameter in our method, the distance threshold (percentile), that needs to be tuned. The distance threshold should be set to balance the trade-off between different metrics like precision and recall. For example, a system where recall is more important than precision, the distance threshold can be set to maximize the recall while meeting the requirement for precision. The duration of a hand washing can range from few seconds to over a minute. The hand washing and other data used in this paper have been collected separately, making it infeasible to detect the duration of a hand washing. Detecting the duration of hand washing using interleaved hand washing and other data is a potential future work.

Hand hygiene compliance can be improved by providing real-time alerts when the user forgets to wash hands. For example, beacons can be placed near the patient beds in the hospital and used to detect the proximity of the doctors

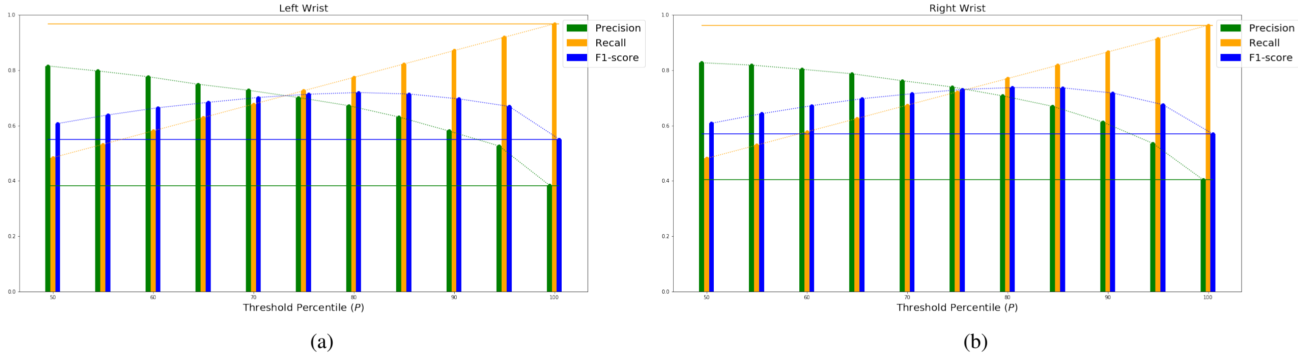

Fig. 9. Precision, Recall and F1-score for different percentiles for (a) left wrist, (b) right wrist. The horizontal lines represents the metrics of the baseline.

or nurses [5]. If they forget to wash hands before entering into or after exiting from the patient room, an alert can be provided immediately. Monitoring hand hygiene practices and real-time reminders would increase hand hygiene compliance. We have not evaluated the performance of our method in improving hand hygiene compliance as well as how it can be generalized to different users. The effectiveness of the solution can be further evaluated by deploying the solution in the real-world. We have not implemented the system on smartwatches. However, recent works [20] show the feasibility of implementing deep neural networks on resource constraint devices. Future works include implementing and evaluating the method on smartwatches. We can place the device on any wrist to detect hand washing since there is no significant difference in the performance. As our solution uses only a single wrist, it is very practical and convenient to be used in the real world.

VI. Conclusion

In this paper, we present a novel solution for hand washing detection that addresses the problem of NULL activities. Our method reduces the false positives by 77% and improves F1-score by 30% compared to the baseline method. The solution is robust against unseen NULL activities, and so it would be very effective in the real-world where people performs a wide range of activities.

Acknowledgment

This work was supported, in part, by NSF CNS-1646470.

References

[1] B. Allegranzi and D. Pittet, "Role of hand hygiene in healthcare-associated infection prevention," *Journal of Hospital Infection*, vol. 73, no. 4, pp. 305–315, 2009.

[2] D. Pittet, B. Allegranzi, H. Sax, S. Dharan, C. L. Pessoa-Silva, L. Donaldson, and J. Boyce, "Evidence-based model for hand transmission during patient care and the role of improved practices," *The Lancet Infectious Diseases*, vol. 6, no. 10, pp. 641–652, 2006.

[3] "CDC HAI prevalence survey," http://www.cdc.gov/hai/surveillance/, last Accessed: 2014-10-07.

[4] M. Ross and J. Guzewich, "Evaluation of risks related to microbiological contamination of ready-to-eat food by food preparation workers and the effectiveness of interventions to minimize those risks," *FDA White Paper, FDA, CFSAN*, 1999.

[5] M. A. S. Mondol and J. A. Stankovic, "Harmony: A hand wash monitoring and reminder system using smart watches," in *proceedings of the 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2015, pp. 11–20.

[6] V. Galluzzi, T. Herman, and P. Polgreen, "Hand hygiene duration and technique recognition using wrist-worn sensors," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. ACM, 2015, pp. 106–117.

[7] H. Li, S. Chawla, R. Li, S. Jain, G. D. Abowd, T. Starner, C. Zhang, and T. Plötz, "Wristwash: towards automatic handwashing assessment using a wrist-worn device," in *Proceedings of the 2018 ACM International Symposium on Wearable Computers*. ACM, 2018, pp. 132–139.

[8] Y. Guan and T. Ploetz, "Ensembles of deep lstm learners for activity recognition using wearables," *arXiv preprint arXiv:1703.09370*, 2017.

[9] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[10] J. Yang, M. N. Nguyen, P. P. San, X. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition." in *Ijcai*, vol. 15, 2015, pp. 3995–4001.

[11] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," *arXiv preprint arXiv:1604.08880*, 2016.

[12] G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and smartwatch-based biometrics using activities of daily living," *IEEE Access*, vol. 7, pp. 133 190–133 202, 2019.

[13] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[14] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," in *Advances in Neural Information Processing Systems*, 2018, pp. 7167–7177.

[15] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The mahalanobis distance," *Chemometrics and intelligent laboratory systems*, vol. 50, no. 1, pp. 1–18, 2000.

[16] D. Pittet, B. Allegranzi, J. Boyce, W. H. O. W. A. for Patient Safety First Global Patient Safety Challenge Core Group of Experts *et al.*, "The world health organization guidelines on hand hygiene in health care and their consensus recommendations," *Infection Control & Hospital Epidemiology*, vol. 30, no. 7, pp. 611–622, 2009.

[17] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," *arXiv preprint arXiv:1610.02136*, 2016.

[18] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," Citeseer, Tech. Rep., 2009.

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[20] S. Bhattacharya and N. D. Lane, "Sparsification and separation of deep learning layers for constrained resource inference on wearables," in *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, 2016, pp. 176–189.