Using accelerometer and gyroscope sensors to detect and differentiate between eating movements associated with different foods

Viktor Tonchev, University of Twente, The Netherlands

Dietary monitoring is a tool used to track and alter the eating habits of individuals, specifically those that are overweight or obese. Automated Dietary Monitoring (ADM) aims to automate this process, in order to make it more accurate and efficient. Existing ADM systems have used various sensors worn on the body to detect eating events, such as gestures, chewing and swallowing, via machine learning algorithms. This research set up an experiment, where an IMU (Inertial Measurement Unit) with accelerometer and gyroscope sensors was worn on the wrist and participants consumed 7 different types of foods. The data from the experiment was used to train classification algorithms using machine learning in an attempt to differentiate individual foods based on the movements recorded by the sensors. The models are effective at recognizing when food is being eat, but the results suggest that they are not sufficient on their own to recognize the food itself.

Additional Key Words and Phrases: Machine Learning, Automatic Dietary Monitoring, Eating habits, Eating detection, Accelerometer, Gyroscope, LSTM.

1 INTRODUCTION

Overweight and obesity, defined as having a BMI over 25 and 30 respectively, are a problem all around the world, with the rate of both rising over recent years [1, 8]. Obesity has been associated with various health risks and overall mortality [2, 13]. Short sleep duration, low dietary calcium intake and high disinhibition eating behavior are all risk factors significantly associated with obesity [4] and the change of such habits is recognized to be an effective measurement to combat obesity [16]. Therefore, keeping track of what and how much one eats plays an important role in weight loss. This can be done manually by keeping a food intake diary either physically or by means of a phone application. The problem with this method is that people have been observed to inaccurately report their intake when recording it manually by over- or underestimating it or just forgetting to do it at all [14]. This is where ADM comes into play, striving to make the process both more efficient and accurate by automating it.

This research will be using an IMU with accelerometer and gyroscope sensors worn on the wrist to identify movements during eating and try to differentiate between the foods being consumed. The sensor is a good choice as it is comfortable, accessible, easy to use, and similar setups have been successfully used to identify eating gestures. In [17] a machine learning framework using a 6-axis inertial wrist-worn censor is suggested, which showed reliable classification of feeding gestures (75% F-measure) and a 94% accuracy of feeding gesture count in an unstructured eating experiment. Dong, et al. used accelerometer and gyroscope sensors on the wrist to detect motion indicating eating activities in a free-living environment. They achieved an accuracy of 81% with a 1-s resolution [5]. Kyritsis et al. detect food intake cycles during a meal using an IMU achieving precision and recall scores of 0.78 and 0.77 respectively [11]. Sen et al. used the accelerometer and gyroscope in a smartwatch to detect eating episodes and mode of eating (hands, chopsticks, or spoon) and then try to recognize the food by using the smartwatch's camera [15]. They detected eating episodes with an accuracy of 97% and eating mode with an accuracy of 85.51%. In their thesis, Mevissen set up an experiment, where a sensor system, consisting of a smartwatch, a piezoelectric sensor and respiratory inductance plethysmography bands, is used for the detection of eating gestures, chewing, and swallowing food respectively [12]. The highest F1-score achieved by the algorithms were 0.82 for the classification of eating gestures (including telling apart eating yogurt and eating a croissant), 0.94 for chewing food and 0.53 for swallowing food.

The difference between this and previous research is that the focus is on differentiating between multiple foods being consumed, instead of just detecting eating movements or periods of eating.

1.1 Problem Statement

Existing research has used accelerometer and gyroscope sensors to detect eating gestures, movement and periods of eating, but none try to differentiate the foods themselves with these sensors alone. This paper aims to expand on previous work, by testing the effectiveness of these 2 sensors further.

1.1.1 Research Question

The problem statement leads us to the following research question:

To what extent can a sensor system, consisting of an accelerometer and gyroscope worn on the wrist, detect and differentiate eating movements for various types of food?

To answer this main question, we have the following sub-questions:

- Can a sensor system, consisting of an accelerometer and gyroscope worn on the wrist effectively differentiate between eating and non-eating movements for various types of food?

TScIT 37, July 8, 2022, Enschede, The Netherlands

•To what extent can a sensor system, consisting of an accelerometer and gyroscope worn on the wrist, differentiate between eating movements associated with various types of food?

To answer this, an experiment was set up, where participants consumed different types of food, while sitting and wearing the IMU device with the accelerometer and gyroscope sensor. The data is then used to create machine learning algorithms to recognize when food is being consumed and to differentiate the type of food consumed.

2 METHODOLOGIES

2.1 Experiment

An experiment was conducted in order to collect data with the sensor to use in the classification algorithms. Participants were asked to consume various types of food, while a video was recorded to establish ground truth. Ethical consent to perform the experiment was granted by the Ethics Committee Computer & Information Science of the University of Twente.

2.1.1 Setup

The sensors were both set to 50Hz and placed on the wrist of the right hand of the subject (all the participants ate with the right hand). Then the subjects were asked to sit at a table, while being recorded by a camera, and consume 7 different types of food:

- Yogurt in a bowl
- Cereal in a bowl
- Piece of bread with hummus on it
- Small croissant
- Grapes
- Pieces of corn in a bowl
- Sliced pieces of cucumber

The idea was to have multiple foods that are eaten in a similar manner to test whether they can be distinguished from one another (the corn and cucumber are there because the experiment was done in collaboration with another paper, which needed the data from them). For example, croissant and bread are both consumed with the hand and have similar movements. Same goes for eating yogurt and cereal with a spoon. Participants were also asked to perform several movements to act as counter-gestures to eating. They were as follows:

- Random movement of the right arm
- Scratching the right cheek with the right hand
- Scratching under the chin with the right hand
- Scratching the back of the head with the right hand

Aside from the random movement, which is there to provide data on movements outside of the experiment, the other gestures were chosen to resemble eating movements, to check how well the algorithm can distinguish them from actual eating. These actions were done in alteration, with participants being asked to perform the counter-gestures in between eating. The full protocol for the experiments can be found in appendix A.

2.1.2 Annotations

The video recordings made during the experiments were used to establish the ground truth to be used for the algorithms. All the different foods have their own label, the random movement and counter-gestures are all under the label of “Other” and everything in between is under the label "Leftover". The annotations were also used to synchronize the data from the sensor with the data from the video. This was done by tapping the sensor 5 times at the beginning of each experiment. The software used for this was ELAN 6.3 [7].

2.2 Data Processing

After the experiments were conducted, the data had to be processed in order to be used for the algorithms. The experiment was conducted with 8 subjects, which amounted to almost 2 hours of video and data. First the data was plotted to figure out when the 5 taps needed to synchronize with the annotations occurred. Then those times were used to synchronize the annotations and the data from the sensors, by figuring out the difference between the two and subtracting it from the annotations to match the sensor data. After that the data was split into windows of 1 second, with an overlap of 50%. When splitting the data, only a small portion of the “Leftover” data was used, as it does not provide that much information and this way the models takes less time to train. Finally, using the annotations, the ground truth was established and recorded for every time window.

2.3 Classification

The final step is to use the collected and processed data to train models for activity recognition. Neural Networks have been successfully used in Human Activity Recognition before, with high accuracy and F1 scores [9]. The best choice for this research is the LSTM (Long Short-Term Memory) network model. This is because this model can learn directly from the data itself and does not require the manual engineering of features, which requires expertise beyond the scope of this paper. The model is also appropriate for long sequences of data, can support multiple parallel sequences of data (such as the x, y and z axes of the accelerometer and gyroscope sensors) and can extract features from the sequences of this data. All of these suggest it to be a good fit for the data. The model was based on the one in [3].

In total 4 models were built: 2 to answer the first sub-question and 2 to answer the second one. The reason for building 2 models for each question is to check how the classification does when it has trained on a subject’s data already compared to when the data is completely new to it. In other words, this helps show the generalizability of the model. The 4 models built are as follows:

1. Model trained and tested on all of the data (we will call this a generic model) to differentiate between eating and non-eating movements
2. Generic model trained to differentiate between movements associated with the different foods
3. Model trained on data from all but one subject and tested on the subject left out, also known as LOSO (Leave One Subject Out) models, to differentiate between eating and non-eating movements

4. LOSO model to differentiate between movements associated with the different foods

The classes for models 1 and 3 are simply “Food” and “Other”. “Food” includes all eating movements, while “Other” contains everything else. The classes for models 2 and 4 are one individual class for every food and one “Other” class that contains everything else.

The train/test splitting for the generic model is done as follows: to make sure there is no data leakage (or at least minimal), due to the 50% overlap in the time windows, every window is assigned an identifying number. Then the data is shuffled, and the assigning of training and testing data starts – there is an 80% chance for a window to be assigned to training and 20% to testing. However, before the random chance, it is checked whether the previous or next window is labeled as testing or training and if it is, this window is labeled as the same, to avoid leakage. Of course, there have to be points in the data where the training and testing data meet. However, with the windows being 1 second, the overlap there is half a second, which compared to the total amount of data is almost insignificant. Aside from this, to ensure all foods have movements in both training and testing, a counter with the total movements in training and testing is kept for each subject, to ensure that at least a minimal number of movements are in each set. Additionally, when evaluating these models, they are run multiple times with different shuffling, to act as a kind of cross-validation.

Since the classification is done using LSTMs, which are stochastic in nature, they give slightly different results when run multiple times, so the models were evaluated by being run 10 times each (more repetitions were not possible due to time constraints) and taking the mean accuracy and F1-score, as well as the standard deviation. During evaluation, the weight of each class is also considered. This is done by counting the instances of each class, comparing it to the other classes and taking it into account when calculating the evaluation scores.

3 RESULTS

Table 1 shows the results for all of the models.

The class names were shortened because of the size of the matrix (otherwise they would overlap each other). Here are what the shortened names stand for:

- Oth – Other class
- Yog – Yogurt class
- Cer – Cereal class
- Br – Bread class
- Cro – Croissant class
- Gr – Grapes class
- Corn – Corn class
- Cuc – Cucumber class

The generic model trained to differentiate only between eating and non-eating movements had accuracy and F1-score both of 0.939, with a standard deviation of 0.018 for both as well.

Table 1. Accuracy and F1-score of the 4 models

Model	Average Accuracy	Average F1-score
Generic for just eating	0.939 (+/- 0.018)	0.939 (+/- 0.018)
Generic for different foods	0.555 (+/- 0.020)	0.555 (+/- 0.018)
LOSO for just eating	0.858 (+/- 0.095)	0.859 (+/ 0.092)
LOSO for different foods	0.402 (+/- 0.051)	0.393 (+/- 0.063)

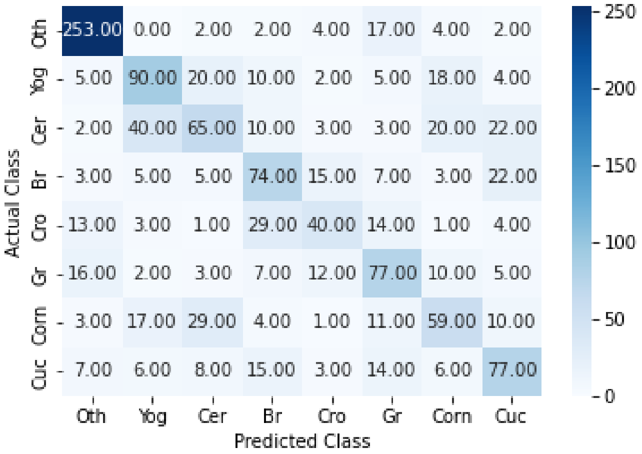


Fig. 1. Confusion matrix of generic model trained to differentiate between the different foods

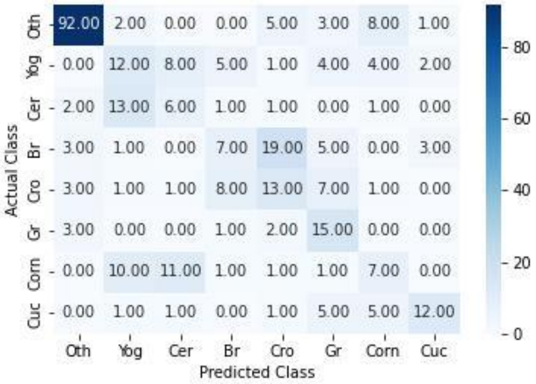


Fig. 2. Confusion matrix of LOSO model trained to differentiate between the different foods

The generic model trained to differentiate between the different foods had an average accuracy of 0.555, with a standard deviation of 0.020, and an average F1-score of 0.555, with a standard deviation of 0.018, Figure 1 shows a confusion matrix for 1 of the instances of this model.

The LOSO model trained to differentiate only between eating and non-eating movements had an average accuracy of 0.858, with a standard deviation of 0.095, and an average F1-score of 0.859, with a standard deviation of 0.092.

The LOSO model trained to differentiate between the different foods had an average accuracy of 0.402, with a standard deviation of 0.051, and an average F1-score of 0.393, with a standard deviation of 0.063. Figure 2 shows a confusion matrix for 1 of the instances of this model.

4 DISCUSSION

The purpose of this research was to figure out whether gyroscope and accelerometer sensors worn on the wrist are sufficient on their own to differentiate between the eating movements associated with different foods.

4.1 Differentiating eating and non-eating movements

Looking at the results, the models had no problem differentiating between eating and non-eating movements. The generic model did so with almost perfect accuracy, while the LOSO models still retained a very good score. This is in-line with previous research and shows that these sensors can be used to detect eating movements for various types of foods. Sen et al. achieved an accuracy of 97% in classification of eating episodes. The difference is slight but can be explained by the higher sample rate used (100 Hz), as well as the fact that they applied smoothing to their original results of 92% (they check if the next two frames and previous 2 frames are different from the current frame and adjust based on it). Dong et al. achieved an accuracy of 81% in detecting eating activities. The difference can be explained by the free-living conditions used in their experiment, which is not done in this research.

4.2 Differentiating between different foods

The results make it clear that the models had a hard time telling apart the different foods, especially when it comes to foods that have similar eating movements. If we look at figures 1 and 2, we can see how the models had no problem differentiating the foods from non-eating movements, but very much had trouble telling them apart from other foods. We can also see those foods eaten in similar a manner (yogurt and soup for example) were confused with each other more often than with the other foods. Mevissen achieved an accuracy of 0.82 when it came to differentiating eating gestures, which included telling apart eating yogurt and eating a croissant. This higher score can be explained by the limited number of foods, as well as the different movements performed when eating these foods. As we can see in figures 1 and 2, the models in this research very rarely confused these foods as well. Overall, the results suggest that these 2 sensors are not sufficient to differentiate between many types of food on their own, especially those similar in their manner of eating.

4.3 Generalizability

The results for the generic and LOSO models were not too far off from each other, suggesting that the models may be generalizable. Of course, the models for recognizing different foods did not work well in this case, but for future research it might be good to note that it should be possible to create generalizable models.

5 CONCLUSION

Using gyroscope and accelerometer sensors worn on the wrist can be used to effectively differentiate between eating and non-eating movements for multiple types of food, with varying similarity. However, when it comes to differentiating between the foods themselves, the 2 sensors prove insufficient to do so on their own, especially when it comes to foods with similar eating movements. The models proved to be generalizable, so for similar research in the future it might be worth building such models.

Using extra sensors to differentiate the foods themselves might be a good approach. For example, a small camera worn somewhere on the body (on a necklace perhaps), can be used, as image recognition has been shown to be effective in this matter[10] and has even been used to estimate calories[6], which should prove important for the future of ADM. The camera can be used in combination with the smartwatch: when the smartwatch detects eating gestures, the camera can turn on and record video or take pictures. There is also the need for doing similar research for left-handed people. Furthermore, this research does not take into account different eating styles: people might eat rice with a fork, spoon, or chopsticks, for example.

REFERENCES

[1] Abarca-Gómez L, Abdeen ZA, Hamid ZA, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. The lancet. 2017;390(10113):2627-2642. https://doi.org/10.1145/161468.161471

[2] Body-mass index and mortality among 1.46 million white adults. New england journal of medicine. 2011;365(9):869-869. doi:10.1056/NEJMx110060

[3] Brownlee J, 2018. LSTMs for Human Activity Recognition Time Series Classification. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/how-to-develop-rnn-models-for-human-activity-recognition-time-series-classification/> [Accessed 2 July 2022].

[4] Chaput JP, Leblanc C, Pérusse L, Després JP, Bouchard C, Tremblay A. Risk factors for adult overweight and obesity in the Quebec Family Study: have we been barking up the wrong tree?. Obesity (Silver Spring). 2009;17(10):1964-1970. doi:10.1038/oby.2009.116

[5] Dong Y, Scisco J, Wilson M, Muth E and Hoover A, "Detecting Periods of Eating During Free-Living by Tracking Wrist Motion," in IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 4, pp. 1253-1260, 2014, doi: 10.1109/JBHI.2013.2282471.

[6] Ege T, Ando Y, Tanno R, Shimoda W and Yanai K, "Image-Based Estimation of Real Food Size for Accurate Food Calorie Estimation," 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2019, pp. 274-279, doi: 10.1109/MIPR.2019.00056.

[7] ELAN (Version 6.3) [Computer software]. 2022. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

[8] Finucane MM, Stevens GA, Cowan MJ, et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9·1 million participants. Lancet (london, england). 2011;377(9765):557-567. doi:10.1016/S0140-6736(10)62037-5

[9] Joshi S and Abdelfattah E, Deep Neural Networ 2021 IEEE 12th Annual Information Technology,ks for Time Series Classification in Human Activity Recognition. Electronics and Mobile Communication Conference (IEMCON), 2021, pp. 0559-0566, doi: 10.1109/IEMCON53756.2021.9623228.

[10] Kong F and Tan J, "DietCam: Regular Shape Food Recognition with a Camera Phone," 2011 International Conference on Body Sensor Networks, 2011, pp. 127-132, doi: 10.1109/BSN.2011.19.

[11] Kyritsis K, Tatli CL, Diou C and Delopoulos A, "Automated analysis of in meal eating behavior using a commercial wristband IMU sensor," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 2843-2846, doi: 10.1109/EMBC.2017.8037449.

[12] Mevissen S. A wearable sensor system for eating event recognition using accelerometer, gyroscope, piezoelectric and lung volume sensors. Master's Thesis. University of Twente. 2021. http://essay.utwente.nl/88431/1/Mevissen_BME_TNW.pdf

[13] Prospective Studies Collaboration. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. The lancet. 2009;373(9669):1083-1096. doi:10.1016/S0140-6736(09)60318-4

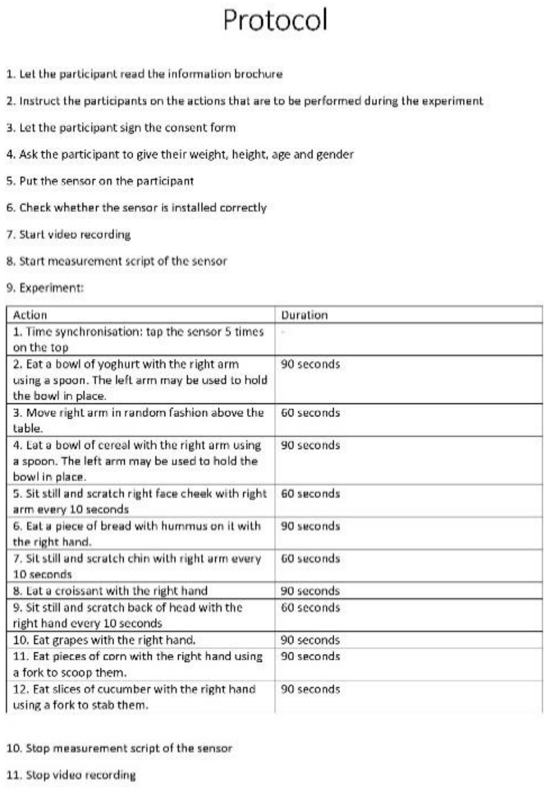
[14] Rumpler WV, Kramer M, Rhodes DG, Moshfegh AJ, Paul DR. Identifying sources of reporting error using measured food intake. The faseb journal. 2007.21(5). doi:10.1096/fasebj.21.5.A310-b

[15] Sen S, Subbaraju V, Misra A, Balan RK and Lee Y, "The case for smartwatch-based diet monitoring," 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), 2015, pp. 585-590, doi: 10.1109/PERCOMW.2015.7134103.

[16] World Health Organization. (2004). Global strategy on diet, physical activity and health. World Health Organization. https://apps.who.int/iris/handle/10665/43035

[17] Zhang S, Stogin W, Alshurafa N. I sense overeating: motif-based machine learning framework to detect overeating using wrist-worn sensing. Information fusion. 2018;41:37-47. doi:10.1016/j.inffus.2017.08.003

A PROTOCOL



Protocol

1. Let the participant read the information brochure
2. Instruct the participants on the actions that are to be performed during the experiment
3. Let the participant sign the consent form
4. Ask the participant to give their weight, height, age and gender
5. Put the sensor on the participant
6. Check whether the sensor is installed correctly
7. Start video recording
8. Start measurement script of the sensor
9. Experiment:

Action	Duration
1. Time synchronisation: tap the sensor 5 times on the top	-
2. Eat a bowl of yoghurt with the right arm using a spoon. The left arm may be used to hold the bowl in place.	90 seconds
3. Move right arm in random fashion above the table.	60 seconds
4. Eat a bowl of cereal with the right arm using a spoon. The left arm may be used to hold the bowl in place.	90 seconds
5. Sit still and scratch right face cheek with right arm every 10 seconds	60 seconds
6. Eat a piece of bread with hummus on it with the right hand.	90 seconds
7. Sit still and scratch chin with right arm every 10 seconds	60 seconds
8. Eat a croissant with the right hand	90 seconds
9. Sit still and scratch back of head with the right hand every 10 seconds	60 seconds
10. Eat grapes with the right hand.	90 seconds
11. Eat pieces of corn with the right hand using a fork to scoop them.	90 seconds
12. Eat slices of cucumber with the right hand using a fork to stab them.	90 seconds

10. Stop measurement script of the sensor
11. Stop video recording