BoxerSense: Punch Detection and Classification Using an IMU

YOSHINORI HANADA[1] ANNA YOKOKUBO[1] GUILLAUME LOPEZ[1]

Abstract: Maintaining healthy living requires habitual physical activities. Nonetheless, staying motivated to work out regularly is challenging for most people. To solve this problem, automated personal supporting systems could help. This paper presents boxercise, a fitness standard exercise that mainly includes shadow-boxing exercises. The paper introduces punch activity detection and classification methods using acceleration and angular velocity signals recorded using a single smartwatch on the participant's rear hand wrist. The proposed method is evaluated on our 10 participants aged between 17 and 53 years old (8 male and 2 female, age 27.8 ± 12.8). As a result, we achieved 98.8% detection accuracy, 98.9% classification accuracy with SVM in-person-dependent (PD) case, and 91.1% classification accuracy with SVM in person-independent (PI) case. In addition, we estimated the real-time performance of each classification method and found out all our methods could classify a single punch in less than 0.1 seconds. The paper also discussed some points of improvement towards a practical boxercise supporting system.

1. Introduction

Encouraging people to perform physical exercise regularly plays a key role in maintaining our health and quality of life. In fact, Hammer et al. found out that frequent physical exercise in a week reduce the risk of psychological distress [1]. However, in practice, maintaining a regular physical exercise a lifelong habit is challenging [2], [3]. As a result many people failed to maintain the recommended levels of exercise [4]. Some hindrances include location and time constraints, a lack of knowledge on appropriate exercise intensity, and poor performance and motivation due to monotony and fatigue. To solve these issues, wearable technologies can help. In recent years, the number of wearable devices available at a reasonable cost have increased including smartphones and smartwatches. This promoted the development of applications for tracking exercise to support people's health.The most used wearable device for people today is without a doubt smartphones. The smartwatch is also a wearable computing device that runs various applications to tracking exercise. In recent years, due to the growth in technologies such as the long-lasting rechargeable battery, high-performance central unit (CPU) and graphic processing unit (GPU), it has become possible to embed high-performance computers in watches making the demand of smartwatches increased. The smartwatch is also embedded with sensors that provide fitness or healthcare-related functionality, e.g., exercise tracking such as swimming, running, and cycling. Although the state of the art exercise tracking applications in smartwatches and smartphones and other wearable technologies can track exercise intensity from pedometer and heart rate monitor, there are only a limited types of exercise motion that smartwatch applications can detect, classify the activities.

In this paper, we focus on boxercise which is based on the training concepts boxers use to keep fit and propose punch detection and classification methods using acceleration and angular velocity signal recorded using a single smartwatch on participant rear hand wrist. As a result, we achieved 98.8% detection accuracy, 98.9% classification accuracy with SVM in person-dependent (PD) case and 91.1% classification accuracy with SVM in person-independent (PI) case.

2. Related Work

The most relevant previous work for punch recognition is that of Ovalle et al.[5], who classified 4 different taekwondo punches from IMU sensors attached to a right-hand wrist and a microphone. Their purpose of the research was to investigate if it is possible to recognize punches with bare hands and increase the recognition rate by adding audio input that is produced by hitting the mitt. They achieve 94.4% accuracy when using only the IMU sensor. However, the audio signal did not improve its performance. Even though they achieved high accuracy, they only had 3 participants and did not investigate person dependent case. Therefore, their research lacks credibility in punch recognition accuracy. They also used wired communication to transfer the data, which makes the system difficult to apply to real-world applications.

For the Boxercise related work, a Spanish start-up company VOLAVA [6] released Fitness Boxing kit in 2020 which includes a punching bag, gloves, an exercise mat, and an IMU sensor kit. The company made sensors that connect to the Volava boxing mobile app to analyze real-time data

1 Aoyama Gakuin University, Japan

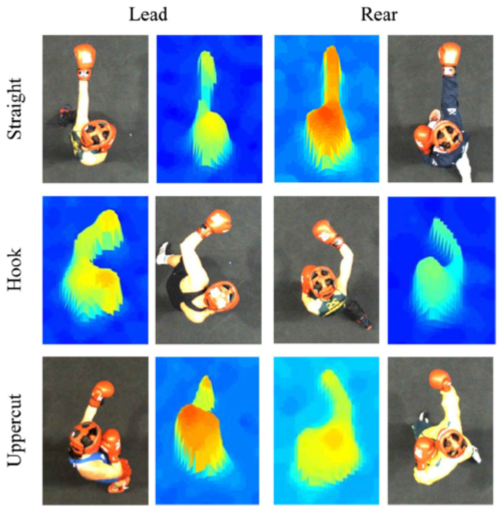

Fig. 1 The six basic boxing punches captured using overhead depth and visible cameras. This includes the straight, hook, and uppercut punches thrown from both the lead and rear hands (quoted from [8])

such as, number of punches, punch force, calories, and heart rate in order to connect sensor metrics with social interaction leaderboard. Their kit, however, includes 3 IMU sensors and requires boxing pieces of equipment which may be too expensive for some people. Their system also did not recognize the type of punches which may be significant information to evaluate the punching technique when creating mobile personal trainers in the future.

Another Boxercise related work is an exergame, Fitness Boxing [7] released by Nintendo Switch in 2018. They utilized Nintendo's Joy-Con motion Controllers for a player to perform punching and dodging maneuvers. In their game, players can personalize the workouts sessions by setting up their fitness goals. By making progress in the game, hit songs for background music and new personal trainers can be gradually unlocked. They can also estimate daily calories burn so that players can track their progress. However, the game only recognizes shadow-boxing punches when the controllers detected a certain motion threshold, and the type of the punch is not identified.

To sum up, the remaining problems of existing literature are poor credibility of punch recognition accuracy, unrealistic communication method for punch recognition when applying real-world application, none of real-world application for boxercise identifying the type of punch. To solve these problems, our research utilized a single IMU sensor embedded in a smartwatch to achieve wireless communication and investigated if it is possible to recognize shadow-boxing movements and achieve high accuracy by testing person dependent cases with 10 participants.

3. Proposed Methods to Detect and Classify Punches

3.1 Overview

The end goal of this research is to build a sensor-based boxercise personal supporting system that can provide instruction, feedback, and gamified experience from boxercise motion and heart rate monitor. To achieve this, we first focused on the detection and recognition of boxercise movements in both real-time and non-real-time.

3.2 Target Activities

Although there are various types of exercise done in boxercise classes, shadow-boxing usually plays an important role for the class. Therefore, we focused on shadow-boxing punches first. According to Kasiri et al.[8] who recognized 6 different punches by using depth image, there are two type of boxer's stance (orthodox or southpaw) and 6 basic type of shadow-boxing punches shown in Figure 1. They includes straight, hook, uppercut for both lead and rear hands respectively. In this paper, we focused on recognizing by rear hand punches with a single smartwatch only assuming that it is too difficult to recognize both hands punches from a single smartwatch on one hand. In our future work, we will test this hypothesis as well as the best sensor locations and numbers.

3.3 Data Collection

We chose an IMU sensor embedded in smartwatch and developed a wear OS application to collect acceleration and angular velocity of shadow-boxing punch from a smartwatch worn on a rear hand wrist. When a user starts the application the user is asked to type their ID. After typing ID, the start button can be pressed whenever they are ready to start recording the sensor values. By clicking the start button, the button text changes to stop. The user can press the stop button to stop recording and send sensor data as a CSV file to the computer via Bluetooth connection. The stop button disappears after it is pressed, and the start button appears when the data transfer completes. We collected 924 punches from 10 participants. Their age range from 17 to 53 years (8 male and 2 female, age 27.8 ± 12.8). Note that in this research, we decided to choose the wrist as a position of the IMU sensor because we assumed that it is the best position to collect punch motion data. We are planning to test this hypothesis in future work.

3.4 Proposed Method for Activity Detection

To detect the punching activity, the process of detection is shown in Figure 2. The algorithm starts with a 3D acceleration signal collected from an accelerometer. Since boxing punches contains both longitudinal and transverse motion, we calculated the synthetic acceleration, which is the norm of the 3D signal. As shown in the second plot of Figure 2, the calculated norm contains noise. Thus, a low-pass filter with a smoothing coefficient of 0.9 is applied to smooth and emphasize the rapid change of punching motion. Finally, segmentation algorithms are applied to the preprocessed data.

Two kinds of thresholds are defined for segmentation shown as the horizontal line of the third plot in Figure 2. The lower threshold is set for detecting the starting point and ending point of an event. We set the value to 9.8 to ignore acceleration caused by gravity. The upper threshold is set to detect the rapid sensor motion. In this paper, we defined that any values above the threshold are a punch. We set the threshold to 20 by testing the value which can completely segment all of the punches from several sessions which is performed by the author of this paper.

The segmentation algorithm starts with detecting the starting point every time preprocessed sensor value exceeds the lower threshold and save the index of the data. After detecting the starting point, the algorithm looks for the sensor value to exceeds the upper threshold. If it did exceeds it, we set boolean value true, representing that upcoming sensor data is produced by a punch motion. If the boolean value is true and the preprocessed sensor value becomes lower than the lower threshold, the data point will be recognized as an endpoint of the punch, and the data between the saved starting point and endpoint is segmented as a punch. If the boolean value is false and the preprocessed sensor value becomes lower than the lower threshold, the starting point is simply deleted from memory which implies no punching event is detected. After segmenting a punch, the saved starting point and endpoint are also deleted from memory.

3.5 Activity Classification

After detecting and segmenting a punch, we tested using correlation-based approach and machine learning approach to classify the activities. For correlation-based method, we used The Dynamic Time Warping algorithm. The Dynamic Time Warping (DTW) calculates the distance between two time-series data that have different lengths. In other words, this algorithm can handle the shape of signals from the identical activity performed at different speeds. To use the DTW algorithm, we collected templates of all targeted activities. For each activity, one of the detected segments of punching motion is randomly selected and extracted its raw 3D acceleration and 3D angular velocity data as a template. Then, we calculated the DTW distance between a extracted signal and template signals of every type of activities of an axis. To know which axis is the best to use for classification we compared the result of using each axes for both angler velocity signal and acceleration signal. Finally, we classify the activity that has minimum DTW distance as the activity. The fact that the algorithm only requires one or a few times of exercises to collect template data makes the data collection process to classify activity makes easier than the machine learning approach which requires a vast amount of activity data to classify with high accuracy.

For machine learning approach, we extracted statistical features from both raw 3D acceleration and 3D angular velocity for each ax within the segment. The features we ex-

Table 1 Extracted Features for each ax of raw 3-D acceleration and 3-D angular velocity

Type of features
mean
median
standard deviation
min
max
25% percentile
75% percentile

tracted are shown in Table 1. We extracted 7 features for each ax, mean, median, standard deviation, min, max, 25% percentile, and 75% percentile in a total of 42 features for each segment. We also labeled each segment as a corresponding punch type to apply supervised machine learning models to classify the type of punches. In this work, we compared three types of classifiers, multi-class Support Vector Machine (SVM), Random Forest (RF), K Nearest Neighbors (KNN) from scikit-learn a machine learning library for python. Since this is only preliminarily researched, we did not optimize hyperparameters, and we chose to set all of the parameters of machine learning models to default values of scikit-learn. Before training data with classifiers we chose, for RF and KNN classifiers which calculate distances between different points in their algorithm, we normalized extracted features between [0; 1] to maintain proportional distances.

Since we are assuming to provide real-time feedback when

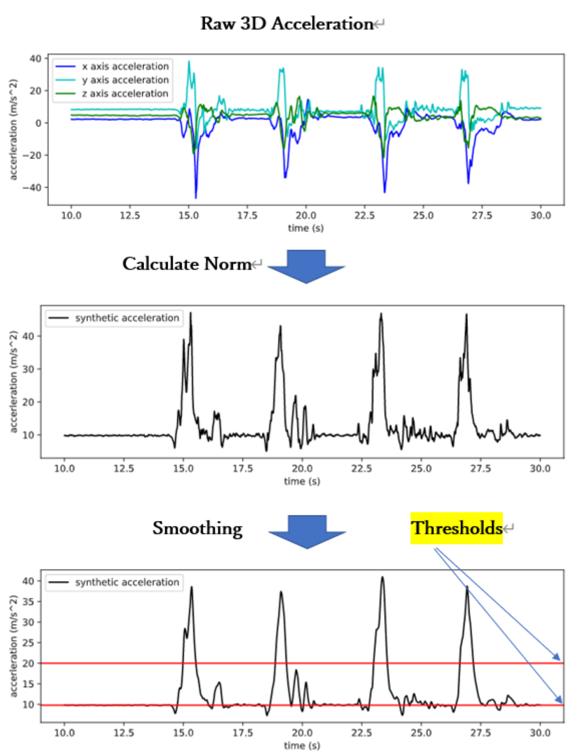

Fig. 2 Example of the processing flow of punch detection applied to 3-D acceleration signal collected during performance of 4 consecutive rear hooks

we develop boxercise supporting system, we estimated the time needs to classify a single punch with each classification methods. For the estimation, we first randomly chose 31 rear straight punches performed by one participant in the experiment and classified all of the punches by using all of the classification method we proposed(DTW, RF, SVM and KNN). Then, we calculated the average of the time required to classify a single punch for each methods. Note that in our experiment, we used a computer with an Intel(R) core(TM) i5-8250U processor running at 1.60 GHz using 8 GB of RAM, running Windows 10 and a programming language of python version 3.7.6.

4. Validation of Proposed Method

4.1 Experiment

We tested the performance of the proposed method on 10 participants using Polar M600 smartwatch which has a gyroscope and accelerometer with 100Hz sampling frequency embedded inside the watch. The participants' ages range from 17 to 53 years (8 male and 2 female, age 27.8 ± 12.8) and include 3 martial art experienced and 7 inexperienced persons. The first experienced person had only attended a few kickboxing classes but had been training by himself for 6 months by watching online lessons. The second experienced person had attended kickboxing classes for few times. The third experienced person has experienced two years of karate.

As we mention in the target activity section which is the subsection of the proposed method, we targeted three types of rear hand punches thrown from orthodox position: 1) rear hand straight, 2) rear hand hook, 3) rear hand uppercut. These three types of punches are collected and labeled through an experiment for each person. The experiment starts with asking the participants whether he/she had learned how to throw a boxing punch before. If they are not experienced, 5 to 10 minutes of punching lesson is taught by an experienced boxer. Then, the participant is asked to wear a smartwatch on the rear hand (right hand) wrist and activate the data collection application we developed which we mentioned in the proposed method section. In a single session, a participant is asked to perform the same type of punch 30 times with 4 seconds interval. Each participant performed 3 sessions in total with different types of punch in each session. As a result, we aimed to create a data set of 900 punches. To notify a participant when to throw a punch we downloaded interval timer application from the iOS app store, which can ring in each interval set by a user. While the participant is performing an experiment, we recorded participants' movements to evaluate the detection algorithm.

4.2 Detection Result

Although we designed the experiment to perform 900 punches, some of the participants accidentally performed extra one punches at the end of the experiment by reacting to the ending bell of the interval application resulting in a total of 924 punches. Out of 924 (307 rear hand straights, 308 rear hand hooks, 309 rear hand uppercut) detected punches, 913 punches were actual punches meaning that the overall accuracy was 98.8%. We had 1 false negative (predicted that the segment is not punched but actually it is) and we had 11 false positives (predicted that the segment is punch but actually it is not). Therefore, the precision of all punches detection was 98.8% and the recall of all punches was 99.9%. The accuracy, precision, and recall of each punch are shown in Table 2.

Table 2 The detection result for each punch type

	Straight	Hook	Uppercut	Total Punches
accuracy	99.0%	98.7%	98.7%	98.8%
precision	100.0%	100.0%	98.4%	98.8%
recall	99.0%	99.0%	99.7%	99.9%

The result of detection showed that it is possible to detect punches with high accuracy with the proposed method. Many of the false positives were caused by the participant's arm lowering movement when the experiment is done and they are asked to press the stop button of the data collection application. Although we set the upper threshold in the third plot of Figure 2 to 20 for this experiment, we found out that if we set the threshold to 30, all of the false positives are eliminated and the accuracy rises to 99.9%. Therefore, we still have rooms to improve detecting accuracy by optimizing the value of the threshold in future work.

4.3 Classification Result

Table 3 shows the results of classification accuracy of all axes for acceleration signal and angular velocity signal resulting from the correlation-based method (DTW). As shown in the table, z-axis acceleration scored the best accuracy followed by x-axis angular velocity. To evaluated the method in detail, we use indices of accuracy, precision, recall, and F1-score for each labels with these two best and second best performing signals. The DTW with z-axis acceleration signal achieved the accuracy of 72.8% and F1-score of 71.2%. On the other hand, the DTW with x-axis angular velocity signal achieved the accuracy of 68.4% and F1-score of 66.6%.

Table 3 Classification accuracy of each axes with acceleration and angular velocity signal resulting from the DTW

Axis	Acceleration	Angular Velocity
X-axis	47.6%	68.4%
Y-axis	56.0%	37.4%
Z-axis	72.8%	56.5%

We evaluated the three types of classifiers (SVM, KNN, RF) by two cases person-dependent case (PD) and person-independent case (PI) for machine learning approach. In the PD case, we conducted 10 fold validation which shuffles the data randomly in each holds. This splits 25% of the data to testing data and 75% of the data to training data. The training data is used to train classifiers, and the trained classifiers are used to predict against the testing data. This process of splitting data and predicting is repeated 10 times and average accuracy is calculated. As shown in Table 4, we achieved 97.8% of f1-score (97.8% of accuracy) with RF, 99.0% of f1-score (99.0% of accuracy) with SVM, 98.5% of f1-score (98.5% of accuracy) with KNN. Therefore, in the PD case, we confirmed the SVM is the best classifier. The confusion matrix of the best classifier is shown in Figure 3.

In the PI case, we conducted leave-one-person-out cross-validation, wherein each fold, 9 participants are used for training, and the remaining one participant was used for testing. After calculating classification accuracy for each person. The average classification accuracy of 10 participants was calculated. As shown in Table 5, we achieved 85.9% f1-score (86.1% accuracy) for RF, 90.6% f1-score (91.1% accuracy) for SVM, and 85.8 f1-score (86.3% accuracy) for KNN. Therefore, in PI case, we confirmed that the best classifier was the SVM. The confusion matrix of the best classifier is shown in Figure 4. The confusion matrix shows that the rear hook has the least true positive rate. It also shows the most confusion occurs when the model accidentally predicted the uppercut but the true label was the rear hook.

The result of the PD case showed that it is possible to classify the 3 types of rear hand punches with high accuracy close to 100% in PD cases. The result of the PI case showed that it is possible to classify the punches slightly over 90% despite a variety of the participants' age and gender. In this work, we did not perform hyperparameter optimization and feature selection, therefore the accuracy of both PD and PI case still have a lot of rooms to improve.

Table 4 Classification results of each models in PD case

Method	Activity	Precision	Recall	F1-score
RF	Rear Straight	97.5%	99.7%	98.6%
	Rear Hook	97.7%	96.4%	97.1%
	Rear Upper	98.4%	97.4%	97.9%
	Macro Average	97.8%	97.8%	97.8%
KNN	Rear Straight	99.0%	99.4%	99.2%
	Rear Hook	98.1%	98.1%	98.2%
	Rear Upper	98.4%	98.1%	98.2%
	Macro Average	98.5%	98.5%	98.5%
SVM	Rear Straight	99.4%	99.7%	99.5%
	Rear Hook	98.7%	98.4%	98.5%
	Rear Upper	99.9%	99.0%	99.0%
	Macro Average	99.0%	99.0%	99.0%

Table 5 Classification results of each models in PI case

Method	Activity	Precision	Recall	F1-score
RF	Rear Straight	92.6%	91.1%	89.6%
	Rear Hook	84.6%	81.1%	79.0%
	Rear Upper	93.1%	89.3%	88.9%
	Macro Average	90.1%	87.1%	85.9%
KNN	Rear Straight	91.0%	93.4%	90.3%
	Rear Hook	85.6%	79.0%	81.6%
	Rear Upper	88.6%	86.5%	85.6%
	Macro Average	88.4%	86.3%	85.8%
SVM	Rear Straight	96.5%	93.1%	93.4%
	Rear Hook	91.4%	84.8%	85.3%
	Rear Upper	92.2%	95.5%	92.9%
	Macro Average	93.4%	91.1%	90.6%

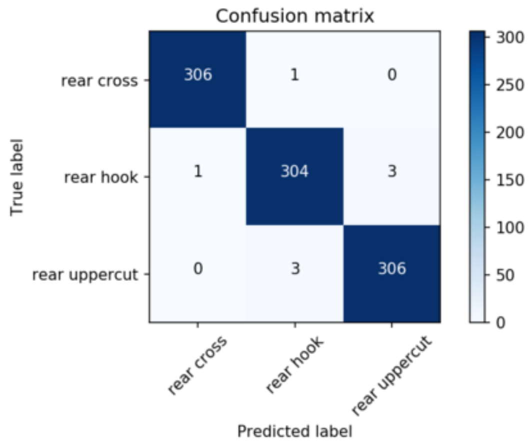

Fig. 3 The confusion matrix of the best classifier (SVM) in person dependent case

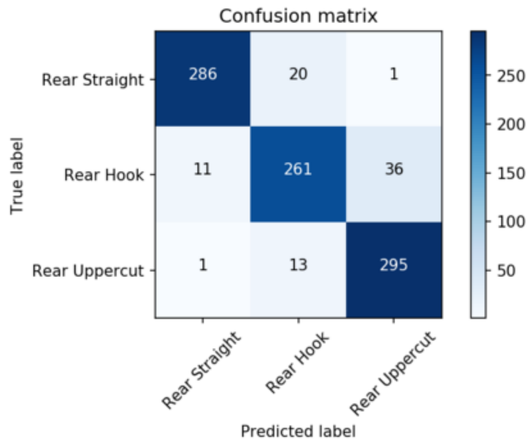

Fig. 4 The confusion matrix of the best classifier (SVM) in person independent case

4.4 Estimated result of real-time classification performance

As shown in table 6, we compared the result of the estimated classification speed performed by each method that we proposed. As a result, all of the method was able to classify a single punch in less than 0.01 seconds and SVM was the fastest method to classify with the record of 0.0107 seconds. The slowest method was DTW and it recorded 0.0992 seconds which is about 0.09 seconds slower than the other machine learning method. This result implies that machine learning perform better at real-time classification performance than correlation-based method(DTW).

Table 6 Estimated real-time classification speed

Rank	Methods	Classification Time
1	SVM	0.0107 seconds
2	RF	0.0114 seconds
3	KNN	0.0120 seconds
4	DTW	0.0992 seconds

5. Conclusion and Future Work

In this paper, we focused on boxercise and proposed punch activity detection and classification methods using acceleration and angular velocity signal obtained by a single smartwatch on participant rear hand wrist. In addition, to develop our research into a boxercise personal supporting system in the future work, we estimated the real-time performance of classification methods and discussed which methods to use for certain types of feedback. For detection of a punch, we achieved the overall accuracy of 98.8%. For classification, with machine learning approach, we achieved 98.9% accuracy with SVM in the PD case, and we achieved 91.1% accuracy with SVM in the PI case. For correlation-based, we used the DTW and achieved accuracy of 72.8% with z-axis acceleration signal. For the result of estimation of real-time performance of classification methods, all of our proposed classification methods could classify a single punch in less than 0.01 seconds.

In the future, we are planning to detect and classify boxercise movements in real-time to give feedback to the user from the sensor values obtained by the smartwatch. To achieve this, we will first investigate the best position of a single sensor and if it is possible to detect and classify more boxercise movements including lead hand punches and dodging maneuvers and improve the accuracy of the methods in non-real-time. We will also work on optimization of SVM and RF hyperparameter and optimization of the DTW template selection. Then, we will build a system that can run the methods in real-time and test their real-time validity by conducting actual trial. Furthermore, we are aiming to build the sensor-based boxercise personal supporting system which includes the features of providing instruction, feedback and gamified experience to the users in real-time to support and promote people to exercise regularly.

References

[1] Mark Hamer, Emmanual Stamatakis, and Andrew Steptoe. Dose-response relationship between physical activity and mental health: the Scottish Health Survey. *British journal of sports medicine*, 43(14):1111–1114, 2009.

[2] Judy Kruger, Heidi Michels Blanck, and Cathleen Gillespie. Dietary and physical activity behaviors among adults successful at weight loss maintenance. *International Journal of Behavioral Nutrition and Physical Activity*, 3(1):17, 2006.

[3] Karen A Schutzer and B Sue Graves. Barriers and motivations to exercise in older adults. *Preventive medicine*, 39(5):1056–1061, 2004.

[4] Carmen D Harris, Kathleen B Watson, Susan A Carlson, Janet E Fulton, Joan M Dorn, and Laurie Elam-Evans. Adult participation in aerobic and muscle-strengthening physical activities—United States, 2011. *MMWR. Morbidity and mortality weekly report*, 62(17):326, 2013.

[5] Juan Quintero Ovalle, Katarzyna Stawarz, and Asier Marzo. Exploring the addition of audio input to wearable punch recognition. In *Proceedings of the XX International Conference on Human Computer Interaction*, pages 1–4, 2019.

[6] VOLAVA FitBoxing Kit Brings Studio Style Fitness Boxing to Home. https://www.movesense.com/news/2020/01/volava-fitboxing-kit-brings-studio-style-fitness-boxing-to-home/.

[7] Fitness Boxing. https://www.nintendo.com/games/detail/fitness-boxing-switch/.

[8] Soudeh Kasiri, Clinton Fookes, Sridha Sridharan, and Stuart Morgan. Fine-grained action recognition of boxing punches from depth imagery. *Computer Vision and Image Understanding*, 159:143–153, 2017.