

Objective monitoring of food-intake using cEEGrids

Seminar Thesis

David Blumenthal
2345611

Lukas Stingl
2397980

At the Department of Economics and Management
at the Institute of Information Systems and Marketing (IISM)
Information & Market Engineering

Reviewer: Dr. Michael Knierim
Second reviewer: M.Sc. Karen Bartholomeyczik

10.02.2022

Contents

1	Introduction	2
2	Related Work	4
3	Methods	6
3.1	Data Collection	6
3.2	Preprocessing	7
3.3	Model selection	8
3.4	Model evaluation	9
4	Results	11
5	Discussion	13
6	Conclusion & Outlook	16
7	Declaration	17
	References	19

Abstract

This study presents an approach to enhance the objective monitoring of food intake using cEEGrids. A 3D-printed around-the-ear electrode array was used to acquire data of participants eating different types of food as well as baseline activities. Signals are processed using different algorithms to create feature sets and to recognize patterns of the different featured activities. The results showed that Random Forests performed best on our data set with average accuracy and f1-score of 85.7% and 85.9%. This approach can be useful to automate and objectively measure ingestive behavior. An imaginable application would be in the treatment of overweight or anorexic patients. It could also be applicable to support Alzheimer's patients, as this disease is also often accompanied by a lack of food and fluid intake.

1. Introduction

*"In general, mankind, since the improvement of cookery, eat above twice as much as nature requires."*¹

One of our biggest health problems as a modern society is obesity. Although the World Health Organisation (WHO) has identified obesity as an global epidemic since 2000, (WHO, 2000), the proportion of overweight people is steadily increasing. According to a 2014 study of the German health institution Robert Koch Institut (RKI), about two-thirds of German men and around half of the women are overweight (23% men & 24% women are highly overweight) (RKI, n.d.). This, however, as indicated by the WHO, is not specific to Germany. The OECD 20 countries have an obesity proportion of 59,6 % (Radtko, 2019).

Obesity is characterized by the accumulation of excess body fat and can be seen as the consequence of a positive energy balance (Sharma & Padwal, 2010, 362). Due to its far-reaching effects, obesity causes various serious medical complications such as type 2 diabetes, strokes, and coronary heart diseases (Pi-Sunyer FX, 1998, 899). Apart from individual sickness, consequential illnesses put immense financial pressure on health and social systems (RKI, n.d.). Although the formation of obesity is more complex than excess energy consumption relative to energy expenditure, excessive food intake and insufficient physical activity are the main known causes (Aronne, Nelinson, & Lillo, 2009, 11). Recent changes in food availability and advertisement have encouraged the overconsumption of high-calorie and high-fat foods, easily accessible through numerous fast-food restaurants. This offer of unhealthy food is further enriched by grocery stores that offer a broad selection of "convenience" food that is highly processed and full of sugar. In combination with modern technological advances that made tasks more efficient and consequently reduced the need for physical activity, the shifts in eating behavior increase the risk of developing unhealthy patterns. (Wright & Aronne, 2012, 730f.).

Both food intake and exercising are essential for maintaining energy balance in the human body. Therefore, to have the best conditions for a healthy life, eating and exercising in the right proportions is essential. The monitoring of exercising is already well researched, and various studies exist. Practical applications and devices have emerged (e.g., Smart Watches and Fitness Trackers) and introduced a method to measure and quantify physical activity objectively. However, methods on the objective monitoring of food intake are lacking behind. To date, food records, food frequency questionnaires, as well as 24-h recalls are the most common methods used in treatment (Johnson, 2002, 63f.). Nevertheless, these methods share the same inherent problems. They rely on self-reported data, which is subjective and therefore, due to biases, often inaccurate (Ioannidis, 2013, 347).

This study aims to present an approach to detect food intake activity using an around-the-ear electrode array. These so-called cEEGrids offer possibilities to capture EEG data in an unobtrusive and concealed way (Debener, Emkes, De Vos, & Bleichner, 2015, 1).

¹Benjamin Franklin

To evaluate the approach, we investigate the following research question in a controlled environment: Can a machine learning classifier distinguish between eating different foods (nuts, sandwiches, and soup) and separate them from defined baseline activities. We defined an experiment where data of different activities is collected, analyzed and classified using different machine learning algorithms. The study then discusses achieved results and assesses whether this approach has the potential to improve the treatment of diseases related to dysfunctional food intake. Finally, this study concludes with an outlook and suggests further research directions.

2. Related Work

There have been various indirect approaches to measure food intake objectively. However, due to the task’s difficulty, most applications remain in a laboratory setting. The following chapter gives an overview of existing approaches and outlines limitations and difficulties. Additionally, it provides the reader with a brief description of cEEGrids.

One approach found in the literature, for example, is to use features that are related to sounds, like chewing or swallowing (Passler & Fischer, 2014; Päßler & Fischer, 2012). These methods are usually vulnerable to being influenced by environmental sounds and are therefore not suited for exterior use (Chung et al., 2018). Other approaches use wrist motion tracking (Dong, Hoover, Scisco, & Muth, 2012) to count the number of bites taken to approximate the calories amount over a "calories per bite method". Participant burden is somewhat limited with this method since the user only has to turn the device on when eating. However, this can lead to limitations when users forget to turn on their devices, a circumstance that might be expected in some social situations. Further, this approach is not yet remarkably robust since eating with both hands can make the assessment inaccurate. (Archundia Herrera & Chan, 2018, 4). Different approaches make use of visual information. This requires the participant to take an image of the food before eating it. The analysis scheme is based on the classification of the food image. The classification result is forwarded to a database containing nutrient density values for different food categories. The approach uses depth images to estimate portion size in order to enhance the calorie approximation (Zhang, Yu, Siddiquie, Divakaran, & Sawhney, 2015). Moreover, some approaches try to deconstruct the food image into its ingredients. The ingredients found are then cross-referenced with a database containing calorie density values. Found results are then added up and get returned to the user (Rhyner et al., 2016). The introduced visual methods share the limitations of all methods requiring participants to take action. Limitations include participants forgetting to take pictures or forgetting to carry their smartphones. Further, the methods rely on up-to-date databases. Keeping such databases up to date could be rather difficult, considering the constantly evolving food possibilities (Archundia Herrera & Chan, 2018, 15). A recent study investigated measuring food intake via analysis of temporalis muscle activity using smart glasses with built-in sensors. The study focuses on distinguishing chewing from other activities (winking, resting, talking, walking, chewing while walking). The results were promising as they reached an average f1-score of 80.5%. However, the study did not distinguish between the two different foods served (jelly, toasted bread). Instead, this study tries to distinguish between eating while walking and sedentary eating, but with only moderate success. Nevertheless, chewing can be separated from the rest of the activities (Chung et al., 2018). In contrast to this study, we try to classify three different types of foods (soup, sandwiches, and nuts) among baseline activities.

The OpenBCI platform offers and creates open-source tools for bio-sensing and neuroscience. They strive to lower the barrier to entry the brain-computer interfacing domain (OpenBCI (2022)). They offer hardware (Cyton board and Daisy shield) that enables the

low-cost mobile recording of biosignals on 18 channels. Two channels are used for reference and ground electrodes. cEEGrids, on the other hand, are flexible, printed Ag/AgCl electrodes positioned in a c-shaped array that fits around the ear. These sensors are designed to be unobtrusive and convenient to acquire EEG data under field conditions. They produce reliable, high-quality EEG recordings over several hours. Combining these two devices forms a Recording System that allows for a less expensive and flexible recording solution due to its open-source nature. The recording system with its two components is displayed in Figure 2.1. The solution reduces costs from 10.000 USD to somewhere around 1.500 USD. (M. T. Knierim, Berger, & Reali, 2021).

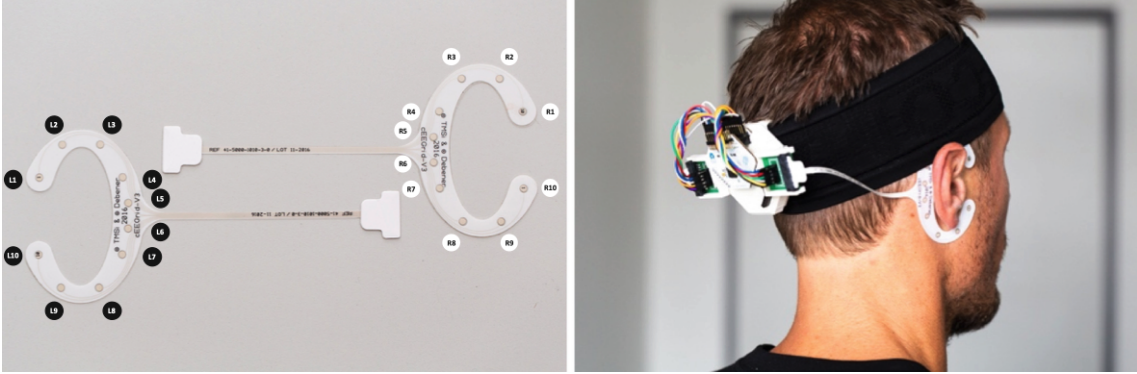


Figure 2.1: The cEEGrid recording system. Left: Two cEEGrid electrodes showing the electrode positions. Right: The OpenBCI biosignal acquisition board with cEEGrids.

This exact setup has already been used to detect daytime bruxism (M. Knierim, Schemmer, & Woehler, 2021) and other reference activities similar to the one in this study. For example, (M. Knierim, Schemmer, & Perusquía-Hernández, 2021) classified different activities like jaw clenching, speaking, and chewing. Further, they tried to classify different types of facial expressions.

3. Methods

In this chapter, we briefly outline our applied research method. First, we started with the design of the experiment, followed by the collection of data using the illustrated recording system from Figure 2.1. Then a brief description of data pre-processing is followed by a section clarifying the model selection.

3.1 Data Collection

To gather data for evaluation, we designed an experiment and implemented the different tasks in PsychoPy, a software solution for creating experiments. Participants conducting the experiment are asked to eat soup, sandwiches, and nuts. Reading a text out loud, yawning, resting with eyes open, resting with eyes closed were implemented as baseline activities. PsychoPy recorded the timestamps of each activities' start and end.

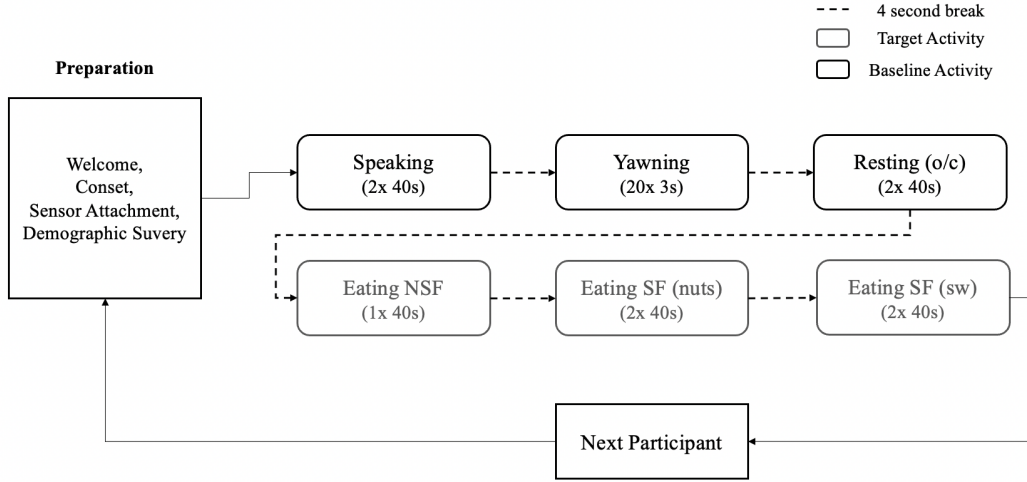


Figure 3.1: Design of the experiment

As shown in Figure 3.1 participants were asked to read a text out loud for 40 seconds. The routine is repeated twice with a break of 4 seconds after each routine respectively. Furthermore, the experiment consists of yawning 20 times for 3 seconds each and resting for 40 seconds, one time with open eyes, one time with closed eyes. Beside the baseline activities, subjects were asked to eat soup once for 40 seconds, nuts and sandwiches for two times 40 seconds each. Due to restrictions caused by COVID-19, only two participants took part in that experiment. To generate enough data to train different models, each participant took part twice. Both participants were male, aged 25 and 26.

As shown in Figure 3.2 participants were connected to an OpenBCI Board. To gather the data, we used the setup designed by M. Knierim, Schemmer, and Perusquía-Hernández (2021) in their research of facial activity recognition. A combination of cEEGrids and OpenBCI amplifiers was used to obtain 16 recording channels.



Figure 3.2: Sensor attachment on participant

3.2 Preprocessing

The recording files were loaded and analyzed using Python. Each channel signal was first mean-centered, then band-pass (FIR 1-62 Hz) and notch (50 Hz) filtered. The continuous signal was cut into the individual activities using the timestamps generated by PsychoPy. The resulting activity measurements were then epoched using non-overlapping one-second windows (125 data points per epoch). As features we choose the sum, maximum, Hurst-exponent (HE) (Hurst, 1951), Petrosian Fractal Dimension (PFD) (Petrosian, 1995), the Higuchi Fractal Dimension (HFD) (Higuchi, 1988) and the Hjorth parameters Activity, Mobility, and Complexity (Hjorth, 1970). These features have been selected due to their feature describing properties and use in the past related works (M. Knierim, Schemmer, & Perusquía-Hernández, 2021; Val-Calvo, Álvarez Sánchez, Ferrández-Vicente, & Fernández, 2019). HE, HFD, and the Hjorth parameters Complexity and Mobility have been calculated using the python library `mne_features`¹. The PFD and Hjorth Activity have been calculated using the python library `eeglib`².

The mentioned features were calculated for each epoch and each channel (16), resulting in a first feature set ($FS1_{signal}$) containing 1840 samples with 128 features. To generate a second feature set ($FS2_{signal+}$) we calculated the first 3 individual components for the whole signal. Individual components are considered to capture the main muscle activity. Following, we calculated the same features as in ($FS1_{signal}$), which resulted in 1840 samples with 152 features.

Due to the unequal length of the different activities, the class frequencies are unevenly distributed, which means we were dealing with an unbalanced data set. The distribution is visualized in Figure 3.3.

¹<https://github.com/mne-tools/mne-features>

²<https://github.com/Xiul109/eeglib>

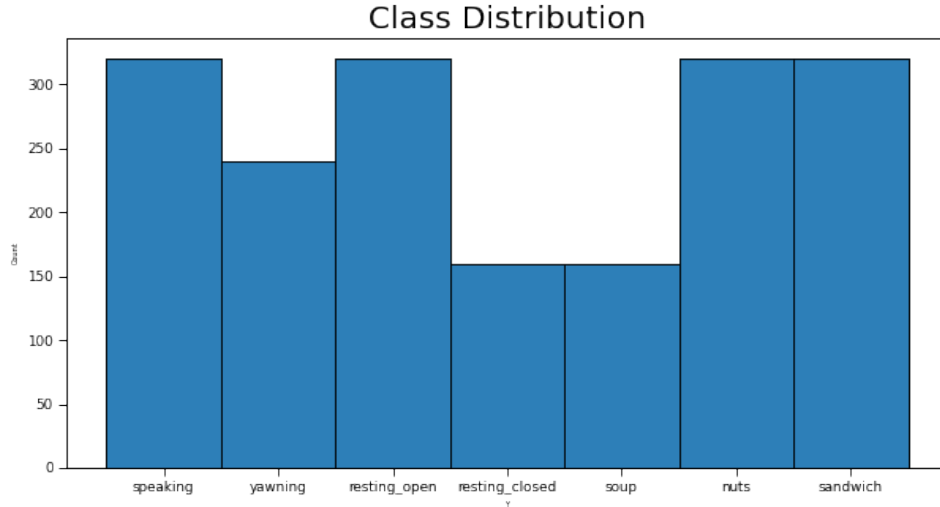


Figure 3.3: Class Distribution of generated feature sets

To account for the data set imbalance we used Synthetic Minority Oversampling (SMOTE), which generates a balanced data set by synthesizing new samples for the minority class (Bowyer, Chawla, Hall, & Kegelmeyer, 2011). After applying the algorithm, the data set was balanced containing exactly 320 samples of each class and 2240 samples in total. The constructed data set was saved as a pickle file. The pickle file could then be imported for model-specific preprocessing and training.

3.3 Model selection

After preprocessing the data, Random Forest, Support Vector Machine, AdaBoost, Logistic Regression, and a Multi-layer Perceptron were tested as classifiers. In our research, we focused on comparing the performance of these different classifiers rather than optimizing them.

The Random Forest classifier can be characterized as an ensemble learning technique using decision trees. Ensemble learning algorithms combine different models and therefore lead to higher prediction performance. Random Forests use two approaches to create randomness. On the one hand, just a subset of data is chosen for each decision tree. On the other hand, the structure of each decision tree is random as well, as the attributes of each tree are also chosen randomly. The classification result is then determined by a majority vote of the individual decision trees. Due to inherent randomness in this algorithm, it is more robust to outliers and has a lower risk of overfitting the training data. The creation of smaller trees leads to faster training but increases proportionally to the number of trees (Breiman, 2001).

Support Vector Machine (SVM) is an approach to separate classes in a multidimensional space. The SVM calculates the optimal hyperplane in a given space to separate two or more classes. For non-linear problems, SVMs use a technique called "kernel trick." This

technique transforms the data in a higher-dimensional space and can again calculate the optimal hyperplanes. As the kernel trick uses kernel functions, SVMs can be used for different applications and are effective in high dimensions. Downsides are the need for external preprocessing, as well as parameter optimization for the optimal kernel (Goodfellow, Bengio, & Courville, 2016).

AdaBoost trains and combines several weak classifiers to receive a strong model. AdaBoost draws a random subset of data from training data. The algorithm then applies weights to each data point of the random subset. In case of a correct classification, weights will be increased, respectively decreased for false classifications. Weight is also applied for every single classifier so that the final classification from AdaBoost concludes from majority voting of the weighted ensemble of weak classifiers. Usually, AdaBoost performs well without costly parameter tuning. On the contrary, AdaBoost has its limitations in noisy data (Freund, Schapire, & Abe, 1999).

Multi-layer Perceptron (MLP) consists of multiple, connected layers of neurons. In general, three different types of layers exist—the input layer, hidden layers, and output layer. The input layer receives the weighted data and forwards it through one or multiple hidden layers to the output layer, resulting in a classification. Each neuron contains a non-linear activation function, which performs a single output. Based on the activation of neurons, the Multi-layer Perceptron is trained using a gradient descent approach, where an error on predictions is calculated and propagated backward through the network. The weight of each neuron will then be adjusted. MLPs perform well with large and complex input data, but computations can be redundant, inefficient, and computationally costly (Goodfellow et al., 2016).

Logistic Regression uses a logistic function to classify input data. The logistic function is described as

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (3.1)$$

The output prediction function returns a probability between 0 and 1. Logistic Regression has a threshold for the classification. For probabilities larger or equal to the threshold Logistic Regression will classify the output as positive, while classifying the output negative with a lower probability than the threshold. Logistic regression is easy to implement and train. The algorithm expects pre-processed data, which increases the computational costs of the overall classification. In addition, Logistic Regression is bound to a linear decision as it is based on a linear model (Goodfellow et al., 2016).

3.4 Model evaluation

Models can be evaluated using stratified k-fold cross-validation. Stratified k-fold cross-validation shuffles the data set and splits the data in k folds. During the first iteration, the model is trained using k-1 folds, whereas one fold is used to evaluate the model. Each iteration uses a different set of training and evaluation data, resulting in an accurate performance estimation. It can be concluded, that the performance of the model is less

biased by the selection of specific train or validation data (Refaeilzadeh, Tang, & Liu, 2016).

To measure the performance of classifiers, accuracy, precision, recall, and f1-score are relevant metrics. The equations are displayed.

$$Accuracy : \frac{TP + TN}{TP + TN + FP + FN}. \quad (3.2)$$

$$Precision : \frac{TP}{TP + FP}. \quad (3.3)$$

$$Recall : \frac{TP}{TP + FN}. \quad (3.4)$$

$$F1 : \frac{2 * Precision * Recall}{Precision + Recall} \quad (3.5)$$

Where TP denotes the true positives, i.e., samples with the actual label positive correctly identified as such by the classifier. TN denotes the true negatives that were correctly identified by the classifier. FP denotes false positives. Samples where the classifier assigned the label positive although the actual label is negative. Vice versa for false negatives denoted by FN .

Accuracy estimates the correct classified instances. The *F1* score is defined as the harmonic mean between precision and recall. Precision estimates true positive classes compared to all positive classified classes (true positive and false positive). In contrast, recall estimates the true positive classified examples compared to the true positive and false negative classified classes. (Flach, 2003).

A Confusion Matrix is used to observe correct and false classified instances. Correct classifications are plotted against predicted outcomes. With a Confusion Matrix errors between classes can be shown (Xu, Krzyzak, & Suen, 1992).

4. Results

We used 5-fold cross-validation to evaluate the performance of the previously introduced algorithms. Since precision and recall have a trade-off, we analyzed accuracy and f1-score to inspect if both precision and recall showed a sufficient result. The classifiers were trained for both subjects together, there was no differentiation between participants. Comparing Table 4.1 and Table 4.2, classification results improved when using SMOTE.

	Random Forest	SVM	MLP	AdaBoost	Logistic Regression
FS1 (accuracy)	79.8%	78.9%	78.0%	49.0%	77.7%
FS2 (accuracy)	79.8%	79.2%	79.2%	51.0%	77.7%
FS1 (f1-score)	78.6%	77.5%	78.0%	41.0%	77.4%
FS2 (f1-score)	79.9%	77.8%	79.1%	41.2%	77.5%

Table 4.1: Results of 5-fold cross-validation for each classifier with unbalanced data set

On the basis of the unbalanced data set, Random Forest showed the best performance, as shown in Table 4.1. The accuracy was 79.8% for both ($FS1_{-signal}$) and ($FS2_{-signal+}$). The f1-score improved 0.3% to 79.9% using ($FS2_{-signal+}$) instead of ($FS1_{-signal}$). In general using ($FS2_{-signal+}$) instead of ($FS1_{-signal}$) leads to a greater or equal accuracy and f1-score, with only AdaBoost as an exception. The SVM and MLP show both nearly 80% accuracy as well, both for ($FS1_{-signal}$) and ($FS2_{-signal+}$). The MLP had a slightly bigger increase in accuracy with ($FS2_{-signal+}$) of 1.2% to 79.2%, compared to the SVM. Logistic Regression reached 77.7% in accuracy for both ($FS1_{-signal}$) and 2. The f1-score increased by 0.1% to 77.5% using ($FS2_{-signal+}$). AdaBoost had the lowest performance among all classifiers for accuracy, with 44.2% and 41.7% for ($FS1_{-signal}$) and ($FS2_{-signal+}$), as well as for the f1-score, with 35.2% and 33.1% for ($FS1_{-signal}$) and ($FS2_{-signal+}$).

	Random Forest	SVM	MLP	AdaBoost	Logistic Regression
FS1 (accuracy)	85.0%	82.5%	84.2%	44.2%	82.2%
FS2 (accuracy)	85.7%	82.8%	84.4%	41.7%	82.9%
FS1 (f1-score)	84.6%	82.3%	84.0%	35.2%	82.1%
FS2 (f1-score)	85.9%	82.7%	82.9%	33.1%	82.8%

Table 4.2: Results of 5-fold cross-validation for each classifier with SMOTE

As shown in Table 4.2, Random Forest, Support Vector Machine, Multilayer Perceptron, and Logistic Regression had an accuracy and f1-score of above 80%, both for ($FS1_{-signal}$) and ($FS2_{-signal+}$) in the oversampled data set. Solely AdaBoost had accuracy scores of 44.2% and 41.7% for ($FS1_{-signal}$) and ($FS2_{-signal+}$). In general ($FS2_{-signal+}$) led to higher accuracy and f1-score. Exceptions are the f1-score of the Multilayer Perceptron and the accuracy and f1-score of the AdaBoost. The best performance among all classifiers yields the Random Forest, with an accuracy of 85.0% and 85.7% for ($FS1_{-signal}$) and ($FS2_{-signal+}$). The f1-score is 84.6% and 85.9% for ($FS1_{-signal}$) respectively ($FS2_{-signal+}$).

In Figure 4.1 the classification results of the Random Forest classifier are represented in a normalized confusion matrix. The correct classified classes are plotted diagonally.

The activity speaking, was assigned the wrong label resting closed in 21 % of the cases and represents the weakest classification result amongst all activities. Further, resting with open eyes and yawning have a confusion rate of 11% and 8%. Especially the eating activities were separated well, with a correct classification of above 90%. There appeared to be little confusion amongst the individual eating activities. Eating sandwiches and eating nuts yield a correct classification of 93%, eating soup was classified with 92% correct. The source code as well as the data set is available on GitHub^[1]

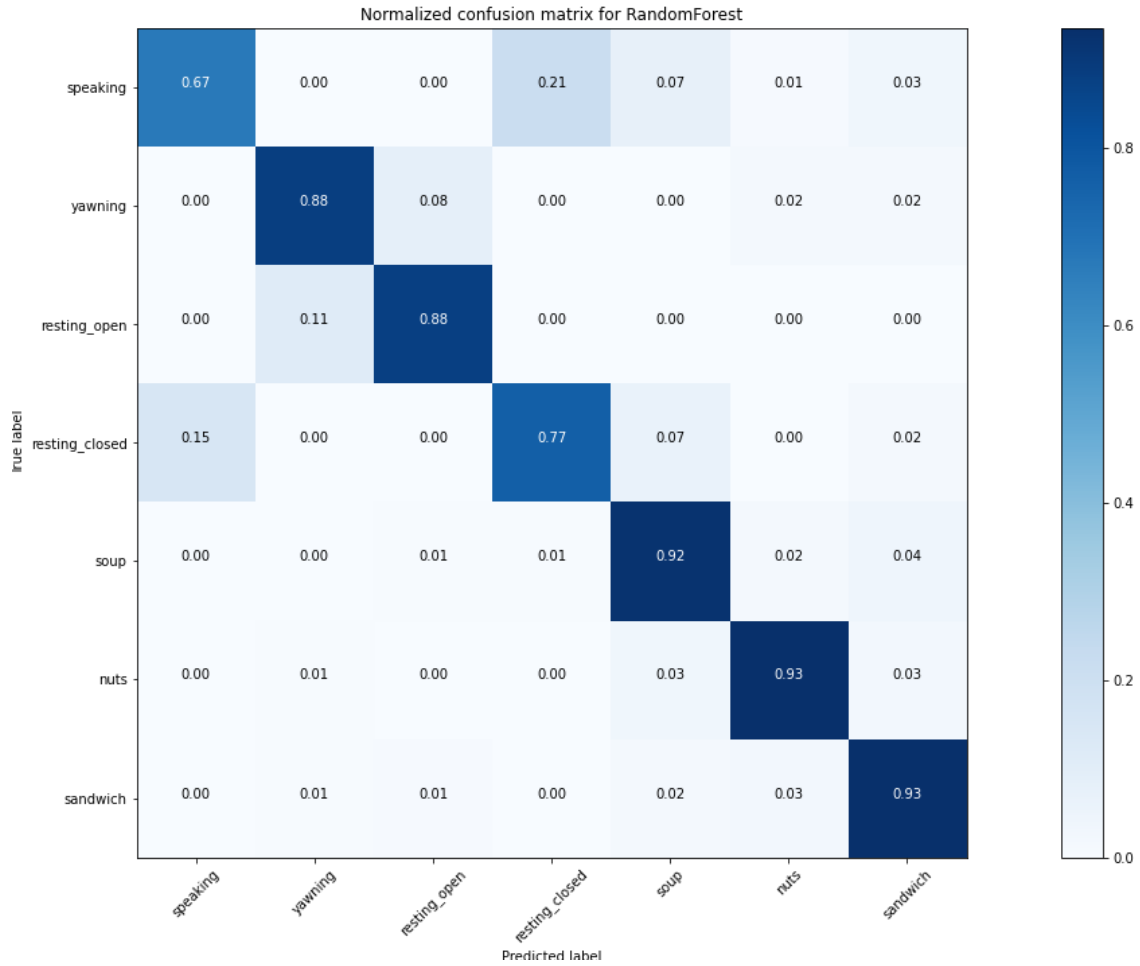


Figure 4.1: Normalized confusion matrix of Random Forest

¹https://github.com/davidvblumenthal/nourriture_cEEG

5. Discussion

Research on classifying food intake is still in its early stages, yet it can become an essential factor for obesity treatment. Our research started with an initial classification of different solid foods and non-solid food, compared to baseline activities speaking, yawning, and resting.

To generate data that is relatively close to the actual distribution, we designed the routines as naturally as possible. For example, the experiments were conducted at different intervals, e.g., participants had to yawn 20 times for three seconds in a row instead of once for 40 seconds. On the other hand, talking was done for a longer period.

We achieved good classification results across all activities. This result was somewhat expectable as Knierim et al. conducted a comparative investigation. Consistent with our findings was furthermore the fact that activities involving larger facial muscles tend to have better prediction scores [M. Knierim, Schemmer, and Perusquía-Hernández (2021)]. The immaculate classification of eating soup could be explained by using a spoon and opening our mouths relatively wide. Since we ate tomato soup without any solid ingredients, we did not chew to swallow. Soup and sandwiches also show good separation. While taking a bite from a sandwich, nuts can be chewed directly with the back teeth. In addition, to take a bite, it is usually required to open the mouth more widely than when eating nuts. As shown in the confusion matrix, Random Forest confused baseline activities frequently. For example, speaking activities were classified as resting with closed eyes. Movements could cause this during the resting phase as the cEEGrids recognize even slight motions, which could have distorted the data on resting. Even in a controlled environment, it is complicated to prevent subjects from moving. However, it represents natural behavior.

With our findings, we were able to tackle limitations of related experiments. In contrast to the work of [Passler and Fischer (2014)] and [Päßler and Fischer (2012)], in which food intake was classified by sounds, our approach is not distorted by external noise. In addition, we have shown that it is possible to classify different foods with cEEGrids. Nonetheless, our approach could not eliminate all limitations from prior research. [Archundia Herrera and Chan (2018)] described in their paper, that it is costly for people to turn their devices on and off before eating. In our setup people need to wear an around-the-ear electrode array, to gather data. This is not yet practical and requires further improvements. In summary, our machine learning model can already classify a limited number of different foods. However, it is debatable if the sheer endless food possibilities can be distinguished with this approach with the needed accuracy. Therefore, a combination with other techniques is recommended. For example, it would be conceivable to supplement the data with images taken via smart glasses.

However, it might not be necessary to classify which food is eaten exactly to generate insights into eating patterns. Therefore the model could already be used to enrich and enhance obesity treatments. A possible application could be the classification of food and fluids compared to not eating and drinking at all. In nursing homes, food intake is

measured by 24-Hour dietary recall, food frequency questionnaires or biochemical markers. Recall and questionnaires are subjective and prone to errors, particularly when completed by elderly people in need of care.

Since we designed the experiment to set the environment as natural as possible, the routines have different lengths. Since we cut the data in one-second epochs, we received an unbalanced data set during feature engineering. The problem with an unbalanced dataset is that the classifier tends to focus its attention on the majority classes, which often leads to bad classification results for the minority class. To overcome this challenge, we over-sampled the minority classes with SMOTE. However, synthetically generated data does not necessarily represent the actual underlying distribution, which can bias the result and lead to a model not being able to generalize on real data. Additionally, oversampling can lead to more noisy data, which can result in an overfitting classifier (Fernandez, Garcia, Herrera, & Chawla, 2018).

As we focused on the general performance of different classifiers, we did not optimize parameters for different classifiers. Therefore classifiers could lead to different results when applying parameter optimization. MLPs are known for their high capacity and state-of-the-art performance on complex problems. However, they require a large amount of training data. The generated data set might not have been large enough for MLPs to learn the underlying pattern, impacting their performance negatively. Especially for small data sets, the generalization can be affected negatively since it could be influenced by the selection of training and test data. Although we tackled this problem with the 5-fold cross-validation, there could be a significant improvement in the classification results when extending observations (Gardner & Dorling, 1998). Even though the Random Forest classifier already performed remarkably well, performance could still be improved. We did not distinguish between features' relevance; those with a low information gain can harm the overall performance. Moreover, unnecessary features increase the size of possible Decision Trees, which correlates with the computational costs. In conclusion, the Random Forest classifier could yield even higher performance when taking information gain of features into account (Rogers & Gunn, 2006). The scores of the 5-fold cross-validated models indicate a very stable and generalized performance. Nevertheless, it has to be taken into account that the data set is not representative since only two participants conducted the experiment and people from different cultural backgrounds form individual eating habits (Fischler, 1980).

In our research, we combined the data sets of each subject. Another approach for further research could be the separation of the subjects. The classifiers could be trained using the data set of one participant to predict the outcome of another participant. This approach could provide information about how well models can generalize.

A different modeling approach could be the use of more modern, powerful algorithms like convolutional neural networks (CNNs) and Long short-term memory networks (LSTMs). Considering their excellent performance on complex time series and other complex data like images, these models could outperform the models presented in this work. However,

since they require immense training data, a more extensive experiment setup would be needed.

Besides selecting the best model, calorie approximation also remains unsolved. It is still not evaluated how calories will be mapped to the eating activities. Carrots and nuts, for example, may contract similar muscles, which could lead to a similar classification. The confusion could be problematic, as they differ significantly in calorie density.

As described earlier, the current hardware is neither comfortable to wear nor easy to set up. However, powerful applications can emerge when sensors become applicable for everyday use. It is conceivable to set up a training pipeline where the sensor is worn in everyday life. By establishing an interface between the subject, EEG Sensor, and the machine learning model, the participant can provide information on what he is currently doing or eating. Extensive amounts of labeled training data can be generated moderately fast and at relatively low costs by leveraging this human-machine-collaboration approach. The model can continuously be updated in the background with the newly generated data and constantly expand its classification classes.

6. Conclusion & Outlook

This work investigated the potential of using cEEGrids to classify eating behavior and distinguish it from other facial activities. To tackle the problem of objectively measuring individuals' food intake (behavior). Obesity is a disease that is becoming more widespread in developed societies and is putting increasing pressure on the health insurance system. Current obesity treatments still relies on data collection methods characterized by subjectivity and offer room for improvement. The experiment showed promising results. Random Forests showed to be the best performing model, achieving an f1-score of 85.9% and an accuracy of 85.7%. Further, even detecting different kinds of foods (nuts, sandwiches, and soup) was possible with very high accuracy. The classifier separated the eating activities almost perfectly, with the worst confusion rate at only 0.04, which occurred between eating soup and eating sandwiches. The high accuracy is impressive considering the short data collection period. Although the cEEGrids can already be worn unobtrusive and non-restrictive, the current form is not yet practical enough for daily and especially not flexible use. Further research should therefore investigate decreasing the size of cEEGrids and making the installation more accessible, so they can be worn in everyday life while still producing reliable measurements. Further, one must reflect that our experiment was conducted in a closed environment and produced a data set, which only contains a limited range of activities. Hence the data set should first be supplemented with a broader range of activities and subjects to evaluate the application in the natural habitat of a human being. Further, it is not yet clear how different granular types of food can be distinguished. Illustrated by an example, it might be challenging to distinguish carrots from eating nuts. However, these two activities differ significantly in the number of calories. Nevertheless, there is much potential to make the measurement of calorie intake less subjective using cEEGrids, for which this paper can play a small part.

Furthermore, the current functionality could be sufficient to detect unhealthy patterns in the food intake of overweight patients. For example, it would be conceivable to monitor eating behavior throughout the day to detect that the person eats conspicuously much at work, eats very frequently during the day in general, or always eats quite late in the evening. Other use cases emerge when taking a somewhat different view away from obesity towards other diseases. People with dementia often suffer from a deregulated food intake since they sometimes no longer know when to eat and whether they have already eaten or drunk. They then cook, eat and drink less (or even more), leading to malnutrition (occasionally overeating) or lack of fluids. Depending on the living situation, patients themselves could be reminded to eat or drink something, or the attending person could be made aware. Other applications could lay in the treatment of anorexia. For example, detecting generic eating events or patterns could supply valuable data and insights to treating physicians or individuals and improve treatment.

As expressed, a deviation from regular food intake in either direction is potentially harmful to health. In both cases, objective monitoring of food intake can improve the situation.

7. Declaration

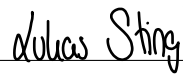
Ich versichere hiermit wahrheitsgemäß, die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des Karlsruher Instituts für Technologie (KIT) zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

Karlsruhe, den 10. 02. 2022



David Blumenthal

Karlsruhe, den 10. 02. 2022



Lukas Stingl

References

- Archundia Herrera, M. C., & Chan, C. B. (2018, August). Narrative Review of New Methods for Assessing Food and Energy Intake. *Nutrients*, 10(8), 1064. Retrieved 2022-01-04, from <https://www.mdpi.com/2072-6643/10/8/1064> (Number: 8) doi: 10.3390/nu10081064
- Aronne, L. J., Nelinson, D. S., & Lillo, J. L. (2009, January). Obesity as a disease state: A new paradigm for diagnosis and treatment. *Clinical Cornerstone*, 9(4), 9–29. Retrieved 2022-01-04, from <https://www.sciencedirect.com/science/article/pii/S1098359709800021> doi: 10.1016/S1098-3597(09)80002-1
- Bowyer, K. W., Chawla, N. V., Hall, L. O., & Kegelmeyer, W. P. (2011). SMOTE: synthetic minority over-sampling technique. *CoRR*, abs/1106.1813. Retrieved from <http://arxiv.org/abs/1106.1813>
- Breiman, L. (2001). *Machine Learning*, 45(1), 5–32. Retrieved from <https://doi.org/10.1023/a:1010933404324> doi: 10.1023/a:1010933404324
- Chung, J., Oh, W., Baek, D., Ryu, S., Lee, W. G., & Bang, H. (2018). Design and Evaluation of Smart Glasses for Food Intake and Physical Activity Classification. *Journal of Visualized Experiments : JoVE*(132). Retrieved 2022-01-04, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5912407/> (Publisher: MyJoVE Corporation) doi: 10.3791/56633
- Debener, S., Emkes, R., De Vos, M., & Bleichner, M. (2015, November). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Scientific Reports*, 5(1), 16743. Retrieved 2022-01-04, from <https://www.nature.com/articles/srep16743> doi: 10.1038/srep16743
- Dong, Y., Hoover, A., Scisco, J., & Muth, E. (2012, September). A New Method for Measuring Meal Intake in Humans via Automated Wrist Motion Tracking. *Applied Psychophysiology and Biofeedback*, 37(3), 205–215. Retrieved from <https://doi.org/10.1007/s10484-012-9194-1> doi: 10.1007/s10484-012-9194-1
- Fernandez, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018, April). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, 61, 863–905. Retrieved from <https://doi.org/10.1613/jair.1.11192> doi: 10.1613/jair.1.11192
- Fischler, C. (1980, December). Food habits, social change and the nature/culture dilemma. *Social Science Information*, 19(6), 937–953. Retrieved from <https://doi.org/10.1177/053901848001900603> doi: 10.1177/053901848001900603

- Flach, P. A. (2003). The geometry of roc space: understanding machine learning metrics through roc isometrics. In *Proceedings of the 20th international conference on machine learning (icml-03)* (pp. 194–201).
- Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780), 1612.
- Gardner, M., & Dorling, S. (1998, August). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15), 2627–2636. Retrieved from [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0) doi: 10.1016/S1352-2310(97)00447-0
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. (<http://www.deeplearningbook.org>)
- Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, 31(2), 277-283. Retrieved from <https://www.sciencedirect.com/science/article/pii/0167278988900814> doi: [https://doi.org/10.1016/0167-2789\(88\)90081-4](https://doi.org/10.1016/0167-2789(88)90081-4)
- Hjorth, B. (1970). Eeg analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3), 306-310. Retrieved from <https://www.sciencedirect.com/science/article/pii/0013469470901434> doi: [https://doi.org/10.1016/0013-4694\(70\)90143-4](https://doi.org/10.1016/0013-4694(70)90143-4)
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116(1), 770-799. Retrieved from <https://ascelibrary.org/doi/abs/10.1061/TACEAT.0006518> doi: 10.1061/TACEAT.0006518
- Ioannidis, J. P. A. (2013). Implausible results in human nutrition research. *BMJ*, 347. Retrieved from <https://www.bmj.com/content/347/bmj.f6698> doi: 10.1136/bmj.f6698
- Johnson, R. K. (2002). Dietary intake—how do we measure what people are really eating? *Obesity Research*, 10(S11), 63S-68S. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1038/oby.2002.192> doi: <https://doi.org/10.1038/oby.2002.192>
- Knierim, M., Schemmer, M., & Perusquía-Hernández, M. (2021). Exploring the recognition of facial activities through around-the-ear electrode arrays (ceegrids). In *Information systems and neuroscience: Neurois retreat 2021. ed.: F. d. davis* (Vol. 52, p. 47-55). Springer. doi: 10.1007/978-3-030-88900-5_6
- Knierim, M., Schemmer, M., & Woehler, D. (2021). Detecting daytime bruxism through convenient and wearable around-the-ear electrodes. In *Advances in usability, user experience, wearable and assistive technology – proceedings of the ahfe 2021 virtual conferences on usability and user experience, human factors and wearable technologies, human factors in virtual environments and game design, and human factors and assistive technology, july 25-29, 2021, usa. ed.: T. ahram* (Vol. 275, p. 26-33). Springer. doi: 10.1007/978-3-030-80091-8_4
- Knierim, M. T., Berger, C., & Reali, P. (2021). Open-source concealed EEG data collection for brain-computer-interfaces - real-world neural observation through openbci

- amplifiers with around-the-ear ceegrid electrodes. *CoRR*, *abs/2102.00414*. Retrieved from <https://arxiv.org/abs/2102.00414>
- OpenBCI. (2022). *About | openbci*. Retrieved from <https://openbci.com/about>. (Accessed on 2022-02-06)
- Passler, S., & Fischer, W.-J. (2014). Food intake monitoring: Automated chew event detection in chewing sounds. *IEEE J Biomed Health Inform.*, *18*, 278-289.
- Petrosian, A. (1995). Kolmogorov complexity of finite sequences and recognition of different preictal eeg patterns. In *Proceedings eighth ieee symposium on computer-based medical systems* (p. 212-217). doi: 10.1109/CBMS.1995.465426
- Pi-Sunyer FX, e. a. (1998, October). Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: executive summary. Expert Panel on the Identification, Evaluation, and Treatment of Overweight in Adults. *The American Journal of Clinical Nutrition*, *68*(4), 899-917. Retrieved 2022-01-04, from <https://doi.org/10.1093/ajcn/68.4.899> doi: 10.1093/ajcn/68.4.899
- Päßler, S., & Fischer, S. (2012). Food intake activity detection using an artificial neural network. *Biomed Tech (Berl)*.
- Radtke, R. (2019). *Übergewicht und Fettleibigkeit nach Ländern weltweit*. Retrieved 2022-01-04, from <https://de.statista.com/statistik/daten/studie/153908/umfrage/fettleibigkeit-unter-erwachsenen-in-oecd-laendern/>
- Refaeilzadeh, P., Tang, L., & Liu, H. (2016). Cross-validation. In *Encyclopedia of database systems* (pp. 1-7). Springer New York. Retrieved from https://doi.org/10.1007/978-1-4899-7993-3_565-2 doi: 10.1007/978-1-4899-7993-3_565-2
- Rhyner, D., Loher, H., Dehais, J., Anthimopoulos, M., Shevchik, S., Botwey, R. H., ... Mougiakakou, S. (2016, May). Carbohydrate Estimation by a Mobile Phone-Based System Versus Self-Estimations of Individuals With Type 1 Diabetes Mellitus: A Comparative Study. *J Med Internet Res*, *18*(5), e101. Retrieved from <http://www.jmir.org/2016/5/e101/> doi: 10.2196/jmir.5567
- RKI. (n.d.). *Gesundheitsmonitoring - Übergewicht und Adipositas* [Government Institution]. Retrieved 2022-01-04, from https://www.rki.de/DE/Content/Gesundheitsmonitoring/Themen/Uebergewicht_Adipositas/Uebergewicht_Adipositas_node.html
- Rogers, J., & Gunn, S. (2006). Identifying feature relevance using a random forest. In *Subspace, latent structure and feature selection* (pp. 173-184). Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007/11752790_12 doi: 10.1007/11752790_12
- Sharma, A. M., & Padwal, R. (2010). Obesity is a sign – over-eating is a symptom: an aetiological framework for the assessment and management of obesity. *Obesity Reviews*, *11*(5), 362-370. Retrieved 2022-01-04, from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-789X.2009.00689.x> (eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-789X.2009.00689.x>) doi: 10.1111/j.1467-789X.2009.00689.x
- Val-Calvo, M., Álvarez Sánchez, J. R., Ferrández-Vicente, J. M., & Fernández, E. (2019). Optimization of real-time eeg artifact removal and emotion estimation

- for human-robot interaction applications. *Frontiers in Computational Neuroscience*, 13. Retrieved from <https://www.frontiersin.org/article/10.3389/fncom.2019.00080> doi: 10.3389/fncom.2019.00080
- WHO. (2000). Obesity: Preventing and Managing the Gloabl Epidemic. *WHO Technical Report Series*, 894. Retrieved 2022-01-04, from <https://apps.who.int/iris/handle/10665/119927> (Section: vi, 133 p. ; 30 cm.)
- Wright, S. M., & Aronne, L. J. (2012, October). Causes of obesity. *Abdominal Radiology*, 37(5), 730–732. Retrieved 2022-01-04, from <http://link.springer.com/10.1007/s00261-012-9862-x> doi: 10.1007/s00261-012-9862-x
- Xu, L., Krzyzak, A., & Suen, C. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3), 418–435. Retrieved from <https://doi.org/10.1109/21.155943> doi: 10.1109/21.155943
- Zhang, W., Yu, Q., Siddiquie, B., Divakaran, A., & Sawhney, H. (2015, May). "Snap-n-Eat": Food Recognition and Nutrition Estimation on a Smartphone. *Journal of diabetes science and technology*, 9(3), 525–533. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/25901024> (Edition: 2015/04/21 Publisher: SAGE Publications) doi: 10.1177/1932296815582222