EXPLORING EARABLE TECHNOLOGY TO DETECT CHEWING MOMENTS IN SEDENTARY COMMON DAILY ACTIVITIES

by

Roya Lotfi

A Thesis Submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

Master of Science

Department of Computer Science

University of Manitoba

Winnipeg

ABSTRACT

The feasibility of collecting various data from built-in wearable sensors has enticed many researchers to use these devices for analyzing human activities and behaviors. In particular, audio, video, and motion data have been utilized for automatic dietary monitoring. In this research, we investigate the feasibility of detecting chewing activities based on audio and inertial sensor data obtained from an ear-worn device, eSense. We process each sensor data separately and determine the accuracy of each sensing modality for chewing detection when using MFCC and Spectral Centroid as features and Logistic Regression, Decision Tree, and Random Forest as classifiers. We also measure the performance of chewing detection when fusing features extracted from both audio and inertial sensor data. We evaluate the chewing detection algorithm by running a pilot study inside a lab environment on a total of 5 participants. This consists of 130 minutes of audio and inertial measurement unit (IMU) data. The results of this study indicate that an in-ear IMU with an accuracy of 95% outperforms audio data in detecting chewing and fusing both modalities improves the accuracy up to 97%.

PUBLICATIONS

---

[1] R. Lotfi, G. Tzanetakis, R. Eskicioglu, and P. Irani. "A Comparison between Audio and IMU data to Detect Chewing Events Based on an Earable Device." In: *11th Augmented Human International Conference*.

ACKNOWLEDGMENTS

I'd like to express my appreciation to my advisor, Dr. Pourang Irani for being an amazing mentor personally and professionally during these two years. I thank Dr. George Tzanetakis, for his continuous guidance during this journey. I thank Dr. Shahin Kamali, Dr. Carson Leung, and Dr. Parimala Thulasiraman for being on my committee. I appreciate Dr. Rasit Eskicioglu for providing me with electronic devices. I thank my family for their support and encouragement. I would not have been able to complete this program without my mother and my father's motivations. I thank my colleagues in the HCI lab for their amazing company during this experience. I would like to acknowledge the Visual and Automated Disease Analytics program for the wonderful educational experience and for funding my research.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The wearable technology market has seen rapid growth, with wearable devices seeing increased rates of acceptance [45]. Various studies have focused on collecting information regarding human activities by leveraging the built-in sensors of wearable devices. In the context of persuasive technology [16], recorded information on human activity can provide insights into behavior and habits in order to provide the users of the wearable devices with strategies for a healthier lifestyle.

According to Statistics Canada, one out of four Canadians aged 18 or more is classified as obese [12]. This issue is not limited to Canada, and the prevalence of obesity has tripled during the past decades [30]. Studies on mindful eating suggest that higher body mass index is associated with a lower rate of mindful eating [26]. For example, the consumed meal proportion was increased by 71% as a result of watching television while eating [40]. Consequently, various methods have been proposed to assist with practicing mindful eating, as a method to prevent and treat obesity and maintain a healthy weight [29].

Sound, heat, body motion, and wrist motion are all generated while eating. In particular, tell-tale sounds are generated from the crushing of food. Several works have successfully detected and distinguished chewing sounds from other body-generated or environmental sounds [3][34][6]. The in-ear microphone is one of the most popular devices used to detect chewing sounds. These microphones can

capture chewing sounds as they are placed close to the users' mouths. In addition, body generated sounds are amplified in the ear canal. Several methods were proven to be successful in detecting swallowing sounds by placing a microphone near the throat [6][36]. Stain sensors, which detect muscle forces, have also been used to capture facial muscle activity while chewing [35][14].

In this thesis, we use the eSense device [23] for the purpose of chewing detection, inspired by the existing works conducted on chewing detection. These studies were based on microphones placed in earbuds and also used accelerometer data to detect the oscillatory movement of the temporalis muscle[1] while engaging in a chewing activity.

The eSense device is an ear-bud platform equipped with a 6-axis IMU (inertial measurement unit), a microphone, and Bluetooth and BLE radios (Figure 1.1). In this research, we investigate the performance of the eSense device and its built-in sensors to detect chewing while performing common sedentary activities such as eating, speaking, watching a movie, and sitting still.

Figure 1.1: eSense device and its specifications.

Our contributions in this research are as follows:

1 A fan-shaped muscle located on top of the jaw. It is one of the mastication muscles and its main function is to move the lower jaw.

- Assessing if the chewing movement of the head can be distinguished from other head-related movements while recording IMU data;
- Assessing if chewing sounds can be distinguished from speaking, watching TV as well as silence;
- Comparing if IMU data offers better signals than audio for detecting chewing when using the same feature extraction and classification techniques;
- Demonstrating the effect of fusing both sensors.

2 RELATED WORK

Several methods have been proposed for automatic dietary and food-intake monitoring based on various sensing modalities. Some examples of these sensing modalities include motion, audio, and video. Several approaches in the literature have employed in-ear devices for the detection of eating. This section discusses the chewing detection methods based on audio and motion data collected from both in-ear and other head-worn wearable devices.

A microphone placed inside the ear canal was used by Amft et al. [3] to explore the possibility of distinguishing between silence, speech, and chewing sounds. An accuracy of 99% was achieved for chewing detection. The authors also measured the intensity of chewing and speech signals when a microphone is placed in different positions of the subjects' head, namely, inner ear, 2 centimeters in front of the mouth, at the cheek, 5 centimeters in front of ear canal opening, collar, and behind the outer ear. They demonstrated that the ear canal is the best placement for the microphone as the chewing sounds have higher intensity compared to speech and environmental sounds in this position.

An in-ear device, equipped with a microphone and PPG sensor, was used by Papapanagiotou et al. to detect chewing and eating events [32]. An accelerometer was also used and, assuming the participants were not eating while engaged in physical activity, data was analyzed to filter out these events. An accuracy of

93.1% was achieved for the detection of chewing events in a semi-controlled lab environment, from the data collected from a total of 14 subjects over a period of 60 hours.

A piezoelectric film sensor, placed below the earlobe, was used by Farooq et al. to detect jaw motions while engaging in different daily activities [14]. Hand-to-mouth gesture sensors were also used to detect bites, and a 3-axis accelerometer was used to capture body movements. Data was collected in a free-living environment for 24 hours with a total of 12 participants. An accuracy of 93% was achieved by fusing all three sensors and extracting both time- and frequency-domain features from the collected data.

EarBit is a head-mounted wearable system that can detect eating activities in a free-living environment [6]. It consists of two IMU devices placed behind the ear and a proximity sensor placed inside the outer ear canal. A microphone was also placed around the neck to detect swallowing activity. The EarBit device was tested by a group of 10 participants over a duration of 45 hours in total, an accuracy of 93% was achieved for eating detection.

In another study, Wang et al. collected data using a single-axis accelerometer attached to the temporalis muscle to detect chewing activities by measuring muscle bulges [42][43]. Accelerometer data were also employed to detect chewing frequency. In a study consisting of 10 participants and a total of 150 hours of recording data, an accuracy of 97% was achieved for the detection of chewing activities.

In yet another study, a 3-axis accelerometer was on the temple of eyeglasses, aiming to detect the oscillatory movement of the temporalis muscle while chewing [28]. Accelerometer data were recorded from 5 participants while they engaged in chewing and non-chewing activities. An accuracy of 73.98% was achieved.

Similar to the study from Meres et al. [28], chewing detection was studied by Farooq et al. [15], using an accelerometer placed in the eyeglasses frame. The SPLENDID dataset [33] was later used by Papapanagiotou et al. [31]. They used convolutional neural networks on audio data and managed to achieve 98% accuracy when distinguishing between chewing sounds from non-chewing sounds. A deep learning pipeline was implemented by Gao et al. to explore the feasibility of chewing detection using the built-in microphone of common headset devices available on the market [19]. A 94-95% classification accuracy for chewing detection from data collected inside the lab environment was achieved, as well as a 94.72% accuracy for real living environment data. Multi-modal sensing based on Google Glass, smartwatch, and in-ear microphone was used by Merk et al. [27] to detect chewing. They achieved a precision of 92% while combining audio and motion modalities on 72 hours of recording data over 5 participants.

Various successful methods have been proposed for chewing detection based on head-worn devices. The result of the previous studies guided us to select the eSense device as a platform to explore its performance in detecting chewing activities. In-ear microphone sensor has been investigated by many other works and has proven to be an effective modality for chewing detection. In addition, the oscillatory movement of facial muscles was also a good indicator of chewing activities. As a result, in addition to considering the audio data, we investigate whether the movement of the facial muscles while chewing are reflected on the in-ear IMU sensor of the eSense device and whether the captured IMU signals can be distinguished from other facial and head movement activities while performing different machine learning classification algorithms.

Table 2.1: Comparison of related works - 1

REFERENCE	SENSORS	PLACEMENT	ADVANTAGES	LIMITATIONS
Amft et al. [2]	Microphone	Inside an earpad	The proposed device reduces ear occlusion. It is viable to be used continuously.	The foam cushion could damp signals in the earpad sensor so the sounds in band 8kHz-16kHz were not captured.
Yang et al. [46]	Piezoelectric	Below outer ear	The proposed method can get meal mass and energy intake.	Limited the long term use as the sensor needs to be attached to the skin and is less socially acceptable.
Liu et al. [24]	Microphone	Outer ear canal	Light-weight and comfortable device that works with Bluetooth	Low detection rate (80%). It uses a camera and can only detect circular bowls and plates.
Gao et al. [19]	Microphone	Outer ear canal	They used off-the-shelve earphones. Detection can be done in many earphones available in the market.	High power consumption due to complicated classification algorithms.

Table 2.2: Comparison of related works - 2

REFERENCE	SENSORS	PLACEMENT	ADVANTAGES	LIMITATIONS
Bi et al. [8]	Microphone	Behind the ear	Low power consumption	The head-mounted device was made with a 3D printer which can be less socially acceptable due to its large size.
Papapanagio et al. [32]	PPG	Earlobe	Its design allows it to be combined with the already explored audio-based chewing sensors.	Physical activities will increase the false positive rate as the sensor works with blood flow.
Passler et al. [34]	Microphone	In-ear and outer ear canal	Low computation costs	It did not consider talking when collecting data.
Steimer et al. [38]	Microphone	In-ear	Can be integrated with the hearing aid, low computation cost	The device does not fit well into the ear.
Bi et al. [9]	Microphone	Neck and throat	Captures high quality signals. Non-invasive	The device needs to be attached to the skin around the throat that could interrupt users' comfort.

Table 2.3: Comparison of related works - 3

REFERENCE	SENSORS	PLACEMENT	ADVANTAGES	LIMITATIONS
Turan et al. [41]	Microphone	Neck and throat	It has a high detection rate of chew and swallow events. It delivers a potential for food intake monitoring in daily life.	The device needs to be attached to the skin around the throat which can interrupt users' comfort.
Wang et al. [42]	Accelerometer (single-axis)	Tempolaris muscle	Non-invasive and preserves privacy	Might interrupt users' comfort. Not socially acceptable due to the placement of the sensor (attach to the face).
Wang et al. [44]	Accelerometer (tri-axis)	Tempolaris muscle	Non-invasive and preserves privacy	Might interrupt users' comfort. Not socially acceptable due to the placement of the sensor (attach to the face).
Farooq et al. [15]	Accelerometer	Temple of eyeglasses	Non-invasive and preserves privacy. It does not require direct attachment.	Low detection rate (F1-score of 87.9)

Table 2.4: Comparison of related works - 4

REFERENCE	SENSORS	PLACEMENT	ADVANTAGES	LIMITATIONS
Sazonov et al. [35]	piezoelectric	Below ear	Non-invasive, low power consumption	Temperature and vibrations affects the performance of the sensor.
Farooq et al. [14]	piezoelectric	Tempolaris muscle	Non-invasive, low power consumption	It negatively affected users' comfort as the sensor needs to be attached to the face with medical tape.
Fontana et al. [18]	Microphone, piezoelectric	Around the neck	Integration of multi-modal, high bandwidth sensor signals and video footage into a single module.	The device interrupts users' comfort as it is attached to the users' neck with a strap.
Fontana et al. [17]	Piezoelectric, accelerometer	Around the collar under ear	The system combines low power multi-modal components and can accurately detect ingestion events.	It uses self-reporting for labeling the ground truth which affects users' behavior throughout the study.
Bi et al. [7]	Microphone, EMG	Behind the ear and neck	Multimodal detection of chewing	Sensors need to be attached to the skin that might interrupt users' comfort.

Table 2.5: Comparison of related works - 5

REFERENCE	SENSORS	PLACEMENT	ADVANTAGES	LIMITATIONS
Papapanagiot et al. [31]	PPG, Microphone, Accelerometer	Earlobe and inside outer ear canal	Low computational costs due to using a PPG sensor that has a low sampling rate.	It can be affected by environmental noise and light. Very precise positioning of sensors are required.
Bedri et al. [6]	Microphone, IMU, proximity	Behind the ear, inside outer ear, back of the neck, around the collar	It can accurately detect and log food intake as well as fast eating behavior.	Proper sensor positioning is required.

3 THEORETICAL BACKGROUND

In this section, we provide an overview of the machine learning and Signal Processing algorithms that have been used for this research. We describe what is known as supervised classification and how it is different from unsupervised and semi-supervised algorithms. We then move into explaining the three supervised classification algorithms that have been used for chewing detection in this research. Finally, we outline the concept of Feature Extraction using Signal Processing and will explain the two Feature Extraction techniques that have been used for this research, namely, Spectral Centroid and Mel-frequency Cepstral Coefficient (MFCC).

3.1 MACHINE LEARNING MODELS

Machine learning algorithms have three main categories, supervised, unsupervised, and semi-supervised. In supervised learning, the machine learning model is trained based on a set of data that had been labeled previously. Unsupervised methods, the algorithm can discover patterns in unlabeled data. It does not require pre-training as in supervised methods. In semi-supervised learning, the data is partially labeled; It falls between supervised and unsupervised learning. In addition, machine learning models fall between two categories of parametric and

non-parametric methods. In parametric methods, the parameters of a particular function are estimated by training. So it has an assumption about the behavior of the data and it estimates a finite set of parameters during the process of training. On the other hand, non-parametric models make fewer assumptions about the distribution of the data and the function to be estimated [10].

1. *Supervised methods*

   Supervised machine learning algorithms have also two categories of multi-class and binary class classifiers.

   Multi-class classifier:

   This method is used when the number of states is specified and the algorithm is trained to detect the state boundary. This approach requires a diversity of training data to cover all possible states. Examples of classifiers used for this method include Decision Trees, Nearest Neighbour, Naïve Bayes, Gaussian Mixture Model (GMM), and Conditional Random Field (CRF) [4].

   Binary classifier:

   Binary classifiers, classify data points into two categories based on the specified classification rules. These classifiers decide as to whether or not a data point has a specific characteristic. For example, in medical applications, the role of binary classifiers is to decide whether a patient with some symptoms has a certain disease [22]. In the case of this research, this method can predict whether a person is eating at the present moment or not. Some commonly-used classifiers for binary classifications are Support Vector Machine, Decision Trees, Random Forest, Logistic Regression, and Bayesian Networks.

2. *Unsupervised methods*

   They are more flexible when dealing with an unexpected variety of data since no pre-training is required. The likelihood ratio method, for example, is based on the idea that two intervals that belong to the same state are likely to have the same probability density [4]. What is common between different unsupervised methods is that they learn and make inferences based on previous data. In other words, they look at the past to predict the future.

3.1.1 *Logistic Regression Classifiers*

Logistic Regression Classifier is a parametric and binary classifier and outputs a prediction for only two categories. It fits training data points to an s-shaped function (Figure 3.1) known as sigmoid (Eq. 3.1) to estimate the parameters in a way that the error is minimized.

$$\mathrm{Sigmoid}(x) = \frac{1}{e^{-x} + 1} \tag{3.1}$$

By taking a minimum value of 0 and a maximum value of 1, the sigmoid function suits the purpose of binary classification. After fitting a sigmoid function to a set of observations and estimating the parameters, the sigmoid function gives a value between 0 and 1 for a new observation x. So it can give a probability for whether the new data point belongs to a category or class 1 or 2.

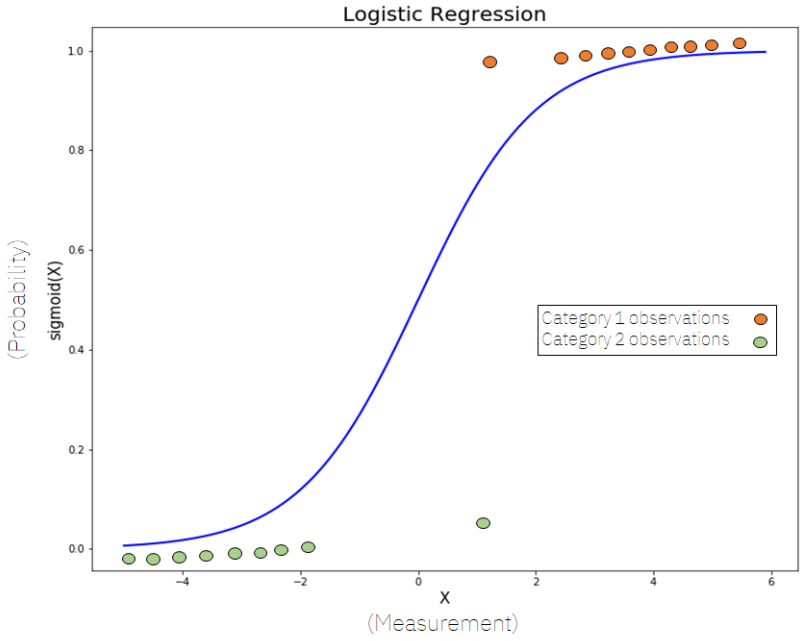

Figure 3.1: Logistic Regression classifier: the sigmoid function estimated into two sets of observations [49]

3.1.2 *Decision Tree Classifiers*

Decision Trees are non-parametric machine learning models the components of which are nodes, edges, and leaves where nodes represent an attribute, edges represent a specific value of the attribute which is connected to the edges and leaves are the outputs. For example, in Figure 3.2 the Decision Tree predicts whether a person is fit by splitting the attribute *age* into two categories less or more than 30 years old. The next split happens for attribute *exercising in the mornings* and *eating a lot of pizza* which both have two categories of *yes* or *no*.

The split of an attribute value is made by calculating the entropy (Eq. 3.2) within each split region. Where p_{mk} represents the percentages of data in the m'th region

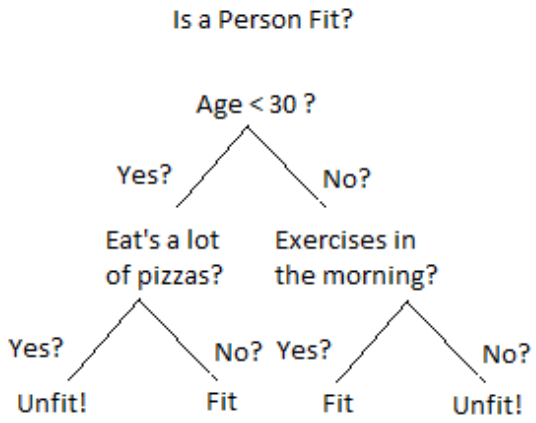

Figure 3.2: An example of a Decision Tree that predicts whether a person is fit [11]

that is from the k'th class. It can be shown that the value of entropy is near zero if p_{mk}'s are either near zero or near one.

$$H(X) = -\sum_{k} p_{mk} \log p_{mk} \tag{3.2}$$

3.1.3 *Random Forest Classifiers*

Random Forest classifiers consist of multiple Decision Trees in which each tree is trained with a random set of data points and a random set of variables. The final prediction of Random Forest is the average or median of the prediction of Decision Trees within the forest (Figure 3.3). Therefore, it reduces the over-fitting. Same as decision trees, the Random Forest is a non-parametric method but they are stronger modeling techniques compared to decision trees as they use the concept of *the wisdom of the crowds*.

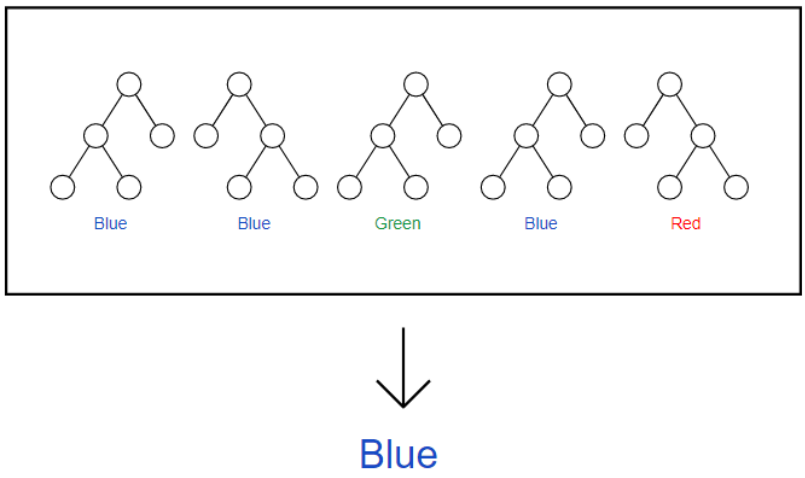

Figure 3.3: An example of a Random Forest: the final output is blue as the majority of the Decision Trees voted for blue [48]

3.2 SIGNAL PROCESSING FEATURE EXTRACTION

Feature extraction is the process of transforming raw data into a smaller and manageable input for processing. In this research, the raw data points are in the form of a signal. Generally speaking, there are two types of feature extraction in digital signal processing:

- Time-domain features: the time-domain features interpret how the signal behaves over time. These features are easy to interpret and can be obtained by simple methods like calculating the signals average, variance, power, zero crossings, etc.

- Frequency-domain features: the frequency-domain features measure how the signals behave over a range of frequencies (Figure 3.4). These features are obtained after transforming the time-domain data into the frequency-domain using a Discrete Fourier Transform. The mapping of f_j, $j = 0, \ldots, N-1$,

into $c_k = \frac{1}{N}\sum_{j=0}^{N-1} f_j e^{-ijk2\pi/N}$, $k = 0, \ldots, N-1$, is called the *Discrete Fourier Transform* (DFT).

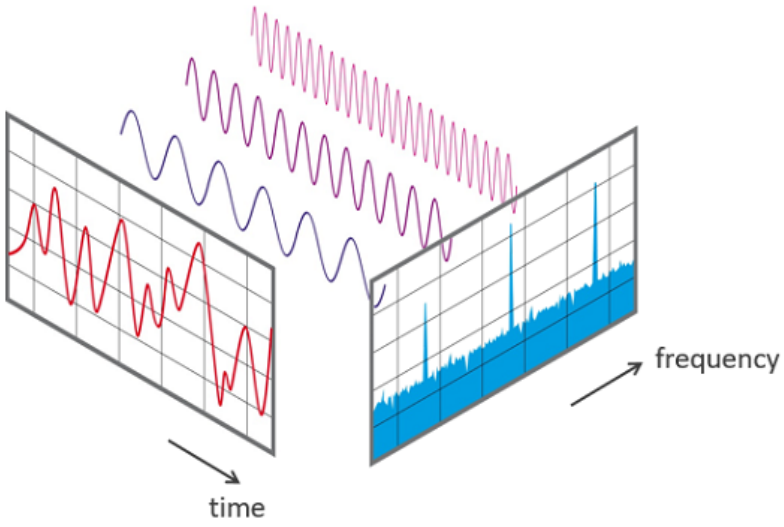

Figure 3.4: Representation of a signal in time and frequency-domain [5]

The frequency-domain features might be harder to interpret, however, they provide useful information such as the pitch and the melody in an audio signal.

In this research, we use mean, variance, and power as time-domain features. For frequency-domain features, we use the two commonly-used signal processing features, namely Spectral Centroid and Mel-frequency Cepstral Coefficients. In the following section, we explain each of the methods briefly.

3.2.1 *Spectral Centroid*

Spectral Centroid is a measure that characterizes the center of the mass in the power spectrum. The higher value of SC means the high-frequency constituent

components of a signal are dominant, in other words, the more energy of the signal is concentrated around higher frequencies. It is defined as follows, where N is the number of bins, f and M[f] are the frequency and magnitude of frequency f, respectively.

$$\text{Centroid} = \frac{\sum_{f=0}^{N} fM[f]}{\sum_{f=0}^{N} M[f]} \tag{3.3}$$

3.2.2 *Mel-frequency Cepstral Coefficient*

Mel-frequency Cepstral Coefficient (MFCC) [25] is a widely-used feature in digital signal processing for different tasks including, speech and speaker recognition. In simple words, Mel-scales converts actual frequencies to what a human can hear. The human brain can better understand the changes in frequencies when the frequency of a sound is lower. For example, the difference between sounds produced in 400Hz and 500Hz is easier to perceive than sounds produced by 1000Hz and 1100Hz even though the difference between both frequencies is 100Hz. The non-linear transformation of frequencies was obtained by Volkmann and Newmann in 1937 (Eq. 3.4) where f is the frequency.

$$Mel(f) = 2595 \times \log(1 + \frac{f}{700}) \tag{3.4}$$

So in the MFCC feature extraction, the signal is windowed into short frames (assuming the audio does not change in short frames), the power spectrum of each short frames are obtained by calculating DFT, Mel spectrum is obtained by applying Mel filter bank on the power spectrum (weighted sum of the spectrum

and Mel filter bank 3.5) and finally, the Mel spectrum is reversed back to time-domain using Discrete Cosine Transform.

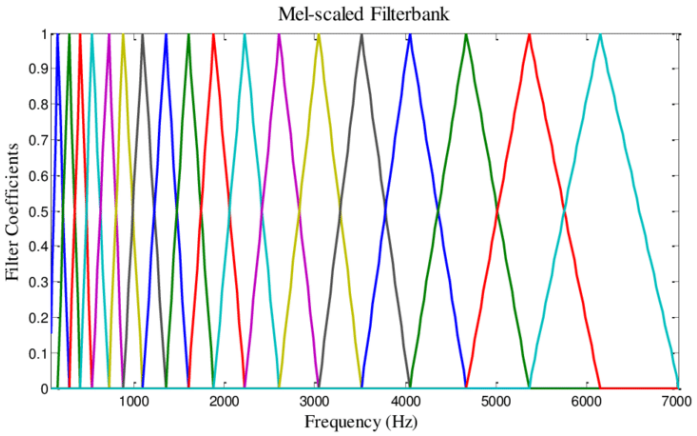

Figure 3.5: Mel filter bank basis functions using 20 Mel filters [47]

4 PROPOSED METHODOLOGY FOR CHEWING DETECTION

We used a machine learning Pipeline [13] for chewing detection. First, we introduce the pipeline consisting of two steps of Feature Extraction and Classification based on solely audio data, after which we propose a pipeline for IMU data and finally we merge two pipelines to utilize both IMU and audio data to detect chewing.

4.1 CHEWING DETECTION BASED ON AUDIO DATA

This section describes the pipeline we used to detect chewing activities based on audio data. As we pointed out previously, we record audio data with a 48 kHz sampling rate. We framed the signal into 3-seconds non-overlapping time windows. For the feature extraction phase, we divide the 3-second time frame into 45 partitions and calculated Spectral Centroid [20] (Eq. 3.3) for each partition. We implemented the following steps on the audio signal to extract Spectral Centroid as a feature vector:

1. Partition the signal into short 3-second frames.
2. Calculate the Discrete Fourier Transform [37] for each 3-second frame and obtain the Spectrum of each frame.

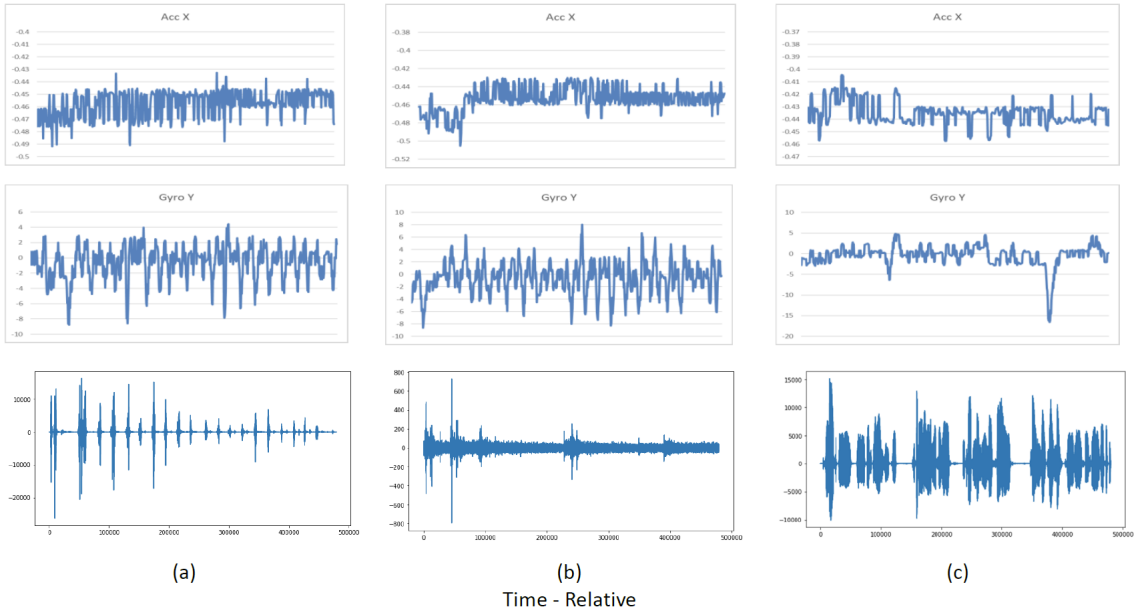

Figure 4.1: Sample of IMU and audio signals (a): chewing pretzel (b): chewing banana (c): speaking

3. Calculate the Spectral Centroid (Eq. 3.3) and return the 45-dimensional feature vector representing the 3-second time windows.

The final output of the algorithm is a 45-dimensional feature vector. To demonstrate the effectiveness of spectral centroid to distinguish chewing and non-chewing activities, we plotted the mean and standard deviation of the 45-dimensional feature vector on a subset of audio recordings.

As Figure 4.2 shows, there is a slight overlap between two classes of chewing and non-chewing activities although the separation is visually observable. Therefore, we decided to use this method for feature extraction step.

As well as Spectral Centroid, we used MFCC to extract features of the audio signals. We extracted 45 cepstral coefficients of each 3-second time window using the MFCC algorithm. The steps of this algorithm are:

1. Partition the signal into 3-second frames.

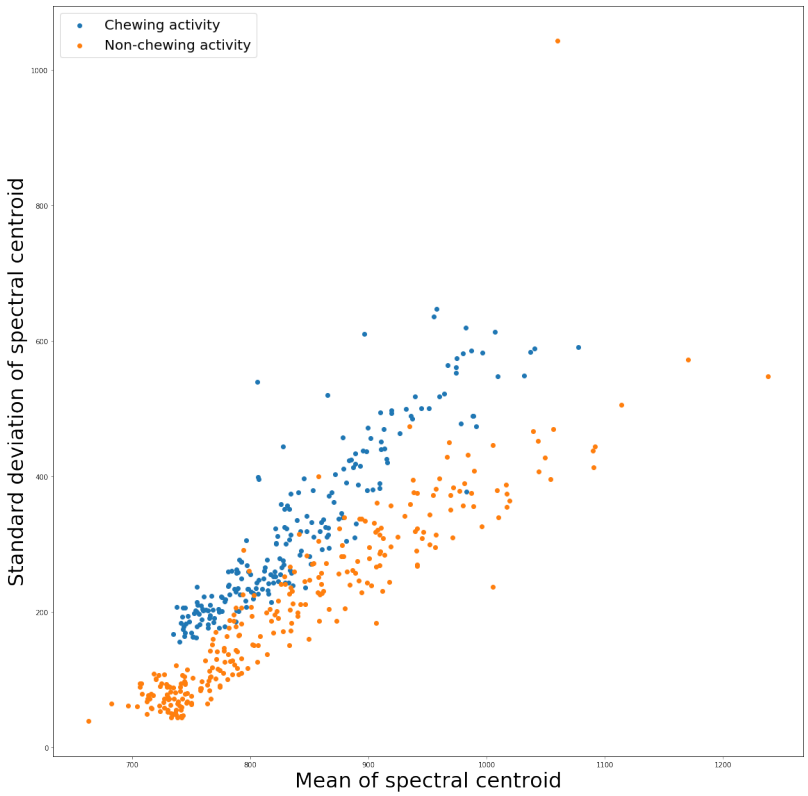

Figure 4.2: Visual exploratory analysis of spectral centroid of chewing and non-chewing audio data

2. Calculate the Discrete Fourier Transform for each 3-second frame.

3. Apply Mel filter bank [39] with 45 filters on the magnitude spectrum obtained from step 2 and extract a 45-dimensional feature vector.

4. Calculate the logarithm of the filterbanks obtained from step 3.

5. Apply the Discrete Cosine Transform [1] of the logarithm of the filter bank obtained in step 4.

We selected a few samples of audio data and visualized the MFCC coefficients with the aforementioned specification to get a sense of whether this feature can be useful to detect chewing audio data from non-chewing. In other words, we

explored the visual representation of MFCC to figure out how likely it is for classifiers to detect and distinguish two classes of chewing and non-chewing successfully. Figures 4.3, 4.4, and 4.5 show audio signal, MFCC coefficients, and the normalized MFCC coefficients where the mean and variance values for each coefficient dimensions are set to zero and one, respectively.

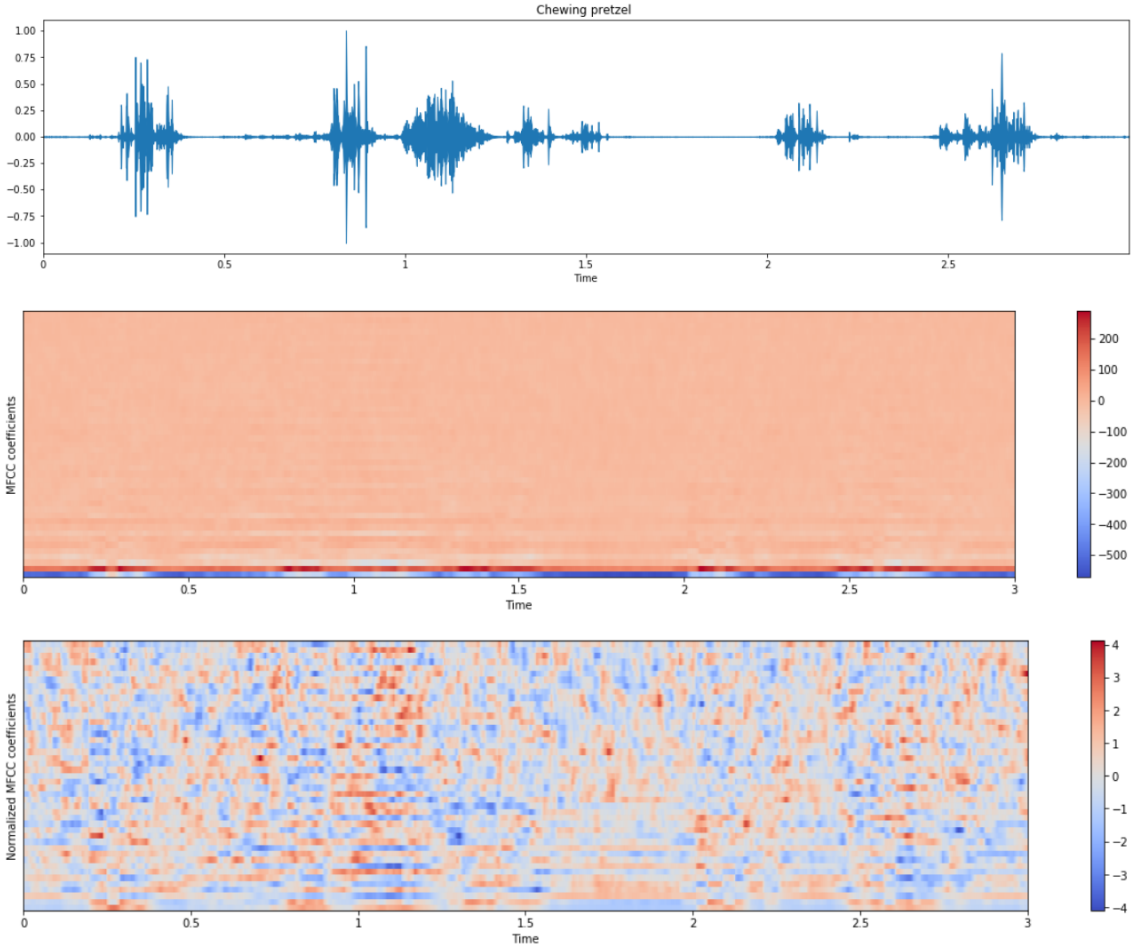

Figure 4.3: Visual exploratory analysis of MFCC coefficients of the audio recording while chewing pretzel

The 2-dimensional visualization of MFCC coefficients with the x-axis representing the time and y-axis representing the coefficients did not provide a good indication of the difference between chewing and other activities. To simplify the

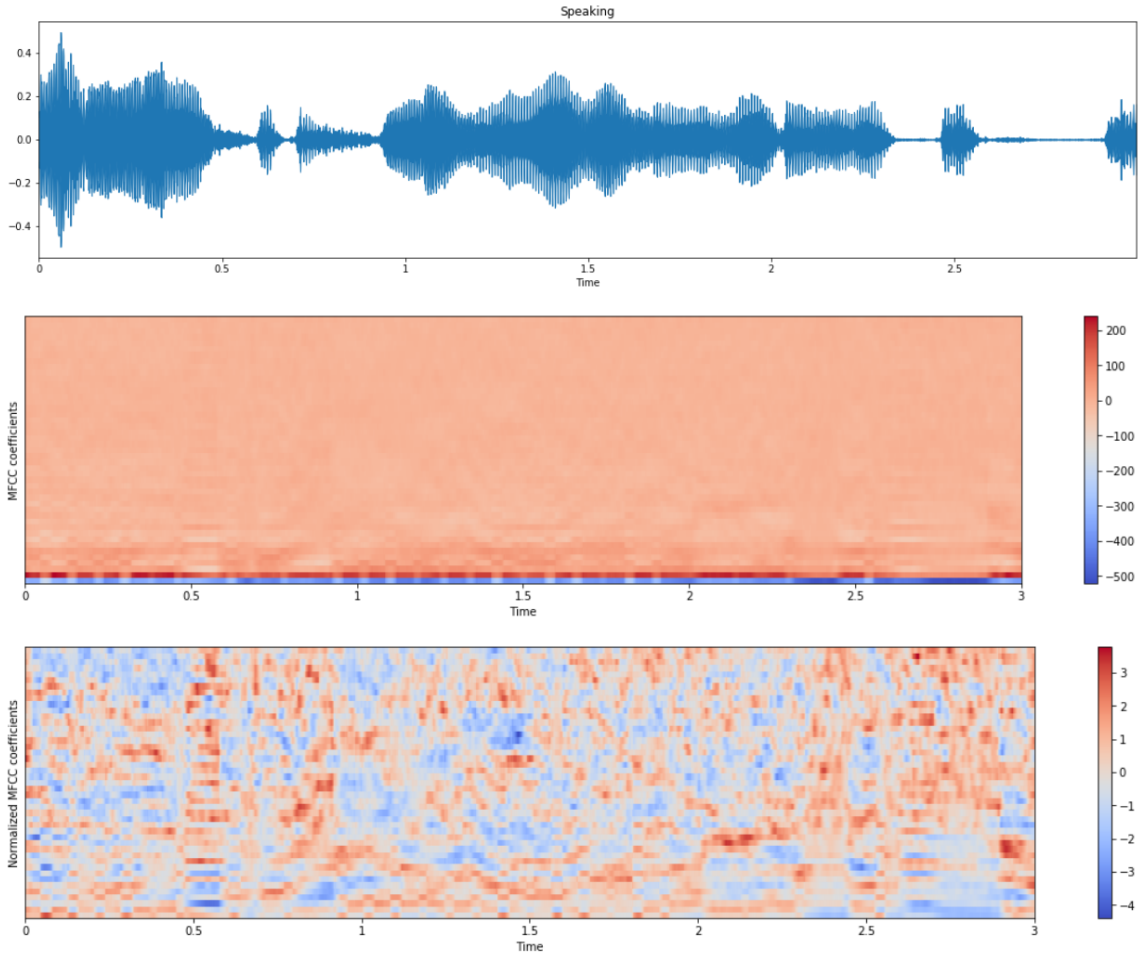

Figure 4.4: Visual exploratory analysis of MFCC coefficients of the audio recording while speaking

visualization, we took the mean value of the MFCC coefficients over time. The result is shown in Figure 4.6. This figure shows a clear distinctive difference between speaking and watching movies with chewing audio data. It can be seen that chewing has a more dominant fluctuation which corresponds to each individual crushing of food with teeth while chewing.

Figure 4.6 indicated that using MFCC is a good feature to differentiate between chewing and non-chewing activities. Although taking the average of the coefficients

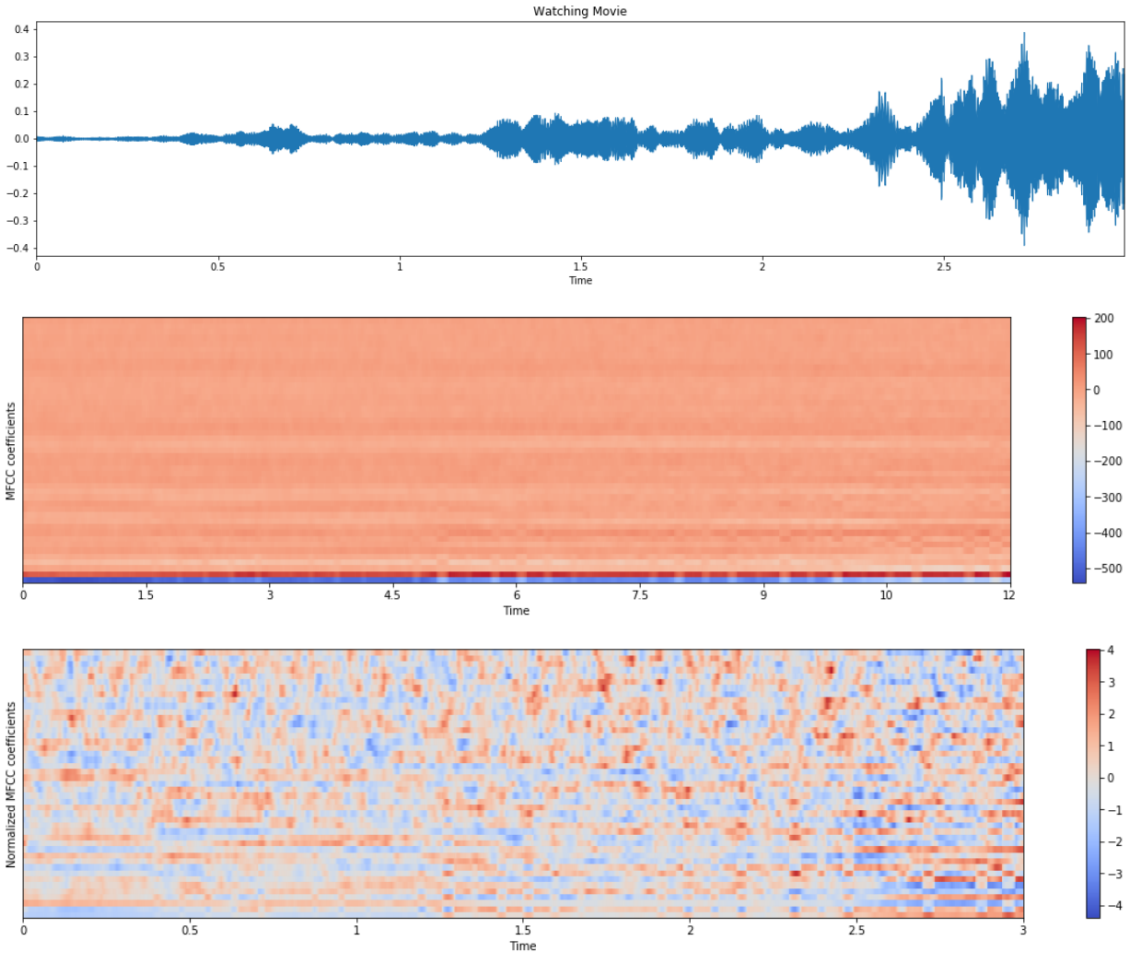

Figure 4.5: Visual exploratory analysis of MFCC coefficients of the audio recording while watching movie

removes a lot of information, there is enough left to distinguish the two activity sets.

We obtain a 90-dimensional feature vector in which the first 45 elements of it are calculated based on Spectral Centroid and the second 45 elements are MFCC coefficients. We empirically examined MFCC and Spectral Centroid and decided to use them as features since they had promising results. We used the obtained 90-dimensional feature vector to train three classifiers: Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF) [22]. Figure 4.7 shows this pipeline.

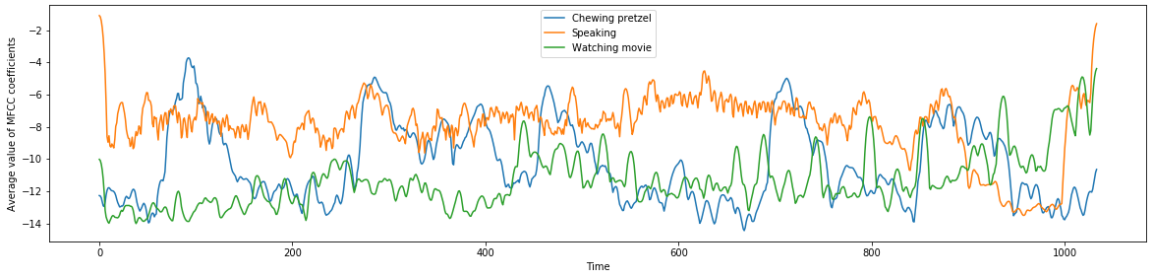

Figure 4.6: The average value of MFCC coefficients corresponding to figures 4.3, 4.4, 4.5

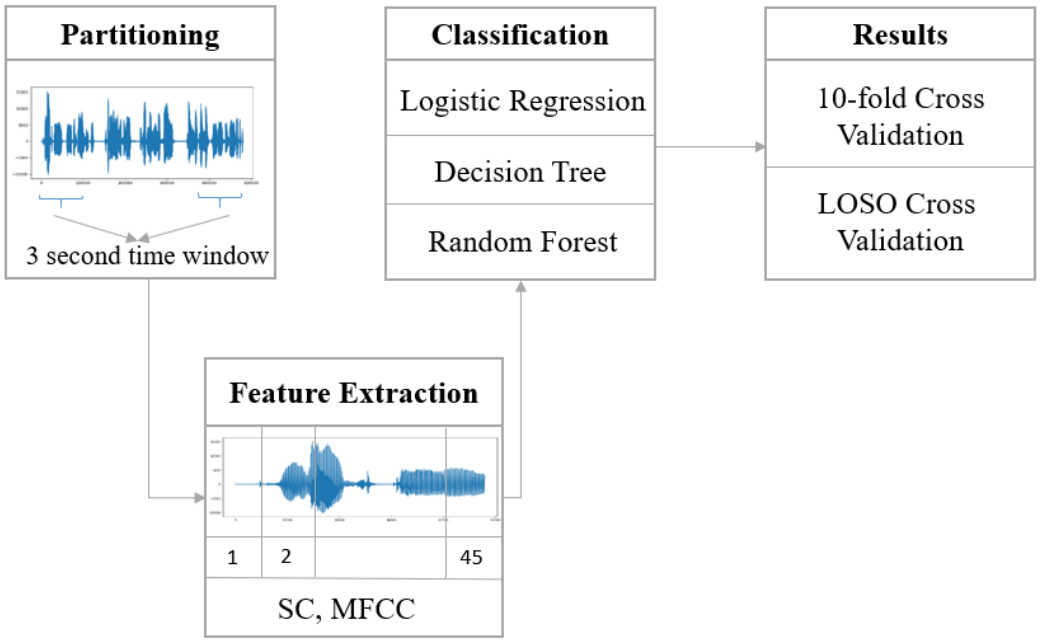

Figure 4.7: Pipeline for chewing detection based on audio data. The pipeline consists of feature extraction and classification.

4.2 CHEWING DETECTION BASED ON IMU DATA

Both time and frequency-domain features were extracted from signals collected from the IMU sensor. We set the sampling rate to 60 Hz for collected IMU data. Time-domain features are the mean, variance, and power of each of the 6 axes of the IMU signals over the 3-second time window (total of $6 \times 3 = 18$ features).

The frequency-domain features are the Spectral Centroid over the 3-second time windows for every 6 axes of the IMU sensor, from which a total of 6 values is obtained. To certify that spectral centroid suits the classification of chewing and non-chewing based on IMU, we took the mean and standard deviation over 6 values. This result is shown in Figure 4.8 which the boundary between two classes of chewing and non-chewing activities is separable.

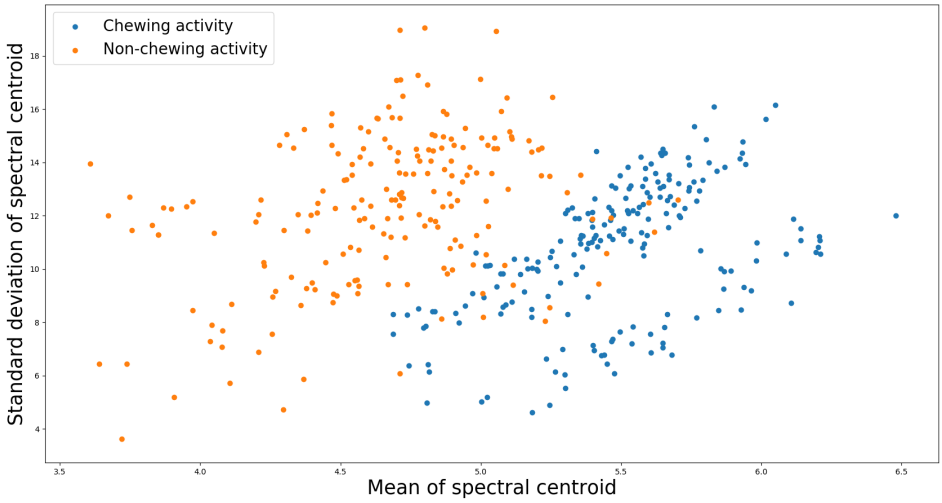

Figure 4.8: Visual exploratory analysis of spectral centroid of chewing and non-chewing IMU data

MFCC with 12 coefficients over the 3-second time windows was obtained over 6 axes of the IMU (total of $12 \times 6 = 72$ features). Although MFCC was designed for audio analysis, we tried it empirically on IMU and the result was satisfactory. Since the sampling rate of IMU data is 60 Hz, we chose 12 coefficients; meanwhile, for audio data, we chose a larger number of coefficients of 45. The obtained 96-dimensional feature vector was then used to train classifiers that were also used to detect chewing by audio. These classifiers are Logistic Regression, Decision Tree, and Random Forest. Figure 4.9 shows this pipeline.

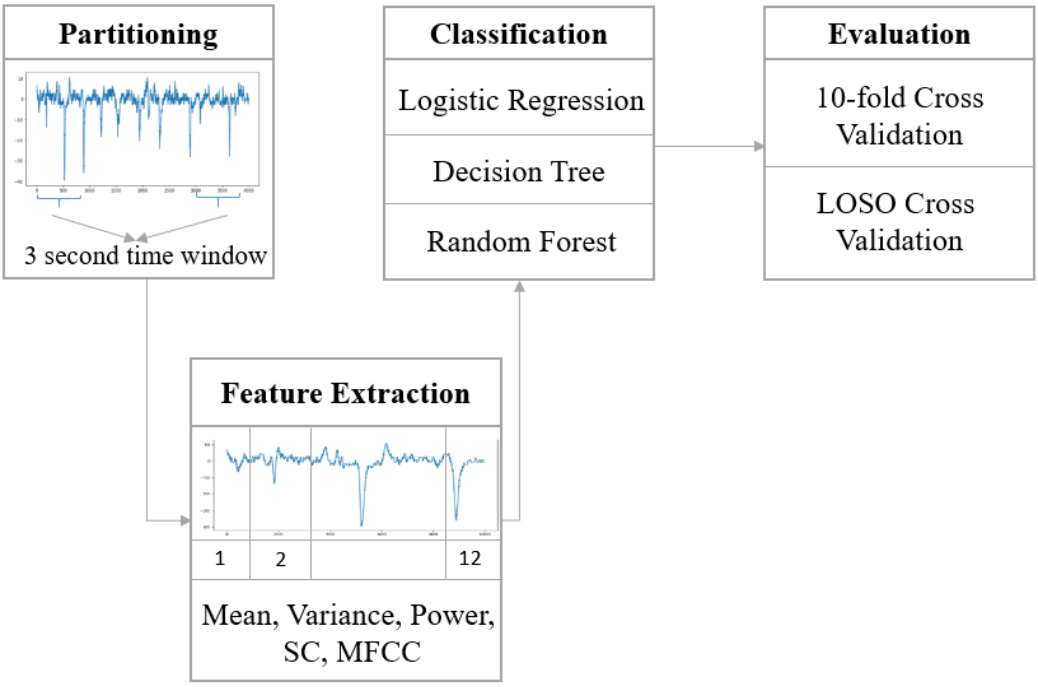

Figure 4.9: Pipeline for chewing detection based on IMU data. The pipeline consists of feature extraction and classification.

4.3 CHEWING DETECTION BY FUSING IMU AND AUDIO DATA

To measure the performance of fusing both sensing modalities to detect chewing, the extracted 96-dimensional feature vector from the IMU and 90-dimensional feature vector from audio were concatenated. The obtained 186-dimensional feature vector was fed into three classifiers namely, Logistic Regression, Decision Tree, and Random Forest. In Section 6, we discuss the performance of all three methods for detecting chewing based on several evaluation techniques.

5 DATASET CHARACTERISTICS AND COLLECTION METHOD

This section describes the data collection method, procedure, as well as dataset characteristics. The collected dataset is based on *common sedentary activities*. We defined *common sedentary activities* as the set of sedentary activities that most individuals perform throughout the day. We did not involve activities that require much body movements since we are concerned with detecting chewing from head-related activities.

The eSense device transmits IMU data via BLE and audio data via Bluetooth. We developed an android application (Figure 5.2) to collect accelerometer, gyroscope, and audio signals from the eSense device. We set the sampling rate to 48 kHz for audio and 60 Hz for IMU. We used a Samsung Galaxy S9 phone to record

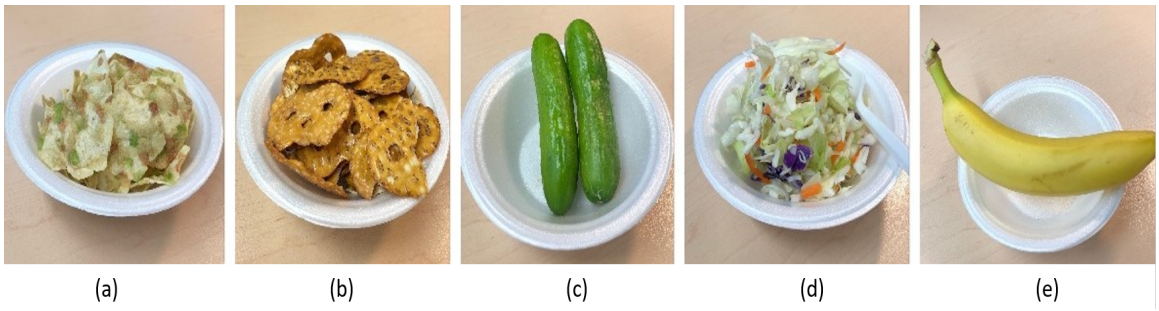

Figure 5.1: Foods from left to right: crispiest to the softest. (a): Chips, crispy (b): Pretzels, crispy (c): Cucumber, crispy and juicy (d): Salad, crispy and juicy (e): Banana, soft

data locally on the phone memory. The developed application has three main buttons on the screen, one for connecting the mobile application to the eSense device via Bluetooth. After connecting the device, the *STARTRECORDING* button activates. Pressing this button will start recording audio data in a PCM file and the IMU data in a CSV file in a real-time manner. After pressing this button, the *STOPRECORDING* button activates and recordings can be terminated by pressing this button.

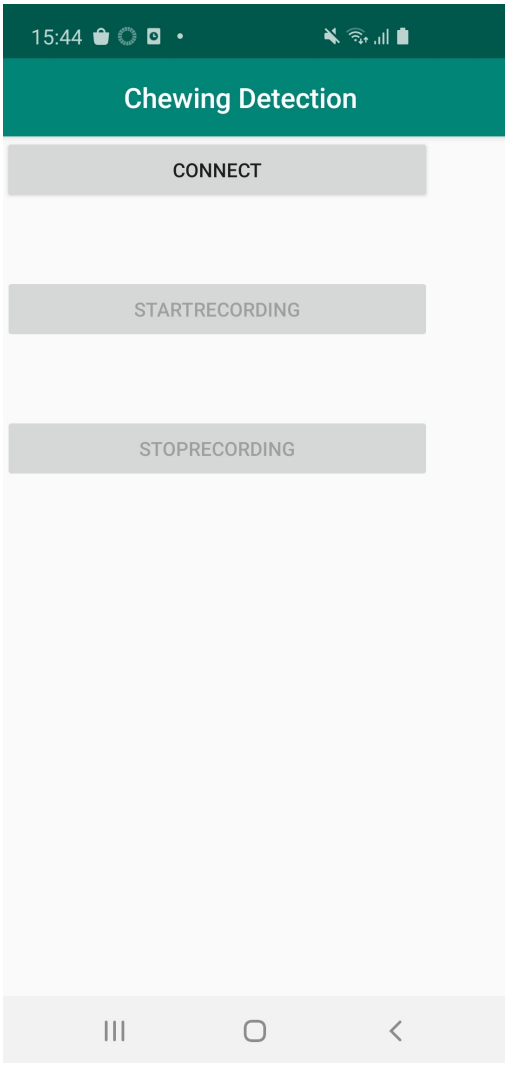

Figure 5.2: The developed android application to record audio and IMU data from eSense.

We recorded the participants' videos throughout the study in order to label the ground truth of each activity performed by them. Figure 5.3 shows the angle of the camera for recording participants. We labeled a sequence of data as chewing when the participant puts the food inside the mouth and initiates chewing. To synchronize the eSense sensor recordings and video recordings, we asked participants to nod three times while saying "recording started."

Figure 5.3: Participant while wearing the eSense device and performing the chewing activity

We conducted the recording experiments in the form of a pilot study at a local university, within a lab environment with a total of 5 subjects (4 male, 1 female, age 25-29). Participants were required to eat each of the foods mentioned in Table 5.1 for 150 seconds. We used different types of foods with different textures as different sounds are produced when chewing different food textures. Figure 5.1 demonstrated the foods that we used in this study. In addition, food texture contributes to a different level of vibrations on the face, mouth, and teeth, and these vibrations can be detected in both IMU and audio signals [21]. Figure 4.1 shows a sample of signals over 3-second time windows, collected from IMU and audio data

Table 5.1: The foods consumed by participants in the recording experiment - Sorted by the level of crispiness - Top to bottom: crispiest to the softest

Eating Activities	Duration
Chips	150 seconds
Pretzel	150 seconds
Cucumber	150 seconds
Salad	150 seconds
Banana	150 seconds

while chewing 2 types of foods and speaking. The participants were also required to speak, read, drink, and watch a short movie. The duration of each activity is mentioned in Table 5.2. The rationale behind the aforementioned activities was to explore whether chewing sounds and motions are distinguishable from other mouth and head-related activities as well as not having any activities. Overall, each participant spent 13 minutes eating and 13 minutes on non-eating activities. After finishing the recording experiment, we encountered some missing points while recording both IMU and audio. This was due to Bluetooth disconnection which happened every once in a while. We removed those missing values. The expected number of data points for each set of chewing and non-chewing was 1300 data points. However, by removing the missing values, we got 820 data points for chewing and 905 data points for non-chewing activities. The pie chart in Figure 5.4 shows the percentage of class chewing and non-chewing. The number of data points in each class are approximately the same so we conclude that the dataset is balanced.

We used the collected data mentioned in this step to analyze the performance of chewing detection based on different modalities of the eSense device.

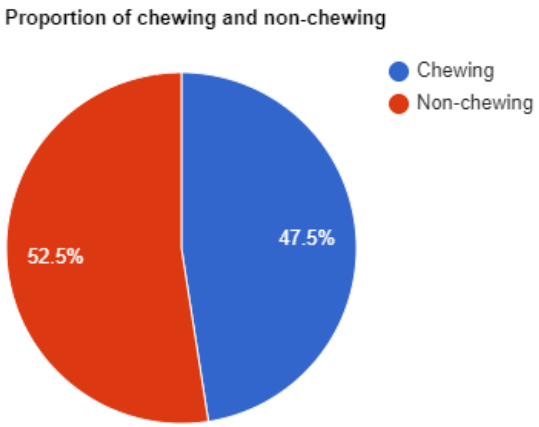

Figure 5.4: Pie chart visualization of the proportion of two classes of chewing and non-chewing

Table 5.2: Sedentary activities performed by the participant in the recording experiment

non-eating Activities	Duration
Speaking/Reading	3 minutes
Watching a short movie	6 minutes
Drinking	2 minutes
Sitting still	2 minutes

6 EVALUATION

In order to compare the performance of each of the sensor data and each detection algorithm mentioned in Section 4, we use 10-fold and Leave-One-Subject-Out (LOSO) Cross-Validation [22] to calculate 4 evaluation metrics: Accuracy (Eq. 6.1), Precision (Eq. 6.2), Recall (Eq. 6.3), and F1-score (Eq. 6.4) which are defined as follows:

$$Accuracy = \frac{TP+TN}{TP + FP + TN + FN} \tag{6.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{6.2}$$

$$Recall = \frac{TP}{TP+FN} \tag{6.3}$$

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{6.4}$$

6.1 EVALUATION OF CHEWING DETECTION BASED ON AUDIO DATA

Tables 6.1 and 6.2 show the result of chewing detection by feeding the audio data into the pipeline mentioned in Section 4.1. The results represent that the Random Forest classifier outperformed both Logistic Regression and Decision Tree based on both 10-fold and LOSO evaluation methods. Figures 6.1 and 6.2 shows the confusion matrix of the best classifier, Random Forest, for two methods of evaluation, 10-fold and LOSO.

Table 6.1: Evaluation of chewing detection using 10-fold cross-validation on audio data

Method	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.85	0.77	0.75	0.76
Decision Tree	0.88	0.80	0.81	0.80
Random Forest	**0.94**	**0.85**	**0.95**	**0.90**

Table 6.2: Evaluation of chewing detection using LOSO cross-validation on audio data

Method	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.81	0.75	0.68	0.68
Decision Tree	0.82	0.62	0.77	0.68
Random Forest	**0.91**	**0.74**	**0.94**	**0.82**

6.2 EVALUATION OF CHEWING DETECTION BASED ON IMU DATA

The result of detecting chewing based on IMU data and the pipeline in Section 4.2 is presented in Tables 6.3 and 6.4. These results suggest that both Logistic

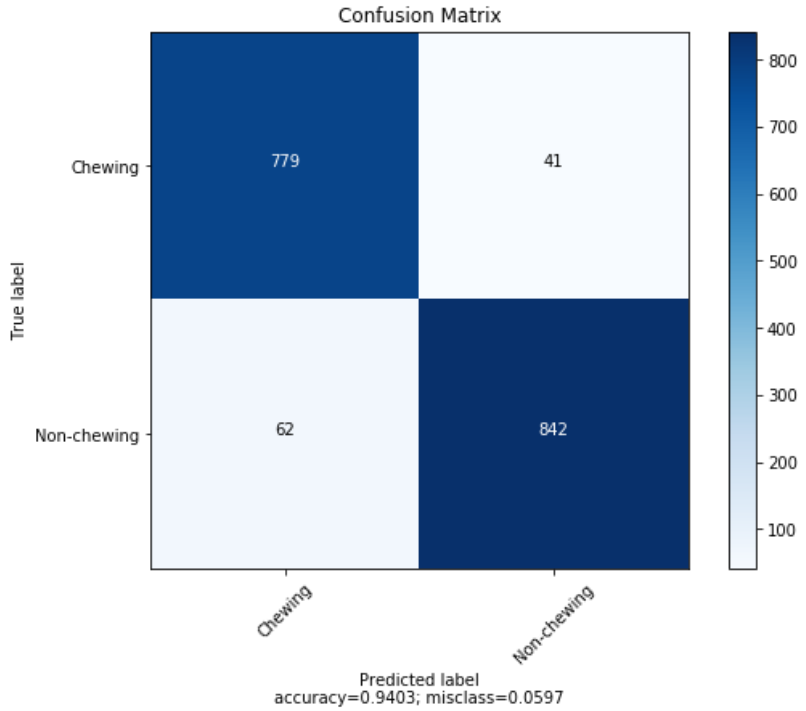

Figure 6.1: Confusion matrix of the Random Forest classifier corresponding to Table 6.1

Regression and Random Forest perform well in detecting chewing. However, the Decision Tree approach did not perform as well as the two other classification algorithms. This result was predictable since the Random Forest classifier is a collection of Decision Trees and can perform at least as good as the Decision Tree method. We represented the confusion matrix corresponding to the Random Forest classifier when evaluation with both 10-fold and LOSO in Figures 6.3 and 6.4.

Table 6.3: Evaluation of chewing detection using 10-fold cross-validation on IMU data

Method	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.93	0.89	0.88	0.88
Decision Tree	0.89	0.82	0.83	0.82
Random Forest	**0.95**	**0.90**	**0.92**	**0.91**

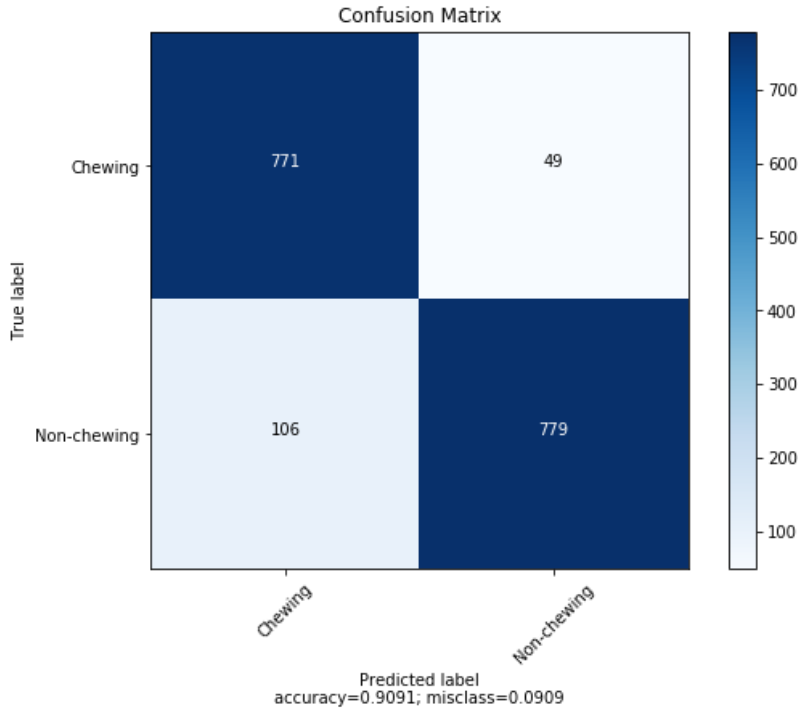

Figure 6.2: Confusion matrix of the Random Forest classifier corresponding to Table 6.2

Table 6.4: Evaluation of chewing detection using LOSO cross-validation on IMU data

Method	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.94	0.90	0.91	0.90
Decision Tree	0.88	0.78	0.81	0.80
Random Forest	**0.94**	**0.87**	**0.92**	**0.89**

6.3 EVALUATION OF CHEWING DETECTION BASED ON FUSING IMU AND AUDIO DATA

Finally, the performance of fusing both sensing modalities of IMU and audio based on 10-fold and LOSO Cross-Validation is summarized in Tables 6.5 and 6.6. As the result suggests, Random Forest with an accuracy of 97% performed better than Logistic Regression and Decision Tree. Moreover, the performance of fusing

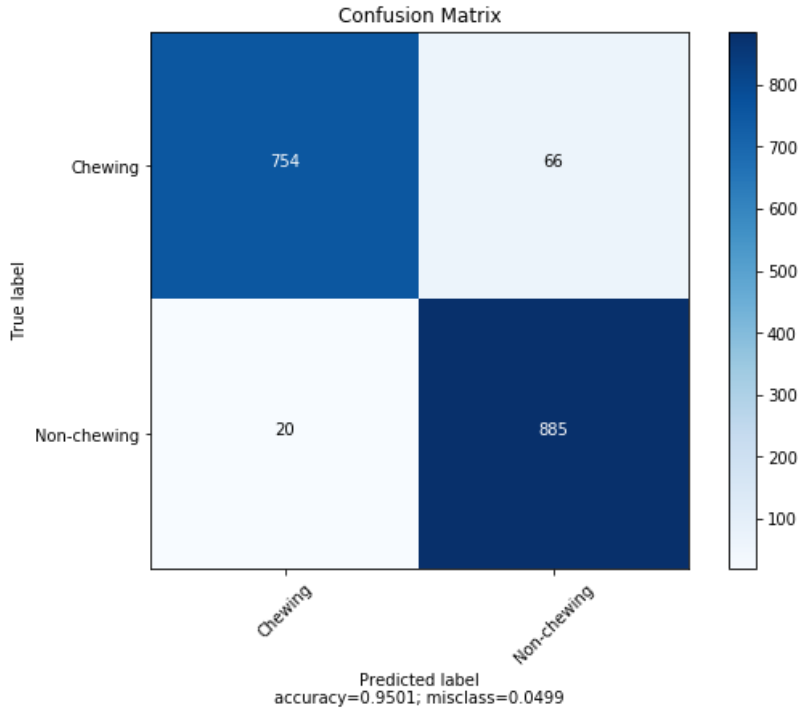

Figure 6.3: Confusion matrix of the Random Forest classifier corresponding to Table 6.3

both modalities resulted in the highest Accuracy, Precision, Recall, and F1-score compared to using single modalities. The confusion matrix for evaluating based on 10-fold and LOSO when detecting chewing by fusing both modalities are brought in Figures 6.5 and 6.6.

Table 6.5: Evaluation of chewing detection using 10-fold cross-validation after fusing both audio and IMU sensors

Method	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.93	0.88	0.88	0.88
Decision Tree	0.92	0.87	0.85	0.86
Random Forest	**0.97**	**0.94**	**0.96**	**0.95**

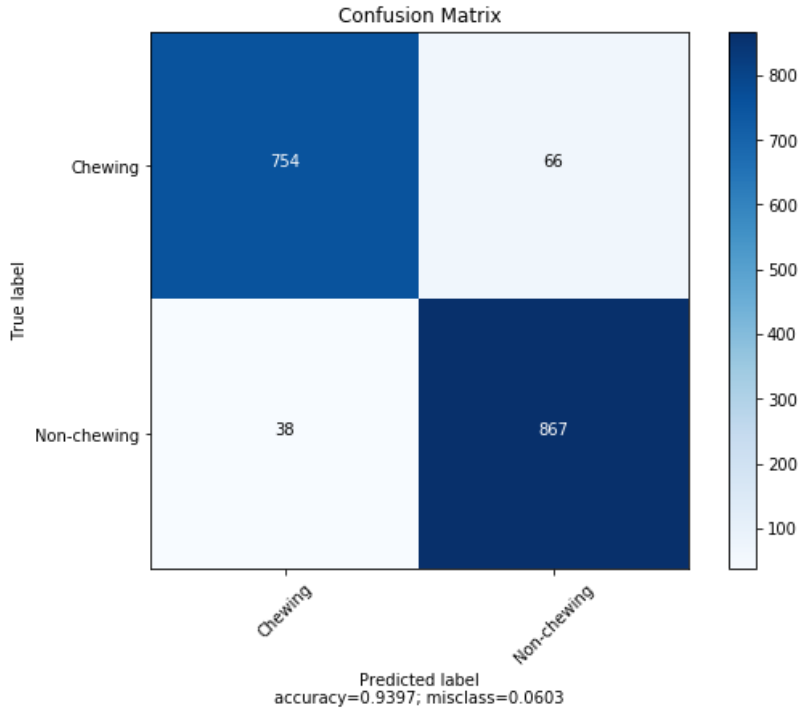

Figure 6.4: Confusion matrix of the Random Forest classifier corresponding to Table 6.4

Table 6.6: Evaluation of chewing detection using LOSO cross-validation after fusing both audio and IMU sensors

Method	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.95	0.89	0.93	0.91
Decision Tree	0.89	0.80	0.83	0.81
Random Forest	**0.97**	**0.94**	**0.96**	**0.94**

6.4 COMPARISON OF SENSING MODALITIES OF IMU AND AUDIO

Figure 6.7 and 6.8 shows the comparison between three combinations of sensing modalities using 10-fold and LOSO evaluations. It can be seen that IMU data outperformed audio data in detecting chewing. This can be due to the fact that soft foods might not be well captured by the microphone since the chewing sound

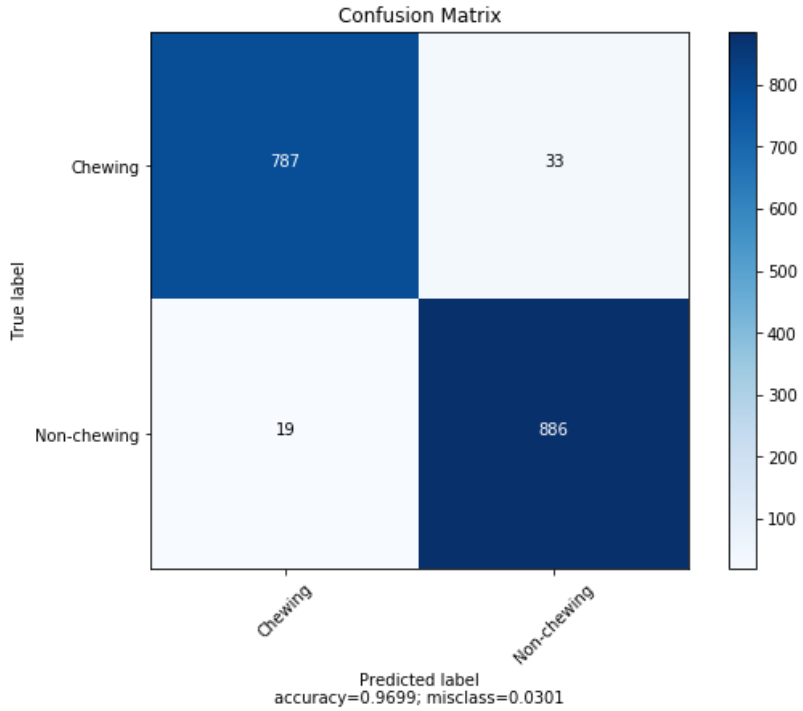

Figure 6.5: Confusion matrix of the Random Forest classifier corresponding to Table 6.5

for soft foods is very quiet. To infer the reason behind this result, we removed the soft food that was used in this experiment and trained and evaluated the Random Forest Classifier based on the three pipelines identified in Section 4. We realized that removing soft foods improves the detection rate of audio data by 4% but does not significantly improve the detection based on IMU data. We concluded that in spite of higher data resolution, the microphone sensor does not perform well for detecting chewing while the subjects are having soft foods. However, IMU data is more robust against the texture of the food and can detect both soft and crispy foods. The improvement of detection of chewing after removing soft foods is shown in Figures 6.9 (10-fold evaluation) and 6.10 (LOSO evaluation).

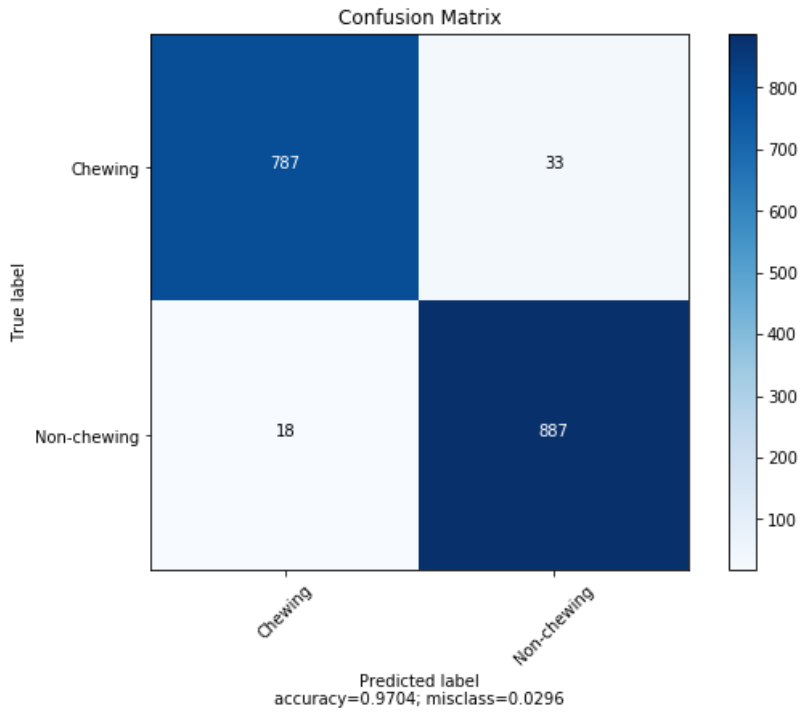

Figure 6.6: Confusion matrix of the Random Forest classifier corresponding to Table 6.6

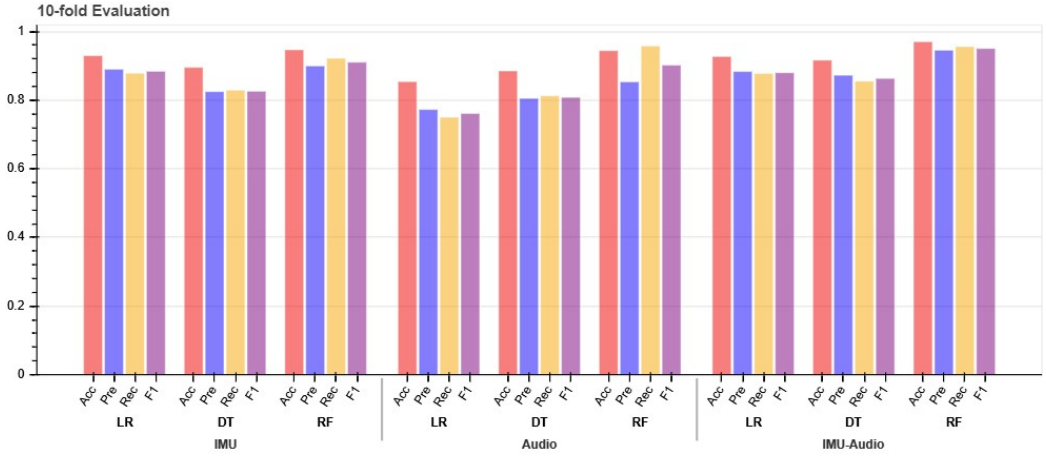

Figure 6.7: 10-fold cross-validation on detection of chewing based on IMU, audio, and combining IMU and audio. LR: Logistic Regression, DT: Decision Tree, RF = Random Forest

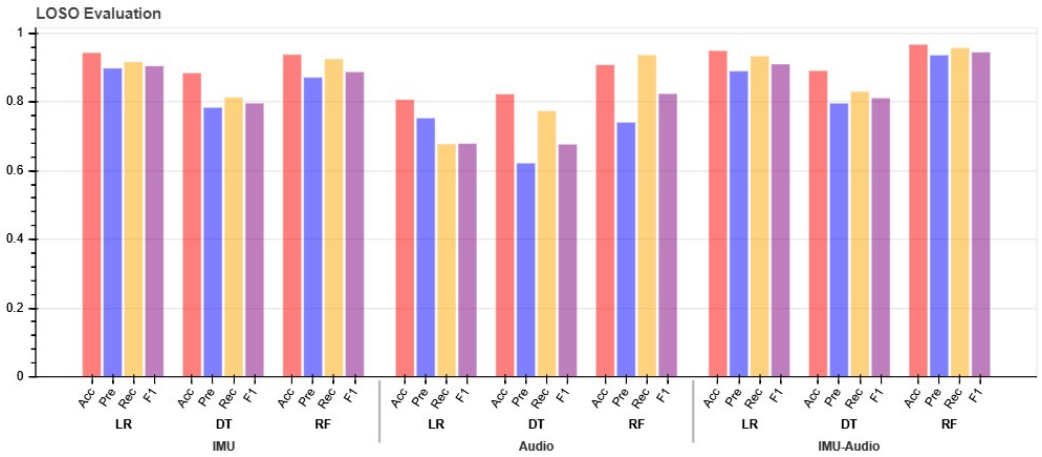

Figure 6.8: LOSO cross-validation on detection of chewing based on IMU, audio, and combining IMU and audio. LR: Logistic Regression, DT: Decision Tree, RF = Random Forest

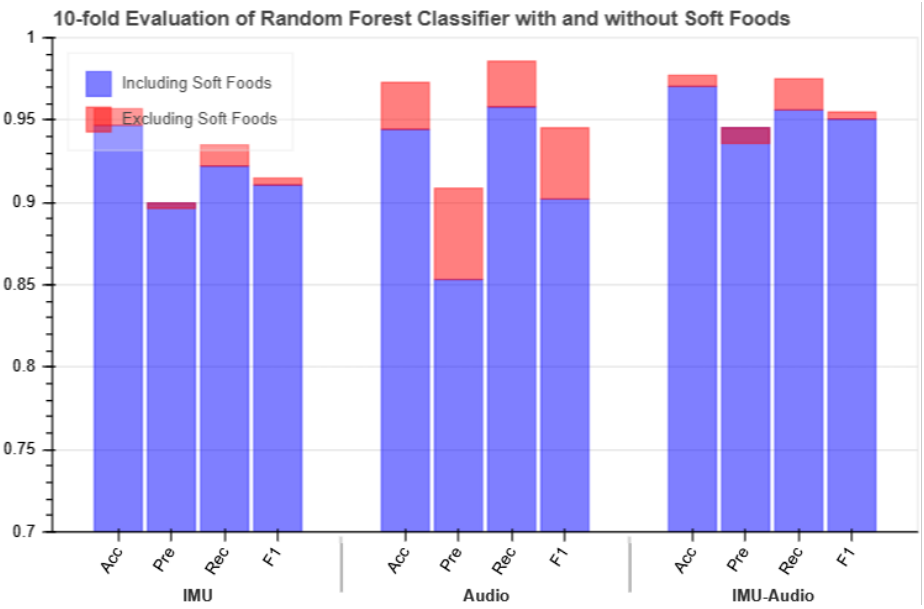

Figure 6.9: 10-fold evaluation of the Random Forest classifier after removing soft foods. As can be seen, removing soft foods improved the performance of audio data but did not significantly improve the performance of IMU and fusing IMU and audio

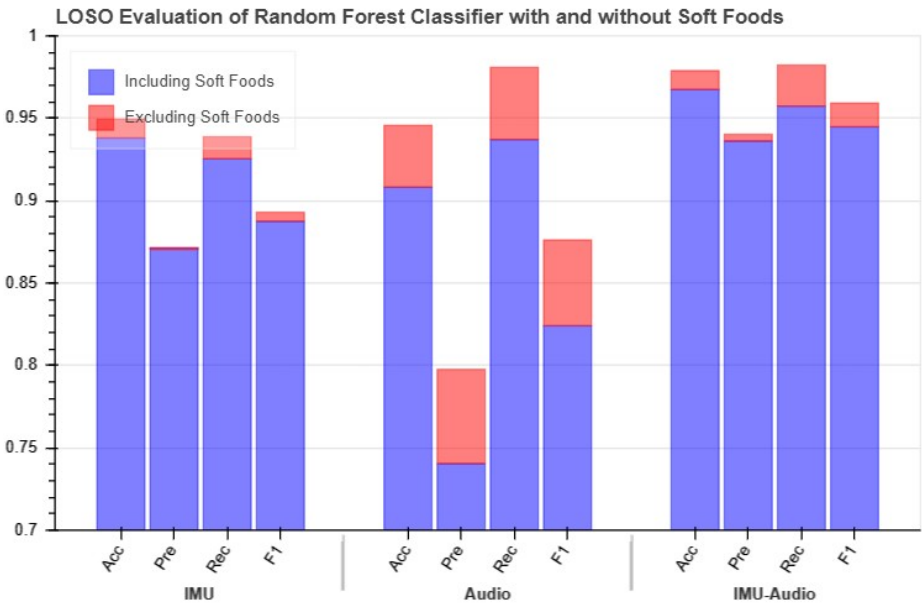

Figure 6.10: LOSO evaluation of the Random Forest classifier after removing soft foods. As can be seen, removing soft foods improved the performance of audio data but did not significantly improve the performance of IMU and fusing IMU and audio

7 FINAL WORD

7.1 DISCUSSION

The first and foremost result of this study is that the movement of the facial muscles while chewing is projected in the ear canal. The built-in IMU of the eSense earable is able to capture these movements and our proposed machine learning approach can distinguish the chewing IMU signals from other head-related signals collected from the eSense.

In addition to achieving a good accuracy for chewing detection, the eSense earable is also easy to wear and does not interfere with eating or daily tasks given its lightweight and small size. Participants did not mention any discomfort while wearing the device and some reported that they even forgot they were wearing the device throughout the study. Based on these feedbacks, we conclude that earables are a good platform for automatic dietary monitoring.

Overall, the in-ear IMU data performed better for chewing detection compared to audio data. The Random Forest classifier was more capable of distinguishing eating from non-eating activities based on both audio and IMU data. In addition to its higher accuracy, IMU data is advantageous over audio data given that it is non-intrusive and preserves privacy as it records neither the subjects' conversations nor other individuals interacting with them. IMU data consumes less power as a

result of lower sampling rates and fewer processing requirements. However, IMU data is noisy and sensitive to movement. For example, in case a participant is both walking and eating, more complicated signal processing techniques are required to filter out movement signals from chewing signals.

7.2 LIMITATIONS

The proposed method for chewing detection has its own limitations like any other work.

1. As the first limitation, only Spectral Centroid and MFCC were used as the frequency-domain features to train ML classifiers and other signal processing feature extraction techniques have not been studied yet.

2. The second limitation of our work is the limited activities performed by the participants, such that the data are collected while the users are performing just one sedentary task at a time. However, in real-life scenarios, people engage in more complicated activities during their mealtime, with watching TV being one of the most common examples.

7.3 FUTURE WORKS

Various approaches can be taken to improve the performance and reliability of the chewing detection method proposed in this research. To deal with the first limitation, we plan to implement more signal processing feature extraction techniques and evaluate the performance based on the new features by adding feature selection components to the proposed pipelines. To address the second

limitation, a more comprehensive dataset can be collected to involve a wider range of human activities to better resemble real-life human behavior.

8 CONCLUSION

In this research, we investigated the feasibility of using earables as a platform for the detection of chewing activities. We evaluated the performance of the built-in IMU and Microphone sensors of the eSense device separately and simultaneously based on three classifiers, namely, Logistic Regression, Decision Tree, and Random Forest. IMU data with an accuracy of 0.94, precision of 0.87, recall of 0.92, and F1 of 0.89 outperformed audio data in detecting chewing. We demonstrated the reason behind the better performance of IMU data is that contrary to audio data, IMU data is robust against soft foods, and removing soft food does not significantly improve the accuracy of chewing detection. Combining both modalities resulted in the highest performance with an accuracy of 0.97, precision of 0.94, recall of 0.96, and F1 of 0.94. Authors in [42] and [43] could achieve the same accuracy, however, they leveraged a device that is attached to the face which might be less socially acceptable than an earable. We identified a number of limitations in our study and proposed future work to fill in the gaps. The ultimate goal of this project is to assist individuals in mindful eating by detecting their eating activities in real-time and providing useful feedback to prevent diet-related diseases.

BIBLIOGRAPHY

[1] Nasir Ahmed, T_ Natarajan, and Kamisetty R Rao. "Discrete cosine transform." In: *IEEE transactions on Computers* 100.1 (1974), pp. 90–93.

[2] Oliver Amft. "A wearable earpad sensor for chewing monitoring." In: *SENSORS, 2010 IEEE*. IEEE. 2010, pp. 222–227.

[3] Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. "Analysis of chewing sounds for dietary monitoring." In: *International Conference on Ubiquitous Computing*. Springer. 2005, pp. 56–72.

[4] Samaneh Aminikhanghahi and Diane J Cook. "A survey of methods for time series change point detection." In: *Knowledge and information systems* 51.2 (2017), pp. 339–367.

[5] NTi Audio. *Fast Fourier Transformation FFT - Basics*. 2019. URL: https://www.nti-audio.com/en/support/know-how/fast-fourier-transform-fft.

[6] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. "EarBit: using wearable sensors to detect eating episodes in unconstrained environments." In: *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1.3 (2017), pp. 1–20.

[7] Shengjie Bi, Tao Wang, Ellen Davenport, Ronald Peterson, Ryan Halter, Jacob Sorber, and David Kotz. "Toward a wearable sensor for eating detection." In:

Proceedings of the 2017 Workshop on Wearable Systems and Applications. 2017, pp. 17–22.

[8] Shengjie Bi, Tao Wang, Nicole Tobias, Josephine Nordrum, Shang Wang, George Halvorsen, Sougata Sen, Ronald Peterson, Kofi Odame, Kelly Caine, et al. "Auracle: Detecting eating episodes with an ear-mounted sensor." In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.3 (2018), pp. 1–27.

[9] Yin Bi, Mingsong Lv, Chen Song, Wenyao Xu, Nan Guan, and Wang Yi. "AutoDietary: A wearable acoustic sensor system for food intake recognition in daily life." In: *IEEE Sensors Journal* 16.3 (2015), pp. 806–816.

[10] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[11] Rajesh Brid. *Logistic Regression Explained*. 2018. URL: https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb.

[12] Statistics Canada. *Canadian Community Health Survey*. 2019. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/190625/dq190625b-eng.htm.

[13] Scikit-Learn Developers. "Scikit-Learn User Guide." In: *Release 0.19* 2 (2018), pp. 214–215.

[14] Muhammad Farooq and Edward Sazonov. "Detection of chewing from piezoelectric film sensor signals using ensemble classifiers." In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2016, pp. 4929–4932.

[15] Muhammad Farooq and Edward Sazonov. "Accelerometer-based detection of food intake in free-living individuals." In: *IEEE sensors journal* 18.9 (2018), pp. 3752–3758.

[16] Brian J Fogg. "A behavior model for persuasive design." In: *Proceedings of the 4th international Conference on Persuasive Technology*. 2009, pp. 1–7.

[17] Juan M Fontana, Muhammad Farooq, and Edward Sazonov. "Automatic ingestion monitor: A novel wearable device for monitoring of ingestive behavior." In: *IEEE Transactions on Biomedical Engineering* 61.6 (2014), pp. 1772–1779.

[18] Juan M Fontana, Paulo Lopez-Meyer, and Edward S Sazonov. "Design of a instrumentation module for monitoring ingestive behavior in laboratory studies." In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2011, pp. 1884–1887.

[19] Yang Gao, Ning Zhang, Honghao Wang, Xiang Ding, Xu Ye, Guanling Chen, and Yu Cao. "iHear food: eating detection using commodity bluetooth headsets." In: *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE. 2016, pp. 163–172.

[20] Tzanetakis George, Essl Georg, and Cook Perry. "Automatic musical genre classification of audio signals." In: *Proceedings of the 2nd international symposium on music information retrieval, Indiana*. 2001.

[21] Shin-ichiro Iwatani, Hidemi Akimoto, and Naoki Sakurai. "Acoustic vibration method for food texture evaluation using an accelerometer sensor." In: *Journal of food engineering* 115.1 (2013), pp. 26–32.

[22] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Vol. 112. Springer, 2013.

[23] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Allesandro Montanari. "Earables for personal-scale behavior analytics." In: *IEEE Pervasive Computing* 17.3 (2018), pp. 83–89.

[24] Jindong Liu, Edward Johns, Louis Atallah, Claire Pettitt, Benny Lo, Gary Frost, and Guang-Zhong Yang. "An intelligent food-intake monitoring system using wearable sensors." In: *2012 ninth international conference on wearable and implantable body sensor networks*. IEEE. 2012, pp. 154–160.

[25] Beth Logan et al. "Mel frequency cepstral coefficients for music modeling." In: *Ismir*. Vol. 270. 2000, pp. 1–11.

[26] Jennifer Mathieu. "What should you know about mindful and intuitive eating?" In: *Journal of the Academy of Nutrition and Dietetics* 109.12 (2009), p. 1985.

[27] Christopher Merck, Christina Maher, Mark Mirtchouk, Min Zheng, Yuxiao Huang, and Samantha Kleinberg. "Multimodality sensing for eating recognition." In: *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ICST (IInstitute for Computer Sciences, Social Informatics and Telecommunications Engineering. 2016, pp. 130–137.

[28] Gert Mertes, Hans Hallez, Tom Croonenborghs, and Bart Vanrumste. "Detection of chewing motion using a glasses mounted accelerometer towards monitoring of food intake events in the elderly." In: *International Conference on Biomedical and Health Informatics*. Springer. 2015, pp. 73–77.

[29] Jessica T Monroe. "Mindful eating: principles and practice." In: *American Journal of Lifestyle Medicine* 9.3 (2015), pp. 217–220.

[30] World Health Organization. *Obesity and overweight*. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight..

[31] Vasileios Papapanagiotou, Christos Diou, and Anastasios Delopoulos. "Chewing detection from an in-ear microphone using convolutional neural networks." In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2017, pp. 1258–1261.

[32] Vasileios Papapanagiotou, Christos Diou, Lingchuan Zhou, Janet van den Boer, Monica Mars, and Anastasios Delopoulos. "A novel chewing detection system based on ppg, audio, and accelerometry." In: *IEEE journal of biomedical and health informatics* 21.3 (2016), pp. 607–618.

[33] Vasileios Papapanagiotou, Christos Diou, Lingchuan Zhou, Janet van den Boer, Monica Mars, and Anastasios Delopoulos. "The SPLENDID chewing detection challenge." In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2017, pp. 817–820.

[34] Sebastian Päßler and Wolf-Joachim Fischer. "Acoustical method for objective food intake monitoring using a wearable sensor system." In: *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. IEEE. 2011, pp. 266–269.

[35] Edward S Sazonov and Juan M Fontana. "A sensor system for automatic detection of food intake through non-invasive monitoring of chewing." In: *IEEE sensors journal* 12.5 (2011), pp. 1340–1348.

[36] Edward Sazonov, Stephanie Schuckers, Paulo Lopez-Meyer, Oleksandr Makeyev, Nadezhda Sazonova, Edward L Melanson, and Michael Neuman. "Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior." In: *Physiological measurement* 29.5 (2008), p. 525.

[37] Julius Orion Smith. *Mathematics of the discrete Fourier transform (DFT): with audio applications*. Julius Smith, 2007.

[38] Konrad C Steimer, Christoph Zimmermann, Jennifer Zeilfelder, Christian Pylatiuk, and Wilhelm Stork. "Portable auricular device for real-time swallow and chew detection." In: *Current Directions in Biomedical Engineering* 2.1 (2016), pp. 129–133.

[39] Stanley Smith Stevens, John Volkmann, and Edwin B Newman. "A scale for the measurement of the psychological magnitude pitch." In: *The Journal of the Acoustical Society of America* 8.3 (1937), pp. 185–190.

[40] Nanette Stroebele and John M de Castro. "Television viewing is associated with an increase in meal frequency in humans." In: *Appetite* 42.1 (2004), pp. 111–113.

[41] MA Tuğtekin Turan and Engin Erzin. "Detection of food intake events from throat microphone recordings using convolutional neural networks." In: *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE. 2018, pp. 1–6.

[42] Shuangquan Wang, Gang Zhou, Lisha Hu, Zhenyu Chen, and Yiqiang Chen. "CARE: Chewing activity recognition using noninvasive single axis accelerometer." In: *Adjunct Proceedings of the 2015 ACM International Joint*

Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers. 2015, pp. 109–112.

[43] Shuangquan Wang, Gang Zhou, Yongsen Ma, Lisha Hu, Zhenyu Chen, Yiqiang Chen, Hongyang Zhao, and Woosub Jung. "Eating detection and chews counting through sensing mastication muscle contraction." In: *Smart Health* 9 (2018), pp. 179–191.

[44] Wei Wang, Xingxing Wu, Guanchen Chen, and Zeqiang Chen. "Holo3DGIS: Leveraging Microsoft HoloLens in 3D Geographic Information." In: *ISPRS International Journal of Geo-Information* 7.2 (2018), p. 60. ISSN: 2220-9964. DOI: 10.3390/ijgi7020060. URL: http://www.mdpi.com/2220-9964/7/2/60.

[45] Xiaojun Wang, Leroy White, Xu Chen, Yiwen Gao, He Li, and Yan Luo. "An empirical study of wearable technology acceptance in healthcare." In: *Industrial Management & Data Systems* (2015), pp. 1704–1723.

[46] Xin Yang, Abul Doulah, Muhammad Farooq, Jason Parton, Megan A McCrory, Janine A Higgins, and Edward Sazonov. "Statistical models for meal-level estimation of mass and energy intake using features derived from video observation and a chewing sensor." In: *Scientific reports* 9.1 (2019), pp. 1–10.

[47] MA Yusnita, MP Paulraj, Sazali Yaacob, R Yusuf, and AB Shahriman. "Analysis of accent-sensitive words in multi-resolution mel-frequency cepstral coefficients for classification of accents in Malaysian English." In: *International Journal of Automotive and Mechanical Engineering* 7 (2013), p. 1053.

[48] Victor Zhou. *Random Forests for Complete Beginners*. 2019. URL: https://victorzhou.com/blog/intro-to-random-forests.

[49] Jaime Zornoza. *Logistic Regression Explained*. 2020. URL: https://towardsdatascience.com/logistic-regression-explained-9ee73cede081.