35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Estimation of Feature Importance for Food Intake Detection Based on Random Forests Classification

Juan M. Fontana, *Member, IEEE*, Muhammad Farooq, and Edward Sazonov*, *Senior Member, IEEE*

Abstract— **Selection of the most representative features is important for any pattern recognition system. This paper investigates the importance of time domain (TD) and frequency domain (FD) features used for automatic food intake detection in a wearable sensor system by using Random Forests classification. Features were extracted from signals collected using 3 different sensor modalities integrated into the Automatic Ingestion Monitor (AIM): a jaw motion sensor, a hand gesture sensor and an accelerometer. Data was collected from 12 subjects wearing AIM in free-living for a 24-hr period where they experienced unrestricted intake. Features from the sensor signals were used to train the Random Forests classifier that estimated the importance of each feature as part of the training process. Results indicated that FD features from the jaw motion signal and TD features from the accelerometer signal were the most relevant features for food intake detection.**

I. Introduction

The study of ingestive behavior of individuals is particularly important to detect and correct patterns of food intake associated with obesity and eating disorders. Obesity is the excessive body fat accumulation caused by a chronic imbalance between energy intake and energy expenditure. The prevalence of obesity among adults was reported to be 35.5% in United States in 2009-2010 [1]. Eating disorders (ED) are serious mental disorders that cause disturbances on eating habits or weight-control behavior of individuals [2]. Anorexia nervosa, bulimia nervosa and binge eating disorder are the most common ED with lifetime prevalence ranging from 0.6 to 4.5% in the United States [3]. Both obesity and ED are medical conditions highly resistant to treatment and that can have severe physical and physiological health consequences [4]. Thus, the implementation of objective and accurate methods for Monitoring Ingestive Behavior (MIB) is critical to provide an adequate assessment of intake particularly in individuals who would most benefit from professional assistance.

Current methods for MIB rely on subjects self-reporting their daily intake (i.e. what, when and how much they ate). Although these methods may be provide reliable information when performing laboratory experiments, their reliability vanishes when subjects are asked to report their intake in free-living situations. Consequently, there is a need for objective, innovative strategies to accurately assess free-living food intake patterns in humans.

Wearable sensor systems have been implemented for automated food intake detection. These systems monitor physiological changes related to food intake using noninvasive sensors that, together with signal processing and pattern recognition algorithms, are used to determine when food is consumed. In [5] the sounds generated during chewing and/or swallowing of food were captured by microphones placed in the ear canal. The acoustic signals were used to develop computer algorithms that achieved food intake detection accuracies between 83% and 86%. In [6] and [7], swallowing sounds were captured by a miniature microphone placed over the throat. The acoustic information was used to create group and individual models to detect periods of food intake [8]. Higher detection rates were observed for individual models suggesting the need for calibration. Food intake detection through monitoring of chewing using a piezoelectric strain gauge sensor was introduced in [9]. An SVM classifier was able to achieve 81% accuracy for single meal experiments. Most of the pattern recognition approaches described in this paragraph achieved acceptable detection accuracies in laboratory experiments, however their performance will most probably be affected in free-living conditions due to the influence of real life situations that are not possible to take into account in a laboratory setup.

The selection of the most representative features from sensor signals is critical to obtain pattern recognition systems that accurately predict food intake in free living. In this study, we used Random Forests ensemble classification technique to estimate the importance of time domain (TD) and frequency domain (FD) features for describing food intake. Features were extracted from signals collected using a wearable sensor system (Automatic Ingestion Monitor, AIM) that monitored jaw motion, hand to mouth gestures and body acceleration. Subjects wore AIM in free-living for a 24-hr period while performing ad libitum intake. Results indicated that the most important features were in the FD for the jaw motion sensor signal. Accelerometer features also presented high level of importance whereas the hand gesture features were not as important as the other sensor features.

II. Methods

A. Data Collection

A total of 12 subjects (6 male, 6 female) participated in this study. The average age was 26.7 y (SD ± 3.7) and the average body mass index (BMI) was 24.39 kg/m^2 (SD ±

The project described was supported by Grant Number R21DK085462 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDDK or the NIH.

Juan M. Fontana (jmfontana@bama.ua.edu), Muhammad Farooq (mfarooq@crimson.ua.edu) and Edward Sazonov (esazonov@eng.ua.edu) are with the Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL 35487 USA.

*Corresponding author (phone: 205-348-1981)

3.81). Subjects did not present any medical condition that would affect normal food intake. The study was approved by the Internal Review Board at The University of Alabama and subjects read and signed an informed consent document.

Each subject was asked to wear the sensor system AIM in free-living for a 24-hr period where they accomplished ad libitum intake and were able to perform daily living activities without restrictions. AIM consists of a wireless module that integrates 3 different sensor modalities:

1- A *jaw motion sensor* to detect characteristic motion of the jaw during chewing [9], [10]. This sensor was attached directly below the ear using medical adhesive.

2- A *hand gesture sensor* to detect hand-to-mouth (HtM) gestures associated to bites. It consists of a RF transmitter worn on the inner side of the dominant arm and a RF receiver in the wireless module operating in RFID frequency band of 125Khz.

3- A *tri-axial accelerometer* located in the wireless module to detect body acceleration.

The jaw motion signal was acquired at a 1 kHz sampling frequency whereas the hand gesture and accelerometer signals were acquired at a 10 Hz and 100 Hz respectively. All sensor signals were quantized with 12-bit resolution and delivered in near real time via Bluetooth to an Android smart phone that acted as a data logger. An example of the signals collected after 24-hr is showed in Figure 1.

A push button was included in AIM as the primary method for self-reporting food intake. Subjects pressed and held the button during chewing to mark food intake. The push button signal was used as the gold standard for training the Random Forests classifier.

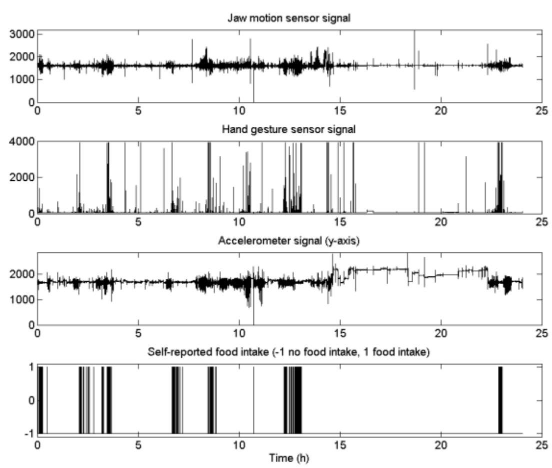

Figure 1. Example of the signals collected after a subject wore the wearable sensor system for 24-hr.

B. Signal Preprocessing and Feature Extraction

The jaw motion signal, $JM(t)$, and the accelerometer signals, $ACC_X(t)$, $ACC_Y(t)$, and $ACC_Z(t)$, were high-pass filtered (0.1 Hz cutoff frequency) to remove the DC component. $JM(t)$, $ACC_X(t)$, $ACC_Y(t)$, and $ACC_Z(t)$ were then normalized to compensate for variations in signal amplitude between subjects. The hand gesture signal, $HG(t)$, was normalized with respect to its maximum value. HtM gestures shorter than 0.25s and longer than 7.5s were removed from $HG(t)$ as it was assumed that they did not belong to food intake activity.

$JM(t)$, $HG(t)$, $ACC_X(t)$, $ACC_Y(t)$, $ACC_Z(t)$, and push button (self-report) signals were divided into non-overlapping 30s epochs that presented a fine time resolution that is suitable for monitoring brief ingestion events such as snacking. A set of TD and FD features were extracted from the signals (Table 1,Table 2, andTable 3) and then combined to create a feature vector $f_i \in \Re^N$ that represented each 30s epoch.

To compute the FD features of $JM(t)$, the frequency spectrum of the signal within an epoch was divided into different frequency ranges. These ranges may potentially contain important information related to different activities that could help to discriminate between intake and no intake episodes (i.e. 1.25-2.5 Hz for chewing, 2.5-10 Hz for walking, and 100-300 Hz for talking [10])

Each feature vector f_i was associated with a class label c_i = {'no food intake'; 'food intake'} for the classification task. A class label belonged to 'food intake' ($c_i = 1$) if at least 10s of the self-report signal within the i-*th* epoch reported food intake; otherwise, it belonged to 'no food intake' ($c_i = -1$).

TABLE 1. TIME AND FREQUENCY DOMAIN FEATURES EXTRACTED FROM EACH EPOCH OF THE JAW MOTION SIGNAL

#	*Description*	#	*Description*
1	Mean Absolute Value (MAV)	20	Energy spectrum in chewing range (chew_ene)
2	Root Mean Squared (RMS)	21	Entropy of spectrum chewing range (chew_ene)
3	Maximum value (Max)	22	chew_ene / spectr_ene
4	Median value (Med)	23	Energy spectrum in walking range (walk_ene)
5	MAV / RMS	24	Entropy of spectrum walking range (chew_ene)
6	Max / RMS	25	walk_ene / spectr_ene
7	MAV / Max	26	Energy spectrum in talking range (chew_ene)
8	Med / RMS	27	Entropy of spectrum talking range (chew_ene)
9	Signal entropy (Entr)	28	talk_ene / spectr_ene
10	Num. of zero crossings (ZC)	29	chew_ene / walk_ene
11	Mean time between ZC	30	chew_entr / walk_entr
12	Num. of peaks (NP)	31	chew_ene / talk_ene
13	Average range	32	chew_entr / talk_entr
14	Mean time between peaks	33	walk_ene / talk_ene
15	NP/ZC	34	walk_entr / talk_entr
16	ZC/NP	35	Fractal dimension (fractal_d)
17	Wavelength	36	Peak frequency in chewing range (maxf_chew)
18	Num. slope sign changes	37	Peak frequency in walking range (maxf_walk)
19	Energy of frequency spectrum (spectr_ene)	38	Peak frequency in talking range (maxf_talk)

TABLE 2. TIME DOMAIN FEATURES EXTRACTED FROM EACH EPOCH OF THE HAND GESTURE SIGNAL

#	*Description*	#	*Description*
1	Num. of HtM gestures within epoch (num_HtM)	6	Wavelength
2	Duration of HtM	7	Wavelength / Duration HtM
3	MAV of HtM	8	Duration HtM / num_HtM
4	Stardard Deviation (SD) of HtM	9	MAV_HtM / Duration HtM
5	Maximum value (Max_HtM)		

TABLE 3. TIME DOMAIN FEATURES EXTRACTED FROM EACH EPOCH OF THE ACCELEROMETER SIGNALS

#	*Description*	#	*Description*
1	MAV of ACC_X (MAV_x)	12	Entropy of ACC_Y
2	SD of ACC_X (SD_x)	13	MAV of ACC_Z (MAV_z)
3	Median of ACC_X (Median_x)	14	SD of ACC_Z (SD_z)
4	Num. of zero crossings for ACC_X (ZC_x)	15	Median of ACC_Z (Median_z)
5	Mean time between ZC for ACC_X	16	Num. of ZC for ACC_Z (ZC_z)
6	Entropy of ACC_X	17	Mean time between ZC for ACC_Z
7	MAV of ACC_Y (MAV_y)	18	Entropy of ACC_Z
8	SD of ACC_Y (SD_y)	19	MAV of ACC_Z , ACC_Z and ACC_Z combined (MAV_xyz)
9	Median of ACC_Y (Median_y)	20	SD of ACC_Z , ACC_Z and ACC_Z combined (SD_xyz)
10	Num. of zero crossings for ACC_Y (ZC_y)	21	Entropy of ACC_Z , ACC_Z and ACC_Z combined (Entr_xyz)
11	Mean time between ZC for ACC_Y		

C. *Random Forests*

Random forests is a ensemble classification technique developed by Breiman [11]. A collection of m_{tree} decision trees are created and the final decision for a test point is obtained by aggregating the results for each tree using majority vote. Trees are constructed using different bootstrap samples of the original dataset. At each node of a tree, a set of m_{try} features are randomly selected from the F available features and the best split is chosen among those m_{try} features. Random Forests performs very well compared to other classifiers [12] with the main advantage of being robust against overfitting.

Two parameters are necessary to set for training the Random Forests classifier: the number of trees in the forest (m_{tree}) and the number of features in the random subset at each node (m_{try}). A grid search procedure was implemented using a reduced dataset to find the optimal set of parameters by varying m_{tree} *={100, 200, 300, ... , 1500}* and m_{try} = *{s/2, s, 2·s, 4·s, 8·s,}* where $s = \sqrt{F} = 8$ as the suggested in [11].

An important characteristic of Random Forests is that it can estimate the relative importance of a feature. This is done by calculating the increase in classification error of the samples that were not included in the bootstrap sample when the values of a specific feature are permuted while all other feature values remain unchanged. Consequently, features with higher increase in error are considered more important than features with lower increase in error. For that reason, Random Forests was used in this study to determine which features contain relevant information to discriminate between food intake and no food intake in free living.

A leave one out cross-validation procedure was used to train a Random Forests classifier. This allowed creating a classifier with data from 11 subjects and validating with data from the subject that was left out, which was completely independent of the training set. A total of 10 runs of the cross-validation procedure were performed using the optimal set of m_{tree} and m_{try} values to obtain more general results.

III. RESULTS

Results of the grid search procedure indicated that m_{tree} = 400 and m_{try} = 32 was the optimal combination of parameters (73.2% classification accuracy). Results of the feature importance estimations for each signal are showed in Figures 2, 3, and 4. The most important feature for $JM(t)$ was the ratio between the entropy of the spectrum in the walking range and the entropy of the spectrum in the talking range. The most important feature for $HG(t)$ was the duration of HtM gestures within the 30s epoch. Finally, the most important feature for the accelerometer signals was the standard deviation of the signal in the y-axis.

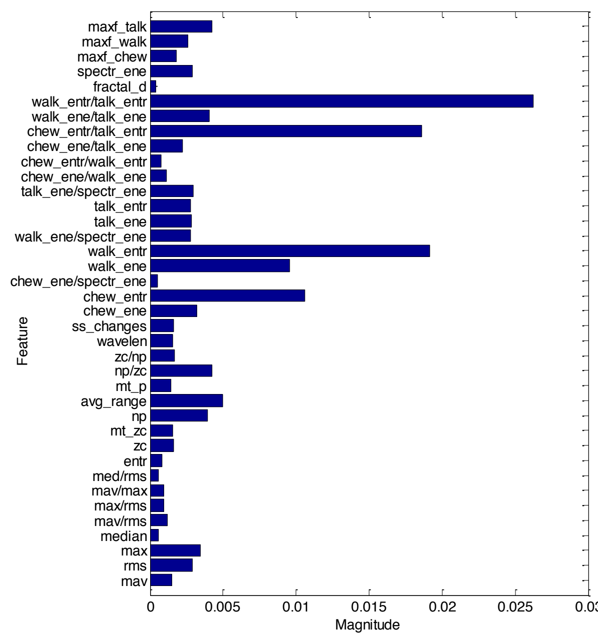

Figure 2. Results of feature importance for the jaw motion sensor signal

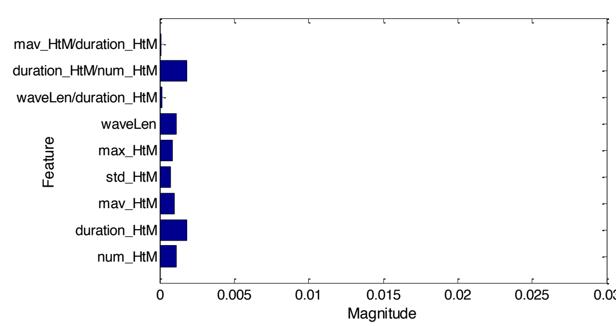

Figure 3. Results of feature importance for the hand gesture sensor signal

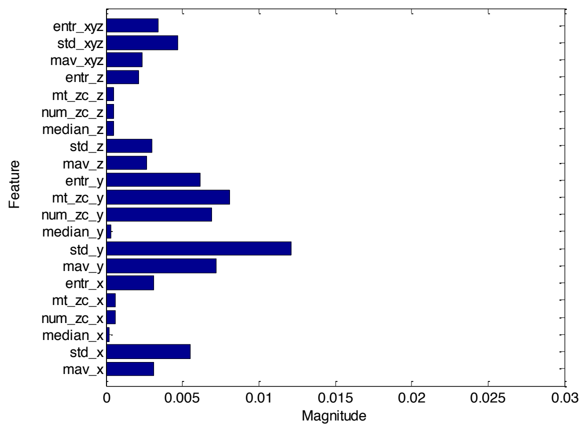

Figure 4. Results of feature importance for the accelerometer signals

IV. DISCUSSION

An accurate and objective detection of food intake in free living is highly desirable to obtain reliable information about dietary intake of individuals. An objective detection can be obtained by using pattern recognition algorithms whereas an accurate detection is more challenging and could be obtained by extracting relevant information from the sensor signals. In this work, a Random Forests classification algorithm was used to estimate the importance of features extracted from three different sensor modalities: a jaw motion sensor to monitor chewing, a hand gesture sensor to monitor bites and an accelerometer to monitor body motion. As expected, the results showed that the jaw motion features presented the most valuable information for food intake detection in free living. Accelerometer signals also presented features with relative high importance compared to jaw motion features. Hand gesture features presented the lowest importance.

Results for the jaw motion sensor signals showed that the most important features were in the FD. This is can be observed in Figure 2, were five FD features presented considerably higher level of importance than the remaining features. They were related to both energy and entropy of the spectrum at the chewing (1.25-2.5 Hz), walking (2.5-10 Hz) and talking (100-300 Hz) frequency ranges. Results also indicated that both chewing and walking ranges appear to be critical to detect intake. The high importance of the walking range was not expected as subjects did not spent much time walking. However, due to the chewing and walking ranges being very close in spectral content, it is possible that the walking range may included information about chewing.

Results of the feature importance estimation for the accelerometer signals showed that the most relevant features were extracted from the y-axis, which represented the acceleration on the frontal plane of the body. The SD, the mean time between ZC and the MAV of the signal within an epoch were the main features selected. These features may help to discriminate between periods of body motion and quietness. Therefore, since many individuals consume their meals mostly in a stationary position, it is reasonable to think that accelerometer features would help to detect food intake.

For the hand gesture sensor signal, the number of HtM gestures and their duration within the epoch were the most important features as estimated by Random Forests. However, these features showed a low importance level compared to the jaw motion and accelerometer features. The hand gesture sensor detected gestures that were not related to food intake (i.e. subjects operating a cell phone while eating). For this reason, hand gesture features by themselves may not provide an accurate food intake prediction.

The results presented in this study may be used in different ways. One of them is to use the most important features to perform a hierarchical classification in order to identify and remove data related to 'no food intake' (i.e. sleeping, working in a computer, watching TV, etc.). In fact, the data collected after 24-hr contained only about 3% of food intake data, thus a reduction in the amount of 'no food intake' data would help to obtain a more balanced dataset for classification. Another alternative is to rank the features according to their importance and use the *N* most significant features in combination with a classifier capable of a better representation of the nonlinear decision boundary and thus achieve a higher food intake detection accuracy.

REFERENCES

[1] K. M. Flegal, M. Carroll, B. Kit, and C. L. Ogden, “Prevalence of Obesity and Trends in the Distribution of Body Mass Index Among US Adults, 1999-2010,” *JAMA*, vol 307, no 5, pp. 491–497, 2012.

[2] C. G. Fairburn and P. J. Harrison, “Eating disorders,” *The Lancet*, vol. 361, no. 9355, pp. 407–416, Feb. 2003.

[3] J. I. Hudson, E. Hiripi, H. G. Pope Jr, and R. C. Kessler, “The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication,” *Biol. Psychiatry*, vol. 61, no. 3, pp. 348–358, Feb. 2007.

[4] D. Sánchez-Carracedo, D. Neumark-Sztainer, and G. López-Guimerà, “Integrated prevention of obesity and eating disorders: barriers, developments and opportunities,” *Public Health Nutrition*, vol. 15, no. 12, pp. 2295–2309, 2012.

[5] S. Päßler, M. Wolff, and W.-J. Fischer, “Food intake monitoring: an acoustical approach to automated food intake activity detection and classification of consumed food,” *Physiol Meas*, vol. 33, no. 6, pp. 1073–1093, 2012.

[6] P. Lopez-Meyer, O. Makeyev, S. Schuckers, E. Melanson, M. Neuman, and E. Sazonov, “Detection of Food Intake from Swallowing Sequences by Supervised and Unsupervised Methods,” *Ann Biomed Eng*, vol. 38, no. 8, pp. 2766–2774, 2010.

[7] E. Sazonov, O. Makeyev, P. Lopez-Meyer, S. Schuckers, E. Melanson, and M. Neuman, “Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior,” *IEEE Trans Biomed Eng*, vol. 57, no. 3, pp. 626–633, 2010.

[8] O. Makeyev, P. Lopez-Meyer, S. Schuckers, W. Besio, and E. Sazonov, “Automatic food intake detection based on swallowing sounds,” *Biomedical Signal Processing and Control*, vol. 7, no. 6, pp. 649–656, Nov. 2012.

[9] E. S. Sazonov and J. M. Fontana, “A Sensor System for Automatic Detection of Food Intake Through Non-Invasive Monitoring of Chewing,” *IEEE Sensors Journal*, vol. 12, no. 5, pp. 1340 –1348, May 2012.

[10] J. M. Fontana and E. S. Sazonov, “A robust classification scheme for detection of food intake through non-invasive monitoring of chewing,” in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 4891 –4894.

[11] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[12] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” in *ICML 2006 - Proceedings of the 23rd International Conference on Machine Learning*, 2006, vol. 2006, pp. 161–168.