MedBuds: In-Ear Inertial Medication Taking Detection Using Smart Wireless Earbuds

Murtadha Aldeer§
WINLAB
Rutgers University
North Brunswick, NJ, USA
maldeer@winlab.rutgers.edu

David Waterworth§
School of Computing
Macquarie University
Sydney, NSW, Australia
david.waterworth@hdr.mq.edu.au

Zawar Hussain
CSE School
UNSW
Sydney, NSW, Australia
zawar.hussain@unsw.edu.au

Tahiya Chowdhury
ECE Department
Rutgers University
Piscataway, NJ, USA
tahiya.chowdhury@rutgers.edu

Christian Brito
ECE Department
Rutgers University
Piscataway, NJ, USA
christian.brito@rutgers.edu

Quan Z. Sheng
School of Computing
Macquarie University
Sydney, NSW, Australia
michael.sheng@mq.edu.au

Richard P. Martin
WINLAB
Rutgers University
North Brunswick, NJ, USA
rmartin@scarletmail.rutgers.edu

Jorge Ortiz
WINLAB
Rutgers University
North Brunswick, NJ, USA
jorge.ortiz@rutgers.edu

***Abstract*—Wireless earbuds are gaining in popularity these days, especially for smart mobile phone pairing. Some of these devices are getting smart as they embed motion sensors to monitor head and mouth movements. The embodiment of these sensors can enable important mobile health applications such as medication adherence monitoring. Existing solutions are often focused on capturing hand gestures associated with medication retrieval and thus they are inaccurate and do not detect medication ingestion. Other solutions use neck-worn systems which make them uncomfortable and socially unacceptable.**

In this paper, we present *MedBuds*, a smart system for medication-taking activity detection using earbud embedded IMUs and a pairing device (e.g., a smartphone). To evaluate our approach, we conducted preliminary experiments examining semi-medication-taking activities (i.e. swallowing) and non-medication-taking activities (speaking and chewing). Our results show the possibility of distinguishing between these activities with more than 84% accuracy. We believe that by coupling MedBuds with other monitoring techniques (e.g. smart pill bottles), the overall performance of medication adherence monitoring systems can be improved.

***Index Terms*—earbuds, pill taking, medication adherence**

I. Introduction

The reduction in size, energy, and cost of computing and communication technologies are enabling new ways to monitor human activity [1]. In particular, emerging platforms such as earables [2], and the Internet of Things (IoT) will provide rich new sensing capabilities, which in turn can automate activity detection. Further analysis of human activity could then be used for other goals, particularly improving human health. One major problem within the healthcare sector is medication adherence. It is defined as "*the extent to which a person-taking medication adheres to a self-administered protocol*" [3]. Medication adherence is dependent on the medication-intake behavior of the patient as directed by the healthcare provider concerning dosage, timing, and frequency. With the rapid advancement in Cyber-Physical Systems (CPS) for healthcare [4], many technologies on medication adherence monitoring have been proposed [5], [6]. Among these, is the use of smart pill bottles that detect pill opening, pill retrieval [5], and even user identification [7]. However, asserting that these systems can determine id the pill was ingested is overly simplistic and requires more rigorous study. So while using technology for medication adherence monitoring is an attractive field, one approach does not close the loop. Hence, the active non-compliance problem may arise, which is when a user fools the systems and discards the medication pill as he does not agree with the medical professional's treatment [3].

To address these challenges, we propose a system based on wireless earbuds. The information from smart pill bottles can tell if the bottle was opened, if a pill was retrieved, and who retrieved it [8]. Additional information from a smart earable can detect when a pill is swallowed a few seconds after opening the cap by using Inertial Measurement Unit (IMU) sensors and associated machine learning algorithms.

Two facts support the use of earbuds such as eSense [9] for recognizing the activities we are targeting. First, the medication taking activity involves swallowing, which can be

Invited paper.
§Equal contribution.

978-1-6654-8203-5/22/$31.00 ©2022 IEEE
DOI 10.1109/CPHS56133.2022.00008

distinguished from other activities by tracking head movement patterns [10]. Second, previous studies [11] observed that lower jaw movement can be detected in the temporomandible joint, which is the point where the mandible (lower jaw) and the temporal bone (on the skull) meet. The joint location is very close to the ear canal, and hence, when the jaw is moving, it is possible to detect the movement of the mandible through the changes in ear canal depressions. Thus, we use earbuds embedding IMU unit to capture jaw movements during swallowing and distinguish it from other non-swallowing activities to confirm ingestion events after the bottle is opened. Data from the earable is acquired by sampling the embedded IMU sensor that results in changes to the acceleration and angular velocity. The data is transmitted to a smartphone for processing using the Bluetooth protocol. Further processing is performed on a PC where classification algorithms are capable of distinguishing between swallowing activities and other types of activities.

The rest of the paper is organized as follows. Section II presents the related work. In Section III, we describe our proposed system while Section IV the evaluation results of MedBuds. Finally, Section V concludes the paper.

II. Related Work

We live in the era of assistive technology and smart devices. Among these, is the development of technology-based solutions for medication adherence monitoring. A comprehensive review by Aldeer et al. [3] presents different technologies for medication adherence monitoring, including smart pill bottles, wearables, computer vision, and proximity sensing. Focusing on wearable sensors, we can place them in two categories, depending on the placement location on the body: **wrist-worn** and **neck-worn**.

Wrist-worn sensors in the form of smartwatches have been employed for detecting the motions associated with pill bottle opening, pill retrieval, and pill pouring into the secondary hands [12]. The works in [13] and [14] utilized smartwatch sensors for a similar goal. A recent work by Cherian et al. [15] used two smartwatches worn on both wrists of the users to record full-day data including medication taking. In general, wearable systems support mobility and accuracy, but, require contact with the subject's body. Also, these solutions can only detect the motion associated with medication intake (e.g., opening the pill bottle, pill retrieval, etc.), and cannot detect the ingestion of the pill.

A neck-worn system for detecting medication ingestion was proposed in [16]. It was designed to use a pendant-style necklace that embeds a piezoelectric sensor, a battery, and an RF module. Skin motion resulting from pill swallowing is captured by the piezoelectric sensor and voltage is generated as a response. Bluetooth technology is used to send the data to a mobile phone and then classification techniques are used for analyzing the data. Another neck-worn sensor is used in [17]. It is based on acoustic technology for detecting medication ingestion. A microphone placed near the throat is used to capture acoustic data resulting from the ingestion activity to detect the ingestion events. One major barrier associated with the neck-worn systems is the user comfort and social acceptance [18]. It requires the sensor to be worn and in contact with the skin during medicine intake.

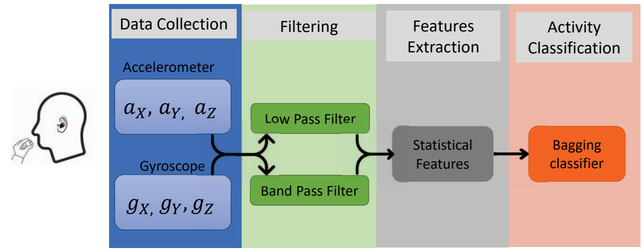
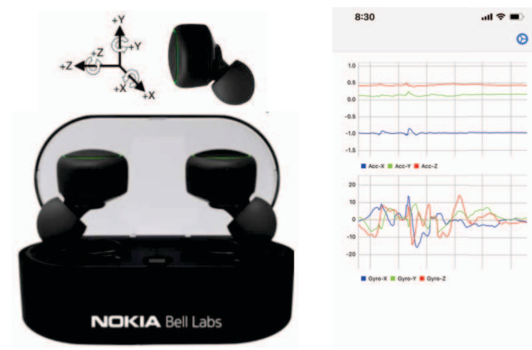


Fig. 1. MedBuds overview.



(a) (b)

Fig. 2. (a) eSense platform [19], and (b) the data collection mobile app.

In this work, we propose a smart earbud-based system that can detect ingestion activity. We investigate the employment of IMU sensors embedded in earbuds to detect swallowing activities (e.g. taking a sip of water, swallowing & drinking, and saliva swallowing) and distinguish them from non-swallowing activities (e.g. chewing and speaking).

III. System Overview

In this section, we describe our proposed system that is based on smart earbuds worn by the user during the medication intake. The architecture of the proposed system is presented in Fig. 1.

A. eSense Platform

MedBuds is based on eSense platform [9]. eSense is an earbud that was designed to enable automatic tracking of a set of head and mouth-related activities including eating, drinking, speaking, etc. It embeds motion and audio sensors (accelerometer, gyroscope, and microphone). It uses Bluetooth Low Energy (BLE) radio to transmit the collected data to a nearby phone.

B. Data Collection

We use eSense with a sampling rate of 100 samples/second that is set for the IMU sensors. Five subjects wore the left earbud only as it houses the IMU. Hence, this does not impact our experiments both right and left joints on the jaw bone move together. Thus jaw movement can be detected equally in both ears. During our data collection, we asked the subjects to perform a set of activities: target activities (drinking a sip of water, swallowing one M&M's® candy with water, and swallowing saliva) and two non-target activities (chewing one candy and swallowing it and speaking).

It is worth mentioning that although the chewing activity is expected to be followed by swallowing, we do not place it with the target (swallowing) activities. This is because we are interested in showing that we can distinguish between these activities using the earable platform. Nonetheless, chewing events can be part of eating food, thus we place it in the non-swallowing class. The speaking activity was performed where the subject reads 2 lines of text of his choice in his preferred language. Each activity is performed 10 times and data is collected and labeled using a smartphone running a data collection application (Fig. 2). Finally, for each activity, a comma-separated values (CSV) file is generated and saved on the phone and upon completion of data collection, all files are transmitted via Bluetooth to an edge device for processing. A total of 200 files that correspond to the number of performed activities are collected.

C. Data Pre-Processing and Feature Extraction

Each raw file contains data for a single activity, for a single subject. The files are read into a raw time-series matrix consisting of a sample per row, each sample consisting of 6 data points from the 3 accelerometer and gyroscope axes. First, we clean the accelerometer data by passing them through a filtering stage. Low-pass and high-pass filters with a 2Hz cutoff are used. We also compute the pitch and roll from the accelerometer data using Eq. 1 and 2.

$$roll = atan2\left(acc_y, acc_z\right) \tag{1}$$

$$pitch = atan\left(-\frac{acc_x}{acc_y \cdot sin(roll) + acc_z \cdot cos(roll)}\right) \tag{2}$$

The original accelerometer and gyroscope data, along with the filtered acceleration data, pitch, and yaw and concatenated into a final time-series matrix. From this matrix we generate statistical, temporal and spectral features by processing each input feature matrix using TSFEL [20], a time-series feature extraction library in Python. Using TSFEL, we use the 27 features from the TSFEL Human Activity Range (HAR) feature set. A window is applied to each activity along the time dimension and the TSFEL feature extraction transforms the window (matrix) into a vector by applying each feature extractor to each series. Each feature vector has an associated activity ID (i.e. chewing, speaking, drinking, etc.) and subject ID (for cross-validation) which is concatenated to produce a final dataset consisting of a feature matrix, a label vector, and a group vector.

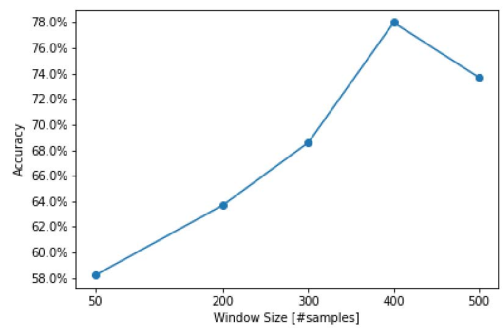


Fig. 3. Selecting the optimal value of window size.

D. Classification Model

We use Leave One Subject Out (LOSO) cross-validation to compute the accuracy, precision, recall, and F1-score. LOSO is used to account for heterogeneity between subjects. LOSO trains one model per subject per cross-validation fold. For each fold, one subject is used as test data and the remaining subjects are used as training data. Achieving high accuracy when using LOSO provides confidence that the model has learned a generalized representation of each activity rather than overfitting to subject-specific idiosyncrasies. It also simulates a setting where we train a model on some subjects and then perform inference on other subjects.

We initially considered a few shallow classification models including, Support Vector Machine (SVM), Naive Bayes, and a decision tree based bagging classifier. We found that the bagging classifier consistently outperformed SVM and Naive Bayes. For this reason we used the bagging classifier for all further experiments. The classifier used is a balanced bagging classifier from Imbalanced-learn [21] with a histogram-based gradient boosted classification tree from scikit-learn [22] as a base estimator. Optuna [23], a hyper-parameter search framework, was used to find the best hyper-parameters using accuracy as the objective. Using Optuna, we identified the window size to be the most important hyper-parameter. The relationship between window size and classification accuracy when all other parameters are held constant is shown in Fig. 3. The peak was obtained at 492 samples, the decline after this point occurs because longer windows resulted in a window size that is longer than the average activity length, resulting in these activities being excluded. The second most significant hyper-parameter was the number of base estimators used by the classifier. We experimented with tuning parameters of the base estimator and selecting additional TSFEL features, but doing so failed to outperform using the defaults in conjunction with tuning the window size and number of base estimators so we performed all further experiments using these two parameters only.

IV. Performance Results

In this section we present the performance of MedBuds. To evaluate the performance of MedBuds, we try to answer the

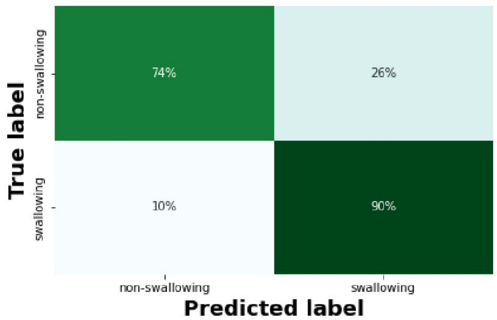


Fig. 4. Confusion matrix of the two-class scenario.

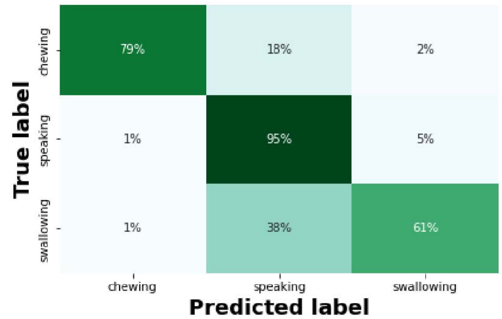


Fig. 5. Confusion matrix of the three-class scenario.

following questions:

- How is accuracy affected by varying the number of classes?
- How is the system performance impacted with respect to the sampling rate?
- How is the system performance changed with respect to the number of sensors used, i.e. accelerometer + gyroscope or accelerometer only?

1) Detection performance of the system for two-class and three-class scenarios : In this section, we investigate the performance of our system under different scenarios. First, we consider a binary classification scenario, where we only distinguish between swallowing and non-swallowing activities. In this scenario we combine swallowing a sip of water, taking candy with water, and swallowing saliva; into one class, and the speaking and chewing activities into one class. Fig. 4 shows the performance of the system. It is noticeable that the system achieves an accuracy of more than 84% for LOSO. This proves that the approach performs well at distinguishing the swallowing-related activities from all other activities. This is the primary use application for medication adherence detection, assuming that the detection of a swallowing event (pill swallowing and/or drinking a sip of water) following the opening of a pill bottle increases the probability of user compliance.

Second, we consider a multi-class classification scenario, where we aim to distinguish swallowing among other classes. Fig. 5 shows the confusion matrix for classifying three classes (swallowing, speaking, and chewing). Hence, similar to the two-class scenario, the swallowing class includes three sub-activities. The results indicate that we can still differentiate the three classes mentioned earlier with accuracy closer to 80%.

Note that for the following sections, we report the results for the three-class scenario.

2) Effect of the sampling rate on classification performance: System lifetime is proportional to the sensor sampling rate [24] making it an important factor in battery-powered systems. To tackles this, we evaluate the impact of the sampling rate on the classification accuracy of our system. To produce datasets with different sampling rates, we down-sampled the

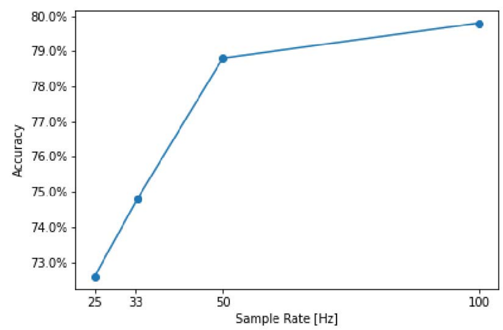


Fig. 6. Effect of sampling rate variation on the performance.

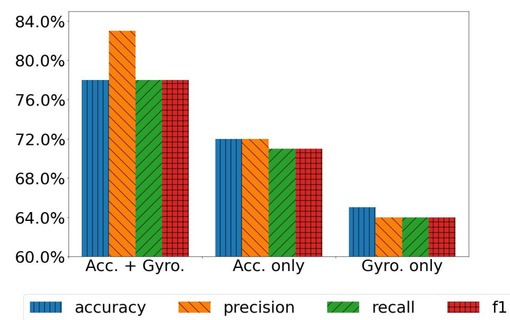


Fig. 7. Individual sensor importance.

data sampling rate to simulate data collected at 50Hz, 33Hz and 25Hz measurement rate, respectively. We applied the same pipeline to these new datasets. However, the window size was adjusted so that its time duration in seconds remains constant. Fig. 6 shows the accuracy for the various measurement rates. We can see that the classification performance does not affect as we lower the sampling rate to the half. However, the performance degrades as the measurement rate decreases, with a sharp drop-off at a measurement rate of 25Hz.

3) Analyzing Individual Sensor Importance: Finally, we performed different experiments to determine the impact of using the accelerometer, gyroscope, and a combination of both sensors on the final classification accuracy. We evaluate the performance of our system by using our original dataset from both sensors, and by generating two new datasets, one from the accelerometer sensor and the other from the gyroscope sensor. We applied the same pipeline in each case.

We report the LOSO performance of the bagging classifier in Fig. 7. The figure shows the class weighted accuracy, precision, recall, and F1-score for the experiments. Notice, the performance of our model decreases when using either the accelerometer or the gyroscope. However, using the accelerometer only, the performance outperformed the gyroscope only case. One takeaway is that the combination of both sensors outperforms the single sensor case.

V. Conclusions and Future Work

Medication non-adherence is a big problem in healthcare. It affects the patients' health and can potentially result in irrevocable complications. Another problem is increasing drug wastage higher healthcare costs. This paper introduces MedBuds, a smart earable system that captures in-ear IMU data that results from jaw movement. Our preliminary results show that swallowing events can be detected using the proposed system, and are distinguishable from other activities involving jaw movement, such as speaking and chewing.

Looking forward, we plan to address other challenges to enhance MedBuds performance further.

- Our dataset was very small as it only included five subjects. In the future, we will conduct experiments with more populations.
- Although our experiments included swallowing activities, we did not perform a real word experiment that involves swallowing medication pills of different sizes. In the future, we will conduct experiments in which we use placebo pills to investigate if we can detect different size pills.
- Finally, we plan to incorporate MedBuds with a smart pill bottle [8] in a two-step verification system for detecting when a pill bottle is opened, who opened it using the smart bottle, and when a pill is consumed using the earable sensor.

Acknowledgment

We would like to thank Nokia Bell Labs, Cambridge, UK for providing us with the eSense earbuds. The work of David Waterworth is funded by the RoZetta Institute, the CSIRO and CIM Pty Ltd.

References

[1] Z. Hussain, Q. Z. Sheng, and W. E. Zhang, "A review and categorization of techniques on device-free human activity recognition," *Journal of Network and Computer Applications*, vol. 167, p. 102738, 2020.

[2] R. R. Choudhury, "Earable computing: A new area to think about," in *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, ser. HotMobile '21, 2021, p. 147–153.

[3] M. Aldeer, M. Javanmard, and R. P. Martin, "A review of medication adherence monitoring technologies," *Applied System Innovation*, vol. 1, no. 2, p. 14, 2018.

[4] S. Amin, T. Salahuddin, and A. Bouras, "Cyber physical systems and smart homes in healthcare: Current state and challenges," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, 2020, pp. 302–309.

[5] M. Aldeer, M. Alaziz, J. Ortiz, R. E. Howard, and R. P. Martin, "A sensing-based framework for medication compliance monitoring," in *Proceedings of the 1st ACM International Workshop on Device-Free Human Sensing*, ser. DFHS'19, 2019, p. 52–56.

[6] S. Faisal, J. Ivo, and T. Patel, "A review of features and characteristics of smart medication adherence products," *Canadian Pharmacists Journal/Revue des Pharmaciens du Canada*, vol. 154, no. 5, pp. 312–323, 2021.

[7] M. Aldeer, R. E. Howard, R. P. Martin, and J. Ortiz, "User identification across multiple smart pill bottle systems: Poster abstract," in *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021)*, ser. IPSN '21, 2021, p. 400–401.

[8] M. Aldeer, "User identification using smart pill bottles: Systems and machine learning models: Phd forum abstract," in *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021)*, ser. IPSN '21, 2021, p. 414–415.

[9] F. Kawsar, C. Min, A. Mathur, and A. Montanari, "Earables for personal-scale behavior analytics," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 83–89, 2018.

[10] A. Ferlini, A. Montanari, C. Mascolo, and R. Harle, "Head motion tracking through in-ear wearables," in *Proceedings of the 1st International Workshop on Earable Computing*, ser. EarComp'19, 2019, p. 8–13.

[11] S. Rupavatharam and M. Gruteser, *Towards In-Ear Inertial Jaw Clenching Detection*, 2019, p. 54–55.

[12] H. Kalantarian, N. Alshurafa, E. Nemati, T. Le, and M. Sarrafzadeh, "A smartwatch-based medication adherence system," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2015, pp. 1–6.

[13] R. Wang, Z. Sitová, X. Jia, X. He, T. Abramson, P. Gasti, K. S. Balagani, and A. Farajidavar, "Automatic identification of solid-phase medication intake using wireless wearable accelerometers," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2014, pp. 4168–4171.

[14] N. Hezarjaribi, R. Fallahzadeh, and H. Ghasemzadeh, "A machine learning approach for medication adherence monitoring using body-worn sensors," in *Proceedings of the 2016 Conference on Design, Automation & Test in Europe*, 2016, pp. 842–845.

[15] J. Cherian, S. Ray, and T. Hammond, *An Activity Recognition System for Taking Medicine Using In-The-Wild Data to Promote Medication Adherence*, 2021, p. 575–584.

[16] H. Kalantarian, B. Motamed, N. Alshurafa, and M. Sarrafzadeh, "A wearable sensor system for medication adherence prediction," *Artificial Intelligence in Medicine*, vol. 69, pp. 43–52, 2016.

[17] T. Olubanjo and M. Ghovanloo, "Real-time swallowing detection based on tracheal acoustics," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 4384–4388.

[18] H. Kalantarian, N. Alshurafa, and M. Sarrafzadeh, "A survey of diet monitoring technology," *IEEE Pervasive Computing*, no. 1, pp. 57–65, 2017.

[19] F. Kawsar, C. Min, A. Mathur, M. Van den Broeck, U. G. Acer, and C. Forlivesi, "Esense: Earable platform for human sensing," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '18, 2018, p. 541.

[20] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "Tsfel: Time series feature extraction library," *SoftwareX*, vol. 11, p. 100456, 2020.

[21] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1–5, 2017.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[23] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery &; Data Mining*, ser. KDD '19, 2019, p. 2623–2631.

[24] M. M. N. Aldeer, R. P. Martin, and R. E. Howard, "Tackling the fidelity-energy trade-off in wireless body sensor networks," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 2017, pp. 7–12.