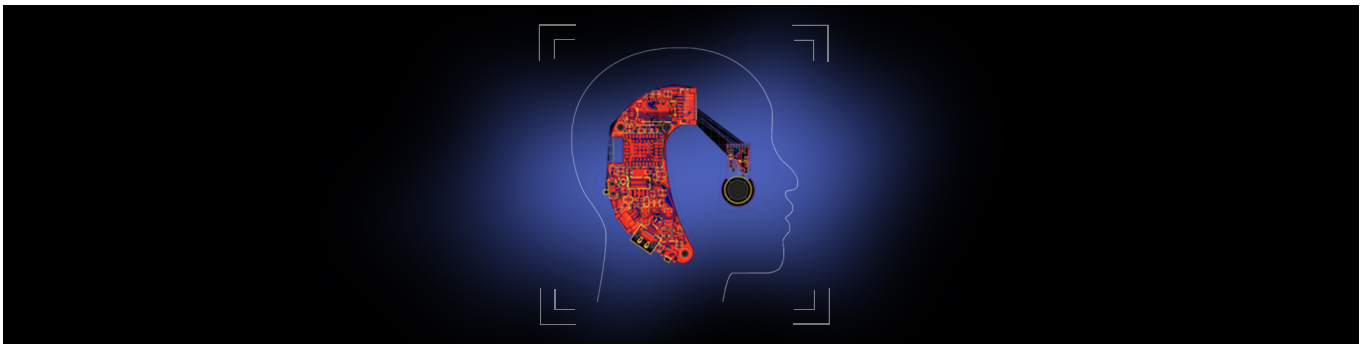
# Real-Time Recognition of Boxing Head Gestures with IMU-Earables: Machine Learning and Dynamic Time Warping

THOMAS SEPANOSIAN, University of Twente, The Netherlands



The rising prominence of earables, wearables meant to be worn around the ear, represents opportunities for novel applications. Previous research showcases the potential of earables in the context of sports; however, a gap is present for boxing, more specifically in recognition of defensive manoeuvres, even outside the realm of earable development. Thus, this paper explores the capability of real-time, IMU-based boxing head gesture recognition using the open-source *OpenEarable* framework through classical machine learning and dynamic time-warping approaches. A dataset was collected consisting of approximately 460 samples of left/right slips, left/right rolls, and pullbacks, by a hobby-level boxer. The results revealed that utilizing dynamic time warping in combination with templates based on barycenter averaging achieves effective results in gesture recognition. During the testing phase, the implemented algorithm achieved a high accuracy score of 99% on the collected dataset. This performance was further validated in a deployed real-world scenario, where the algorithm maintained an overall accuracy of 96% across 50 repetitions per gesture. Additionally, the system demonstrated robustness against variations in gesture execution speed and intensity.

Additional Key Words and Phrases: Earables, Inertial Measurement Unit, Real-time Head Gesture Recognition, Machine Learning (ML), Dynamic Time Warping (DTW), Ubiquitous Computing

## 1 INTRODUCTION

With technology becoming increasingly integrated into everyday life, the evolution from basic wearables to more advanced devices like earables marks a significant shift. Earables are devices designed to be worn in or around the ear, and they extend beyond traditional earphones by incorporating sensors enabling a wide range of applications [17, 22].

Example applications include sensing fine-grained facial expressions [27], enabling user authentication through sensing of heart rate, gait and breathing patterns [2], and head gesture recognition through PPG signal readings [15]. The placement of earables on the head makes them well-suited for tasks requiring head gesture recognition, as demonstrated by Xu et al., who facilitate hands-free text entry [29].

In the literature, most of the studies utilize the *Nokia Bell Labs eSense* earbuds; however, with this device's end of life [1], the *OpenEarable* platform[2] has emerged as an alternative. This platform follows a fully open-source approach. This enables researchers and developers to directly modify and enhance both the hardware and software components, thereby facilitating collaborative development and novel research perspectives [23]. To exemplify its potential, applications such as a `Jump Rope Counter`, `Posture Tracker` and `Tightness Meter` have been developed [3].

In sports, the precision and real-time recognition capabilities of these devices can be particularly beneficial. Research into boxing-specific gesture recognition, particularly defensive manoeuvres, remains underexplored despite its potential to significantly enhance training and performance. This gap presents a unique opportunity to leverage the OpenEarable platform to facilitate real-time recognition of gestures and feedback, potentially enhancing training.

This paper investigates the capabilities of the *OpenEarable* platform, focusing on its application in real-time gesture recognition for boxing. The following research questions are posited:

- **RQ 1:** How can classical machine learning and dynamic time warping techniques be applied to IMU data obtained through *OpenEarable* devices to attain real-time and effective boxing head gesture recognition?
- **RQ 2:** How can a boxing head gesture recognition system be integrated into the *OpenEarable* framework to provide real-time feedback?

To explore these research questions, both classical machine learning and dynamic time-warping (DTW) techniques are employed. Classical machine learning models provide a robust foundation for

*TScIT 41, July 5, 2024, Enschede, The Netherlands*


[1] Details about the discontinuation can be found at https://www.esense.io/, accessed in June 2024. A second generation is reportedly under development but is not yet available.
[2] https://github.com/OpenEarable/open-earable , accessed in June 2024
[3] These applications are showcased at https://open-earable.teco.edu/ , accessed in June 2024.

pattern recognition, while DTW offers a complementary approach, potentially yielding novel insights into the research questions.

The remainder of this paper is organized as follows: Section 2 reviews previous studies on earables, head gesture recognition, and boxing gesture detection, identifying the gaps this research aims to address. Section 3 outlines the methods used in this study. Section 4 presents the research results. Finally, Section 5 draws conclusions, addresses the research questions, and discusses potential avenues for future work.

## 2 RELATED WORK

This section reviews the emerging field of earables and explores their applications, particularly in head gesture recognition. Table 1 presents a comparative analysis of relevant prior research in relation to this work.

### 2.1 Earables: A Novel Sensing Platform

Earables, have been recognized for their vast range of potential applications. They are particularly well-suited for head gesture recognition due to their placement and, commonly, integration of an IMU sensor. A comprehensive taxonomy of phenomena, including head gesture recognition, is presented in [22], emphasizing the wide variety of potential applications that can leverage earables.

### 2.2 Head Gesture Recognition

Head gesture recognition is an area of active research across diverse computing platforms. For example, an IMU-augmented hat has been utilized to recognize head gestures in [28]. The authors employed a 3-axis accelerometer, 3-axis gyroscope, and 3-axis compass to collect sensor readings for nodding, shaking, raising up once, raising down once, turning left once, and turning right once. Their decision tree model achieved 94.63% accuracy and executed within 1.54 ms, while the random forest model reached 99.17% accuracy but with a higher execution time of 18.17 ms. These findings highlight the potential for a real-time application, though the physical obtrusiveness of the hat could limit practical use compared to more discreet alternatives like earables. Furthermore, using computationally expensive features such as the average absolute difference could be a bottleneck for on-edge inference.

Similarly, [1] explored user authentication through head gesture characteristics using smart glasses equipped with accelerometers, gyroscopes, and geomagnetic sensors. They collected data for movements such as moving the head in circles, squares, and triangles, segmented using sliding windows, and translated into simple features: minimum, maximum and mean. They observed an EER of 2.4% for authentication and an f1-score of 98.7% for identification, further underscoring the potential for real-time earable applications. However, as with the augmented hat, earables offer a more unobtrusive and pervasive option, particularly as earbuds become increasingly prominent in everyday use.

Building upon the potential of earables, the *Nokia Bell Labs eSense* earbuds have been shown to effectively capture everyday gestures, such as eating, nodding, and head shaking, which can facilitate the detection of social interactions [12]. Advances in gesture recognition also encompass facial expressions like smiling, talking, and yawning, which have been effectively recognized using hierarchical methods applied to inertial signals collected through these earbuds [7].

These technological advancements open new avenues for practical applications, such as developing head gesture-based interfaces. For example, *HeadText* provides an innovative interface for typing [29].

### 2.3 Sports and Earables

Recent research suggests that earables offer promising applications within the sports field. For example, [16] demonstrated how an IMU at the waist, in conjunction with earables, can assist users in performing core training exercises correctly. Furthermore, earables have been used to collect data on basketball dribbling [9] and recognition of fitness exercises [25], showcasing their versatility and potential in enhancing athletic performance.

### 2.4 The Gap: Boxing-Specific Head Gestures

Despite the advancements of earables in head gesture recognition and their applications in sports, there is a notable gap in research, specifically targeting boxing. In boxing, defensive manoeuvres heavily depend on precise and quick head movements. Even outside the realm of earables, when it comes to human activity recognition research in boxing, research most commonly focuses on punches [11, 26], rather than head movements for defensive techniques, despite it being a crucial aspect in competitive boxing [10]. Furthermore, there is a lack of research on the utilization of earables in conjunction with DTW. Studies most commonly focus on the use of machine learning. Additionally, existing literature often does not explore the implementation of an end-to-end system to evaluate real-world efficacy.

This research aims to fill these gaps by exploring the use of earables for real-time head gesture recognition, focusing on boxing-specific movements such as slipping, rolling and pulling back. To do so, this study will extend the current use of the OpenEarable platform, an open-source earable platform.

## 3 METHODOLOGY

In this section, the methodology employed to investigate the capabilities of the OpenEarable platform for head gesture recognition in the context of boxing is outlined. An overview of the methodology is provided in Figure 1. The following subsections discuss the details of the blocks mentioned in the overview.

### 3.1 Gestures

The research focused on three specific boxing manoeuvres: slipping, rolling, and pulling back. These gestures are crucial defensive techniques boxers use to evade punches and position themselves advantageously.

- **Slipping**: This maneuver involves quick lateral movement of the head to either side, approximately the width of a boxing glove, to avoid straight punches aimed at the head. It is typically used to counter jabs and crosses, allowing for immediate counterattacks due to the minimal movement required. This is simulated by performing the slip in front of a boxing glove attached to the roof at head-level height, akin to a slip bag.

Table 1. Summary of Related Work

| Reference | Device | Recognition Goal(s) | Techniques | Results |
|---|---|---|---|---|
| Wu et al. [28] | Head-mounted wearable | Nodding, Shaking, Raising, Bowing, Turning Left/Right | (Weka) J48 & Random Forest models | J48: 94.63% Accuracy, 1.54 ms; RF: 99.17% Accuracy, 18.17 ms |
| Agac and Incel [1] | Smart Glasses | Identification & Authentication through Head movements: Circle, Square, Triangle, ... | Feature extraction (mean, min, max) → Multiple models (RF, Adaboost) | 99.3% f1-score for identification |
| Laporte et al. [12] | Nokia Bell Labs eSense earbuds | Nodding, Speaking, Eating, Standing still, Head shaking | CNN (Accelerometer & Gyroscope) | 80% Balanced Accuracy |
| Pansiot et al. [18] | Custom Earable | Climbing performance: Fluidity, Strength, Endurance, Speed | PCA and Gaussian Mixtures | Polar graph performance representation |
| Mavus and Sezer [14] | Head-mounted wearable | Rotating (Clockwise/Counter), Shaking, Nodding | Dynamic Time Warping | Approx. 80% Accuracy per gesture |
| Li and Hu [13] | Smart Glasses | Nodding, Tilting (Up/Right/Left), Shaking (Left/Right) | Dynamic Time Warping | 100% Accuracy |
| Xu et al. [29] | Custom Earable | Turning Left / Right / Down | K-Nearest-Neighbor & Dynamic Time Warping | Approx. 90% Accuracy |
| Motokawa et al. [16] | Nokia Bell Labs eSense earbuds & Waist-mounted Accelerometer | Core Training Monitoring and Support | Rule based feedback | n/a |
| Strömbäck et al. [25] | Nokia Bell Labs eSense earbuds, among other devices | Exercise Recognition & Repetition Counting | Multimodal Deeplearning | 82% accuracy for the earbud device |
| This study | OpenEarable Earable | Real-time Boxing Head Gesture Recognition | Classical Machine Learning & Dynamic Time Warping | 99% Accuracy, Real-time feedback |

- **Rolling**: Also known as bobbing and weaving, rolling involves moving the head in a circular "U" shape motion. This technique is crucial for dodging hooks and uppercuts. The study simulates this by performing rolls underneath the slipback set at head height, encouraging replicable and consistent rolls.
- **Pulling Back**: Also known as the lean back, involves a quick backward movement of the head to avoid straight punches. Similar to the slip, for this study, lean backs are performed in front of a slip bag.

An illustration of these gestures is provided in Figure 2. Given the variability in boxing styles, this study standardizes the maneuvers by using an orthodox stance, characterized by the left foot forward and hands guarding the face and torso.

## 3.2 Data Collection

The study was conducted with a single participant, the researcher. The participant is a hobbyist skill-level boxer. Thus, their executions of the gestures may differ from those performed by more experienced boxers or those completely unfamiliar with boxing.

Data collection spanned several sessions across multiple days, ensuring a comprehensive dataset. Each session was dedicated to performing and recording all the gestures repeatedly for a predetermined number of iterations. This approach allowed each gesture to be captured in isolation before proceeding to the next. Recordings were made using the OpenEarable dashboard's data recorder, complemented by a screen recording for enhanced clarity. To distinctly mark the beginning and end of each gesture's recording, the participant briefly jumped and moved their head around before starting each set of iterations and after each set of iterations.

The *OpenEarable* device, equipped with an Inertial Measurement Unit (IMU), was used to collect motion data. Using Bluetooth, data was streamed at 50Hz in real-time to the *OpenEarable* web dashboard, deployed locally on a laptop. The IMU data included 3-axis accelerometer, gyroscope, magnetometer readings and timestamps.

*3.2.1 Labeling.* Post-collection, the data was labelled using *Edge-ML*, an open-source and browser-based tool directly supporting the labelling of data obtained through the *OpenEarable* dashboard [4]. The data was inspected together with the recorder video to manually

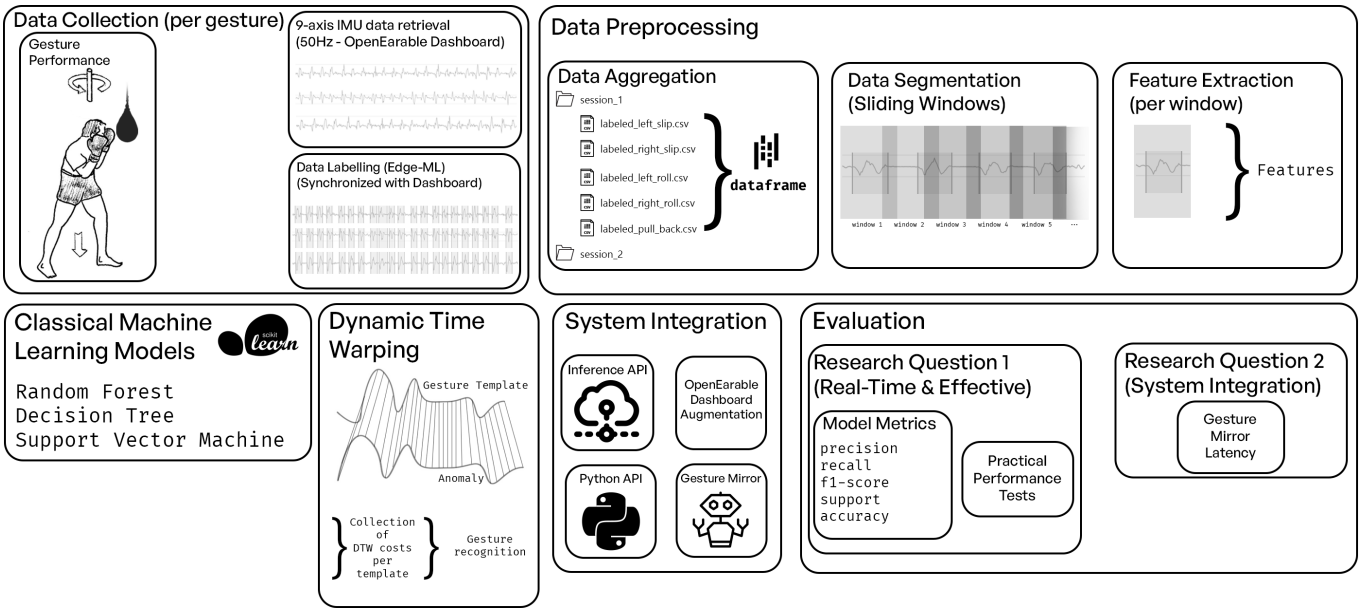



Fig. 1. Methodology Overview

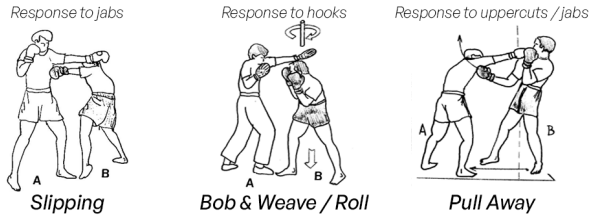



Fig. 2. The target head gestures demonstrated by practitioner B, sparring against opponent A

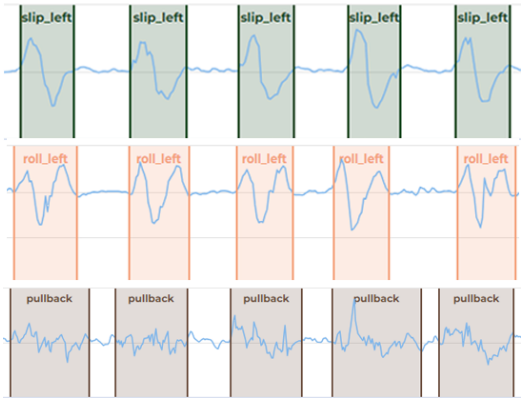



Fig. 3. Labelling of data in Edge-ML (Gyroscope, y-axis (°/s))

label the gestures. An example of labelled data in Edge-ML has been provided in Figure 3.

Table 2. Aggregated Data - Sequences Information

| Gesture | No. of Sequences | Avg. Length (samples) |
|---|---|---|
| Idle | 2296 | 24.50 |
| Left Slip | 460 | 38.16 |
| Right Slip | 459 | 38.59 |
| Left Roll | 458 | 54.83 |
| Right Roll | 459 | 55.95 |
| Pull Back | 459 | 53.44 |

## 3.3 Data Preprocessing

*3.3.1 Aggregation & Segmentation.* The raw, labelled IMU data was exported to CSV format for further processing. Using Python, the data underwent preprocessing, which involved a sliding window approach. During data aggregation of the sessions, the number of sequences and the average sequence length were extracted for every gesture. Statistics for these sequences are provided in Table 2. With this information, several window sizes and overlap sizes were considered, with a window size of 50 and an overlap of 25 being chosen as the most optimal, as apparent from testing multiple configurations which were based on previously mentioned average sequence lengths. Window labels were determined through a majority voting process based on the data within the window. This segmentation process resulted in a collection of windows, with approximately 800 windows for the slips, 1100 for the rolls, and 1500 for the idle gesture.

*3.3.2 Feature Extraction.* To train the models, various features were extracted from the windowed segments of sensor data. The selection of these features was guided by evaluating the performance of different models, manually assessing feature distinction between classes

in Weka [6] (illustrated in Figure 4), and utilizing tsfresh [3]. Considering the requirements for real-time performance and resource constraints of the earable device, the selected features include `minimum`, `maximum`, `standard deviation`, `root mean square`, `10th quantile`, and `absolute maximum` from both the accelerometer's x-axis and the gyroscope's y-axis.

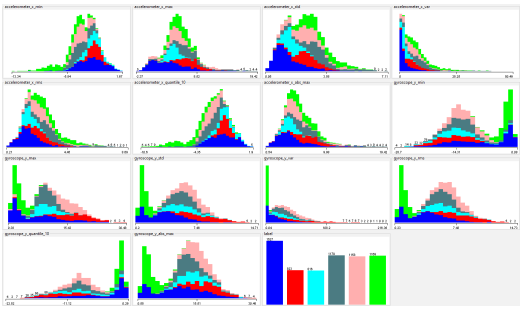


Fig. 4. Feature Visualization using Weka

### 3.4 Classical Machine Learning Approach

The extracted features were used to train various classical machine-learning algorithms. The algorithms evaluated included `Random Forest`, `Decision Tree` and `Support Vector Machine`. These algorithms were chosen for their superior performance in classification tasks. The SciKit-Learn library [19] was used to implement and evaluate the models. All data used for model development and testing underwent a train test split, with a 70:30 ratio, respectively. A collection of Jupyter notebooks concerning the development of these models is made available in a public GitHub repository [4].

### 3.5 Dynamic Time Warping Approach

In addition to classical machine learning techniques, this study also employed Dynamic Time Warping (DTW) for recognizing gestures. DTW is particularly effective in measuring the similarity between temporal sequences, which is essential in scenarios where the same gestures might be performed at varying speeds and intensities. For example, in boxing, boxers might perform rolls more smoothly depending on the intensity of their opponent's pressure.

However, employing DTW effectively requires careful selection of a reference template that accurately represents the gesture across different users. To enhance the reliability and generalizability of these templates, this study utilized Dynamic Time Warping Barycenter Averaging (DBA) [5, 20, 21]. Unlike simple averaging methods, which may skew towards outliers or be adversely affected by noise, DBA constructs a more robust central sequence that better preserves the intrinsic properties of the gestures being analyzed. Templates were generated for every gesture, with the templates consisting of the gyroscope's y-axis DBA sequence, followed by the accelerometer's x-axis sequence DBA of said gestures, resulting in a single template per gesture. This approach is depicted in Figure 5. A collection of Jupyter notebooks is made available in the mentioned GitHub repository.


Fig. 5. DTW Template creation for the left slip gesture using DBA (Showcasing only the first 5 sequences, instead of all 460)

### 3.6 System Integration

Integrating the boxing gesture recognition system into the OpenEarable framework consisted of multiple procedures. The OpenEarable platform has a wide range of software accessible, including a dashboard [5], and a Flutter library [6], which serve as ways to communicate with the earable. The aforementioned approaches to recognizing boxing head gestures were developed in Python. To assist with gesture recognition development, a Python CLI application, akin to the dashboard and Flutter implementation, has been developed [7].

To make the gesture recognition more accessible and extensible, a Flask [8] implementation has been developed to host an inference API. Given this API, the OpenEarable dashboard has been augmented to provide real-time boxing head gesture recognition by utilizing the API [8]. Visual feedback is provided through the form of a gesture mirror, wherein an animated character performs the same gestures as the user. The augmented dashboard, featuring the added head gesture recognition module, is showcased in Figure 6.

[4]Made available at: https://github.com/Thomas-mp4/EarableBoxingHeadGestureRecognition

[5]https://github.com/OpenEarable/dashboard , accessed in June 2024

[6]https://github.com/OpenEarable/open_earable_flutter , accessed in June 2024

[7]Made available at: https://github.com/Thomas-mp4/EarableBoxingHeadGestureRecognition

[8]Made available at: https://github.com/Thomas-mp4/OpenEarableAugmentedDashboard-BoxingHeadGestureRecognition


Fig. 6. Augmented OpenEarable Dashboard, showcasing the recognition of a left roll in the head gesture recognition module in the bottom right

## 3.7 Evaluation

*3.7.1 Research Question 1.* To evaluate the effectiveness and efficiency of the real-time boxing head gesture recognition approaches, multiple factors were considered:

- **Model Metrics**: For the developed machine learning models, multiple metrics were extracted and analyzed, including insights into the following:
  - Accuracy: Overall correctness
  - Precision: Effectiveness in predicting positives
  - Recall: Effectiveness in identifying true positive
  - F1-score: Representation of harmonic mean of precision and recall
  - Confusion Matrix: Representation of classification errors / mislabeling
- **Practical Performance Tests**: For the most effective approach, real-time practical tests were conducted, to determine real-world applicability.

*3.7.2 Research Question 2.* To evaluate the degree to which integration with the OpenEarable framework has been attained, tests were performed to measure the delay between the gesture being performed in real-time and the gesture being performed in the gesture mirror, using video and screen recordings to measure the difference in frames per second.

## 3.8 Ethical Considerations

The Computer & Information Sciences (CIS) committee reviewed and approved the study's design and methodology. Ethical considerations included the potential misuse of AI and model biases arising from training on data from a single participant.

## 4 RESULTS AND DISCUSSION

In this section, the results are discussed in relation to the research questions. First, the performance of the classical machine learning models is discussed. Then, the performance of the dynamic time-warping approach is presented, including the differences between the two approaches. Finally, the results of integrating with the OpenEarable framework are shown, highlighting the practical performance of the gesture recognition system.

*4.0.1 Recognition Performance with Classical Machine Learning Models.* Multiple machine learning models have been developed, including a decision tree, random forest, and support vector machine. The models' confusion matrices are presented in Figure 7.

Important to note is that apart from a set random number generator, the default hyperparameters were employed for both the Random Forest and Support Vector Machine algorithms. However, for the decision tree model, a parameter-tuning exercise was conducted: with the `max_depth` parameter being set to 5, and the `min_samples_leaf` parameter set to 300, in an effort to achieve a model that generalizes well. This decision was based on a manual inspection of results and the usage of GridSearchCV [9]. The window size and overlap size, 50 samples and 25 samples, respectively, were chosen in a similar fashion.

The classification report results of the best-performing model, the random forest model, are provided in Table 3. A recognition system with accuracies of approximately 70% and higher, with most miss-classifications being made within the same gesture types (e.g. left and right rolls) in theory could be adequate for real-time applications, assuming the system is not used in critical contexts. However,

[9] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html, accessed in June 2024

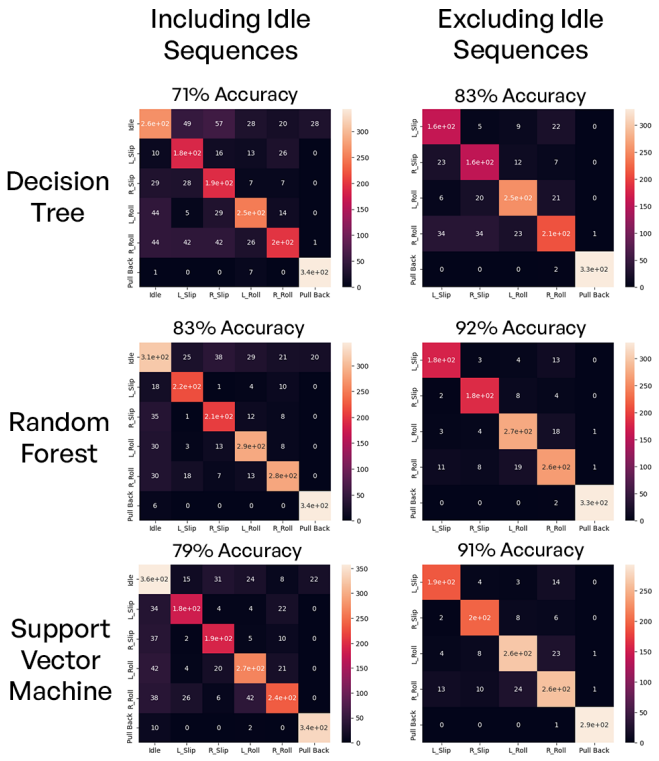



Fig. 7. Machine Learning Models - Gesture Recognition Confusion Matrices

Table 3. Classification Metrics Summary (Random Forest excluding idle sequences

| Label | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Left Slip | 0.92 | 0.90 | 0.91 | 195 |
| Right Slip | 0.92 | 0.93 | 0.93 | 197 |
| Left Roll | 0.90 | 0.91 | 0.91 | 300 |
| Right Roll | 0.88 | 0.87 | 0.87 | 303 |
| Pullback | 0.99 | 0.99 | 0.99 | 332 |
| **Accuracy** | 0.92 (1327 Support) | | | |
| **Macro Avg** | 0.92 Precision, 0.92 Recall, 0.92 F1-Score | | | |
| **Weighted Avg** | 0.92 Precision, 0.92 Recall, 0.92 F1-Score | | | |

when these models were deployed in a real-time environment, none performed as one would assume given the testing results. Despite developing a Python CLI application to circumvent a potential delay being caused by the JavaScript implementation utilizing the Python Flask inference API, results remained poor, with a right slip never being recognized, and, apart from the idle class, running into frequent miss-classifications.

This discrepancy between real-time performance and testing performance could be due to poor generalizability of the models. Given that gestures were performed only approximately 460 times by a hobbyist-level boxer, the data potentially lacks the capability to spawn models that can classify gestures that slightly deviate from how they were performed prior. Furthermore, the feature visualization in Figure 4 suggests potentially that the features do not differentiate between the classes distinctly enough. Thus, this approach did not sufficiently satisfy the research goal of attaining real-time gesture recognition despite the promising results in testing.

Table 4. Classification Metrics Summary (DTW with DBA templates)

| Label | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Left Slip | 0.98 | 1.00 | 0.98 | 460 |
| Right Slip | 0.99 | 1.00 | 0.99 | 459 |
| Left Roll | 1.00 | 0.98 | 0.99 | 458 |
| Right Roll | 1.00 | 0.98 | 0.99 | 459 |
| Pullback | 0.99 | 1.00 | 1.00 | 459 |
| **Accuracy** | 0.99 (2295 Support) | | | |
| **Macro Avg** | 0.99 Precision, 0.99 Recall, 0.99 F1-Score | | | |
| **Weighted Avg** | 0.99 Precision, 0.99 Recall, 0.99 F1-Score | | | |

*4.0.2 Recognition Performance with Dynamic Time Warping.* Next to classical machine learning models, the applicability of dynamic time warping was also explored. The performance of DTW is displayed in Figure 8. The DBA templates were generated using all samples per gesture. Additionally, the classification report for recognition using DTW with the DBA templates is provided in Table 4. For the implementation of DTW, `fastdtw` was used, which provides optimal or near-optimal alignments with an O(N) time and memory complexity [24].

The testing results significantly improve with the use of the DBA-based templates, compared to randomly selecting samples out of the dataset to use as templates. Given that DTW needs two sequences to compare, this approach requires detecting the start and end of anomalies, as opposed to the machine learning approach, which utilized sliding windows. The metrics present great potential for gesture recognition. When deployed in a real-time environment, the DTW approach performs as promising as the testing results do. Practical testing yielded an overall accuracy of 96% across 50 repetitions of each gesture, with misclassification only occuring for the left roll (8 misclassifications as left slip), and the right roll (2 misclassifications as left slip). Classification remained accurate when performing gestures with different speeds, but showed sensitivity to stylistic deviations from the template (e.g. varying head orientations while rolling or drastically different speeds). The DTW approach, compared to the machine learning approach, thus more closely satisfies the research goals posed by research question 1.

Important to note is that the exceptions to the performance are the times when anomaly detection triggers unnecessarily due to sensitive thresholds, resulting in an attempt to recognize a gesture even though the correct gesture is idle. Furthermore, because the entire sequence of data needs to be recognized before attempting to recognize it, predictions can only be made after an anomaly, including the time it takes to recognize the end of an anomaly.

*4.0.3 OpenEarable Framework Integration.* The integration of the boxing head gesture recognition system into the OpenEarable framework consisted of augmenting the dashboard and developing an

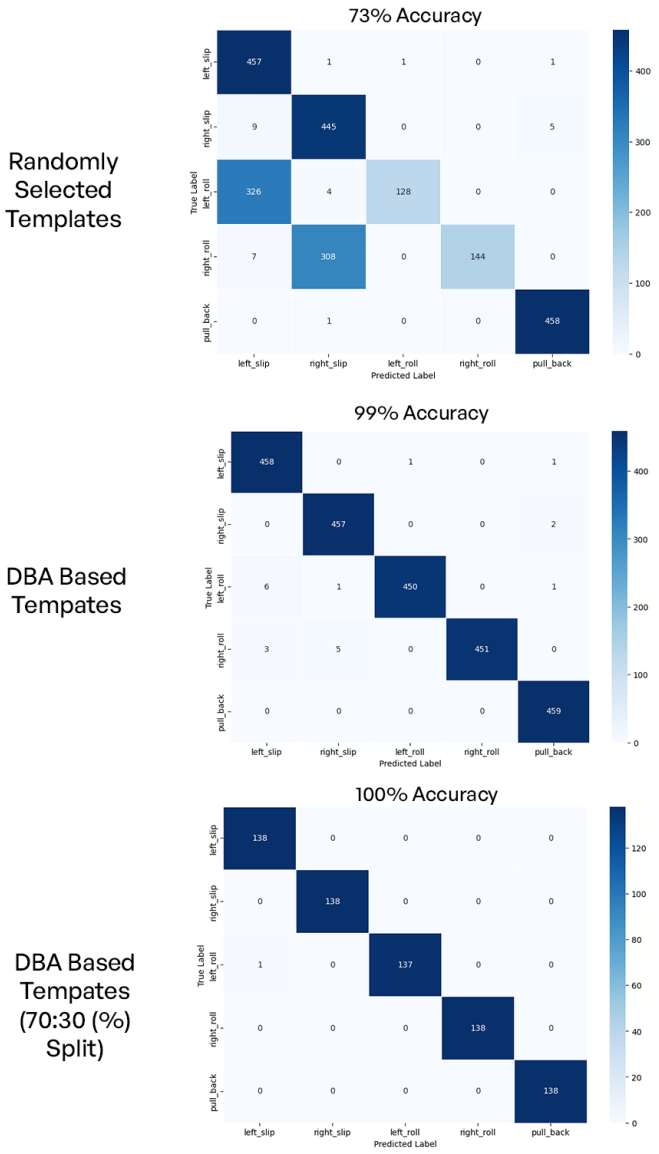



Fig. 8. Dynamic Time Warping - Gesture Recognition Confusion Matrices

inference API. The OpenEarable dashboard was augmented by creating an additional module in the UI, which mirrors the user's behaviour nearly instantaneously. Inspecting a video recording of the augmented OpenEarable dashboard, recorded at 60 frames per second, predictions are approximately made 50 frames after the anomaly ends (ignoring the moment the anomaly has been recognized as ending).

Furthermore, the UI has been designed so that the user is clearly notified of gesture recognition. Figure 6 showcases the recognition of a right roll, including the highlighting of the label indicator, and the gesture mirror performing the gesture. Figure 9 showcases all gesture mirror models.

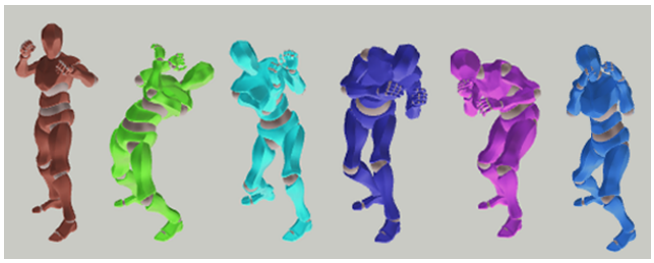


Fig. 9. All model movements featured in the augmented dashboard, mid animation - From left to right: Idle, Left Slip, Right Slip, Left Roll, Right Roll, Pullback

The research goals posed by research question 2 are answered using the developed Python CLI application and the augmented OpenEarable dashboard.

## 5 CONCLUSION

This study explored using the IMU sensor embedded in the OpenEarable device for real-time boxing head gesture recognition. Gestures of interest were the left/right slip, the left/right roll, and the pullback, for which approximately 460 samples were collected from a hobby-level boxer.

Two approaches were explored to achieve effective gesture recognition: classical machine learning and dynamic time warping. While both performed well during theoretical testing, results showed that dynamic time warping, in combination with barycenter averaging-based templates, performed superiorly in real-time deployed contexts. Testing metrics for this approach showed an accuracy of 99%, which was closely matched by its performance in deployed contexts, where performance was not affected by varying speeds or intensities of performed gestures.

The gesture recognition system was integrated into the OpenEarable framework through a Python CLI application and the augmentation of the OpenEarable dashboard, powered by an inference API. Real-time feedback was near-instantaneous, approximately displaying the recognized gesture within a second.

In conclusion, this study offers valuable insights into the potential of earables for head gesture recognition, particularly for sports-related applications such as boxing. Future research into other recognition approaches or further extensions of the OpenEarable framework has the potential to further expand the capabilities of earables and their utilization.

## 6 ACKNOWLEDGEMENTS

The author would like to thank Özlem Durmaz for her guidance and support throughout this research.

During the preparation of this work, the author utilized ChatGPT 4 and GitHub Copilot for assistance with debugging, translating code to different languages, and using various frameworks and libraries. After using these tools and services, all content was thoroughly reviewed and edited as needed. The author takes full responsibility for the final outcome.

## REFERENCES


[1] Sumeyye Agac and Ozlem Durmaz Incel. 2023. User Authentication and Identification on Smart Glasses with Motion Sensors. *SN Computer Science* 4, 6 (Sept. 2023), 761. https://doi.org/10.1007/s42979-023-02202-4

[2] William Cheung and Sudip Vhaduri. 2020. Context-Dependent Implicit Authentication for Wearable Device Users. In *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, London, United Kingdom, 1–7. https://doi.org/10.1109/PIMRC48278.2020.9217224

[3] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W. Kempa-Liehr. 2018. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing* 307 (Sept. 2018), 72–77. https://doi.org/10.1016/j.neucom.2018.03.067

[4] edge-ml developers. 2024. edge-ml: Open source web based machine learning framework for MCUs. https://github.com/edge-ml/edge-ml. Accessed: 2024-06-14.

[5] Germain Forestier, François Petitjean, Hoang Anh Dau, Geoffrey I Webb, and Eamonn Keogh. 2017. Generating synthetic time series to augment sparse datasets. In *Data Mining (ICDM), 2017 IEEE International Conference on*. IEEE, 865–870.

[6] Eibe Frank, Mark A. Hall, and Ian H. Witten. 2016. *The WEKA Workbench* (fourth ed.). Morgan Kaufmann, Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques".

[7] Shkurta Gashi, Aaqib Saeed, Alessandra Vicini, Elena Di Lascio, and Silvia Santini. 2021. Hierarchical Classification and Transfer Learning to Recognize Head Gestures and Facial Expressions Using Earbuds. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. ACM, Montréal QC Canada, 168–176. https://doi.org/10.1145/3462244.3479921

[8] Miguel Grinberg. 2018. *Flask web development: developing web applications with python.* " O'Reilly Media, Inc.".

[9] Alexander Hoelzemann, Henry Odoemelem, and Kristof Van Laerhoven. 2019. Using an in-Ear Wearable to Annotate Activity Data across Multiple Inertial Sensors. In *Proceedings of the 1st International Workshop on Earable Computing*. ACM, London United Kingdom, 14–19. https://doi.org/10.1145/3345615.3361136

[10] Safet Kapo, Said EL Ashker, Anida Kapo, Ekrem Colakhodzic, and Husnija Kajmovic. 2021. Winning and losing performance in boxing competition: a comparative study. *Journal of Physical Education and Sport* 21 (May 2021), 1302–1308. https://doi.org/10.7752/jpes.2021.03165

[11] Soudeh Kasiri-Bidhendi, Clinton Fookes, Stuart Morgan, David T. Martin, and Sridha Sridharan. 2015. Combat sports analytics: Boxing punch classification using overhead depthimagery. In *2015 IEEE International Conference on Image Processing (ICIP)*. 4545–4549. https://doi.org/10.1109/ICIP.2015.7351667

[12] Matias Laporte, Preety Baglat, Shkurta Gashi, Martin Gjoreski, Silvia Santini, and Marc Langheinrich. 2021. Detecting Verbal and Non-Verbal Gestures Using Earables. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. ACM, Virtual USA, 165–170. https://doi.org/10.1145/3460418.3479322

[13] Huaizhou Li and Haiyan Hu. 2024. Head Gesture Recognition Combining Activity Detection and Dynamic Time Warping. *Journal of Imaging* 10, 5 (May 2024), 123. https://doi.org/10.3390/jimaging10050123

[14] Ubeyde Mavus and Volkan Sezer. 2017. Head gesture recognition via dynamic time warping and threshold optimization. In *2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. IEEE, Savannah, GA, USA, 1–7. https://doi.org/10.1109/COGSIMA.2017.7929592

[15] Alessandro Montanari, Andrea Ferlini, Ananta Narayanan Balaji, Cecilia Mascolo, and Fahim Kawsar. 2023. EarSet: A Multi-Modal Dataset for Studying the Impact of Head and Facial Movements on In-Ear PPG Signals. *Scientific Data* 10, 1 (Dec. 2023), 850. https://doi.org/10.1038/s41597-023-02762-3

[16] Nishiki Motokawa, Ami Jinno, Yushi Takayama, Shun Ishii, Anna Yokokubo, and Guillaume Lopez. 2021. Coremoni-WE: Individual Core Training Monitoring and Support System Using an IMU at the Waist and the Ear. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. ACM, Virtual USA, 176–177. https://doi.org/10.1145/3460418.3479325

[17] Nhan Nguyen, Avijoy Chakma, and Nirmalya Roy. 2021. A Scalable and Domain Adaptive Respiratory Symptoms Detection Framework using Earables. In *2021 IEEE International Conference on Big Data (Big Data)*. 5620–5625. https://doi.org/10.1109/BigData52589.2021.9671796

[18] Julien Pansiot, Rachel C. King, Douglas G. McIlwraith, Benny P. L. Lo, and Guang-Zhong Yang. 2008. ClimBSN: Climber performance monitoring with BSN. In *2008 5th International Summer School and Symposium on Medical Devices and Biosensors*. IEEE, Hong Kong, China, 33–36. https://doi.org/10.1109/ISSMDBS.2008.4575009

[19] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, Oct (2011), 2825–2830.

[20] François Petitjean, Germain Forestier, Geoffrey I Webb, Ann E Nicholson, Yanping Chen, and Eamonn Keogh. 2014. Dynamic time warping averaging of time series allows faster and more accurate classification. In *Data Mining (ICDM), 2014 IEEE International Conference on*. IEEE, 470–479.

[21] François Petitjean, Alain Ketterlin, and Pierre Gançarski. 2011. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition* 44, 3 (2011), 678–693.

[22] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. 2022. Sensing with Earables: A Systematic Literature Review and Taxonomy of Phenomena. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (Sept. 2022), 135:1–135:57. https://doi.org/10.1145/3550314

[23] Tobias Röddiger, Tobias King, Dylan Ray Roodt, Christopher Clarke, and Michael Beigl. 2022. OpenEarable: Open Hardware Earable Sensing Platform. In *Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Cambridge United Kingdom, 246–251. https://doi.org/10.1145/3544793.3563415

[24] Stan Salvador and Philip Chan. 2007. FastDTW: Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis* 11, 5 (2007), 561–580.

[25] David Strömbäck, Sangxia Huang, and Valentin Radu. 2020. MM-Fit: Multimodal Deep Learning for Automatic Exercise Logging across Sensing Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (Dec. 2020), 1–22. https://doi.org/10.1145/3432701

[26] Javier Vales-Alonso, Francisco Javier González-Castaño, Pablo López-Matencio, and Felipe Gil-Castiñeira. 2023. A Nonsupervised Learning Approach for Automatic Characterization of Short-Distance Boxing Training. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 53, 11 (Nov. 2023), 7038–7052. https://doi.org/10.1109/TSMC.2023.3292146 Conference Name: IEEE Transactions on Systems, Man, and Cybernetics: Systems.

[27] Dhruv Verma, Sejal Bhalla, Dhruv Sahnan, Jainendra Shukla, and Aman Parnami. 2021. ExpressEar: Sensing Fine-Grained Facial Expressions with Earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (Sept. 2021), 1–28. https://doi.org/10.1145/3478085

[28] Cheng-Wei Wu, Hua-Zhi Yang, Yan-Ann Chen, Bajo Ensa, Yi Ren, and Yu-Chee Tseng. 2017. Applying machine learning to head gesture recognition using wearables. In *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*. IEEE, Taichung, 436–440. https://doi.org/10.1109/ICAwST.2017.8256495

[29] Songlin Xu, Guanjie Wang, Ziyuan Fang, Guangwei Zhang, Guangzhu Shang, Rongde Lu, and Liqun He. 2022. HeadText: Exploring Hands-free Text Entry using Head Gestures by Motion Sensing on a Smart Earpiece. http://arxiv.org/abs/2205.09978 arXiv:2205.09978 [cs].