Detection of chewing motion in the elderly using a glasses mounted accelerometer in a real-life environment

Gert Mertes[1,2], Hans Hallez[3], Bart Vanrumste[1,2] and Tom Croonenborghs[4,5]

***Abstract*— This paper describes a method of detecting an elderly person's chewing motion using a glasses mounted accelerometer. A real-life dataset was collected from 13 elderly adults, aged 65 or older, during meal times in a care facility. A supervised classifier is used to automatically distinguish between epochs of chewing and non-chewing activity. Results are compared to a lab dataset of 5 young to middle-aged adults captured in previous work. K-Nearest Neighbor, Random Forest and Support Vector Machine classifiers are evaluated. All are able to achieve similar performance, with the Support Vector Machine performing the best with an F1-score of 0.73**

I. INTRODUCTION

Nutrition plays an important role in the health status of an elderly adult [1]. In the frailest of people, malnourishment will severely impact both the psychological and physical health of the person, as well as decrease the overall quality of life. It can have a negative effect on the healing process after an illness or surgery, poor wound healing as well as reduce mobility due to decreased muscle mass and reduced cognitive function [2]. Studies have shown that up to $15\,\%$ of community-dwelling and home-bound elderly adults are malnourished and another $45\,\%$ are at risk. This risk of malnourishment increases as the health of the older person deteriorates. Up to $65\,\%$ of hospitalized subjects are malnourished. In care facilities, this number increases up to $85\,\%$ [3].

Early recognition of malnourishment and the monitoring of nutritional status could greatly increase quality of life in both home-bound and institutionalized elderly adults. Traditional monitoring methods, such as daily questionnaires or food diaries, however, are time consuming and typically contain mistakes [4]. A possible method to overcome this disadvantage is to automatically monitor a person's food intake using a wearable device.

In previous work, we already suggested the use of an accelerometer mounted to a pair of glasses to detect the wearer's chewing motion, specifically targeting the elderly population [5]. This method has the advantage of not requiring a drastic change in the routine of the person and, if the sensor is built into the frame, is less stigmatizing. This can aid in user acceptance and makes it more suitable for adaptation by elderly people. Our previous work, however, was limited to a small dataset consisting of 5 young to middle-aged adults recruited on campus. Data from the accelerometer was collected during a meal in a controlled lab environment to study the feasibility of the approach. While test subjects were allowed to talk to the researcher during the meal to incorporate some level of realism into the measurement, interaction with the environment was kept to a minimum. The goal of this work is to evaluate the system proposed in [5] on elderly adults in a real living environment and compare the results to those previously obtained in a lab environment.

This work is funded by DISH and Engineers@CareHome, supported by KU Leuven, Vlaams Brabant and OCMW Leuven.

Corresponding author: gert.mertes@kuleuven.be
[1]KU Leuven, Campus Groep T, e-Media Lab, Belgium
[2]KU Leuven, ESAT-STADIUS, Belgium
[3]KU Leuven, Campus Oostende, ReMI, Belgium
[4]KU Leuven, Campus Geel, AdvISe, Belgium
[5]BWH/HMS/Broad Institute, Boston, USA

II. DATA COLLECTION

Data was collected from 13 elderly adults aged 65 or older in a nursing home. Participants were asked to consume a meal while wearing the capture setup shown in Figure 1. The tri-axial accelerometer is a Shimmer3 IMU sampling at $128\,\text{Hz}$. For presentation purposes, the accelerometer unit is attached using a hook and loop fastener. During data collection, however, the unit was firmly attached using cable ties to ensure a good transfer of the movement and vibrations of the glasses to the accelerometer. The accelerometer is mounted around the temple area and was kept mounted in the same place for all test subjects. Test subjects were also video recorded to annotate the data afterwards. Video and accelerometer data was captured synchronously.

There were no restrictions placed upon the elderly adult during the measurement. While the measurement took place at a separate table that was different from their regular seating position for practical reasons, care was taken to emulate a real world environment as closely as possible. The majority of subjects were measured in pairs and both subjects were allowed to interact with each other, just like they would normally do. Interaction with the researcher was also allowed as well as any other meal related activity such as leaving the table and walking to the counter to get coffee or tea. Figure 1 shows one of the subjects during the data collection process. Subjects were also equipped with wrist mounted Shimmer3 IMU units on both wrists. They are not used in this study, but will be the subject of future work. The medical ethical commission of the KU Leuven university hospital approved this study and all participants gave their written informed consent.

Two main classes were annotated in the data: chewing and non-chewing. An epoch of chewing is the time between when a subject first starts chewing a new bite and when the subject swallows. The non-chewing label contains everything else during the meal, e.g. talking, cutting, drinking, etc. Annotation was done visually using the captured video

978-1-5090-2809-2/17/$31.00 ©2017 IEEE

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on May 11,2020 at 14:04:07 UTC from IEEE Xplore. Restrictions apply.



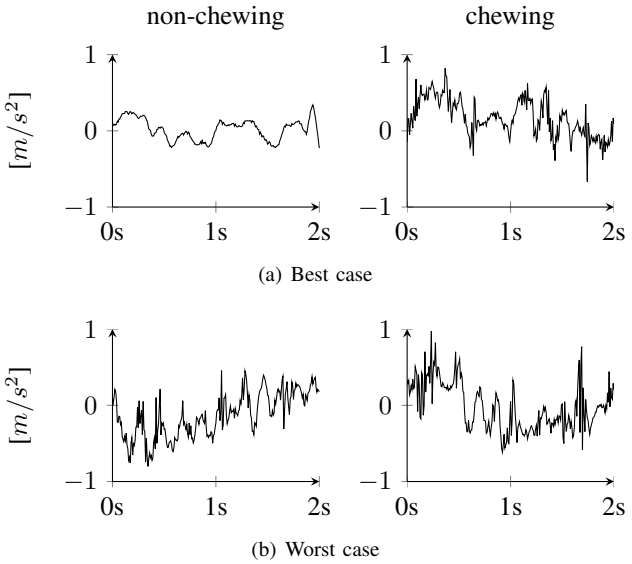
Fig. 1. Left: out-take of the data collection process. Right: setup used for data collection.

image. It is worth noticing that this resulted in an unbalanced dataset, the total time spent doing the non-chewing activity is longer than that of the chewing activity. As seen in Figure 1, subjects were given their regular daily meal, which often included a side dish such as soup. While this activity does contain eating activity, it was assigned the non-chewing label since this study focuses on chewing alone.

A total of 93.8 min of chewing and 315.5 min of non-chewing was recorded. The average recording time per participant was 31.5 ± 6.5 min. A total number of 153 chewing epochs were recorded with an average duration per epoch of 37 ± 57 s. Several subjects had the tendency to eat the entire meal without pauses in between bites, resulting in a high average time per chewing epoch and standard deviation.

We already described the occurrence of distinctive peaks and high-frequency content in the accelerometer signal while a person is chewing in a controlled environment [5]. This is still the case in our real-life data. However, the majority of the real-life data contains more noise and auxiliary movements due to the unconstrained manner of data collection. This is illustrated in Figure 2. Figure 2(a) shows a best case scenario where a visual distinction can still easily be made between both activities. Figure 2(b) shows a worst case scenario where a visual distinction can no longer be made between both activities due to noise and other activities present. The displayed epochs were performed in order, i.e. the shown chewing epoch directly follows the shown non-chewing epoch. From the tri-axial accelerometer data, only the axis aligned mediolateral to the subject is shown.

The real-life dataset contains a single meal measurement from 13 elderly adults and will be used to train and validate a supervised classifier. Two subjects, however, had to be excluded from the dataset. One subject suffered from Parkinson's disease with a severe tremor in the upper limbs. As such, the accelerometer signal was entirely predominated by the erratic movements caused by the tremor, making it impossible to distinguish between epochs of chewing and non-chewing using our proposed method. Another subject was given a personalized meal consisting entirely of soft foods due to a decreased chewing ability, resulting in the subject not chewing the food but directly swallowing it. These two subjects were excluded from the study, resulting in a dataset of 11 subjects. The methods evaluated on the real-life dataset will also be applied to the lab dataset obtained in [5] using the same capture setup. The lab dataset contains



(a) Best case

(b) Worst case

Fig. 2. Illustration of the captured real-life accelerometer data. Only the axis aligned mediolateral to the person is shown.

a single meal measurement from 5 young to middle aged adults in a controlled environment.

III. METHODOLOGY

A. Feature Extraction and Selection

The raw accelerometer signal is first filtered with a linear phase high-pass filter with a cut-off frequency of 0.5 Hz to discard the DC component of the signal. From the tri-axial accelerometer signal, only the signal mediolateral to the subject is used. This choice is based on our earlier findings in [6]. The data of each person is then split up into two signals, each signal containing only the chewing and non-chewing data respectively. These signals are then segmented into non-overlapping windows of 15 seconds. Features from both the frequency and time domain are extracted from each window.

From the time domain, we calculate: (a) number of zero-crossings and (b) 75th percentile value. From the frequency spectrum obtained with a Fast Fourier Transform, we calculate: (c) dominant frequency, (d) 25th percentile and (e) 75th percentile. The choice of features combines those used in [5] and those proposed in [6], which showed promising results on the lab and real-life dataset respectively. A maximum-relevance minimum-redundancy forward feature selection based on [7] was performed on the real-life dataset, which selected (a), (d) and (e) as optimal features.

B. Training and Validation

Three different types of classifiers are evaluated: K-Nearest Neighbours (KNN), Support Vector Machine (SVM) and Random Forest (RF). For the KNN classifier, the number of neighbours k was evaluated between 1 and 10. The highest performance was achieved with $k = 5$. For the SVM, a linear, Gaussian and 2nd order polynomial (quadratic) kernel are evaluated. Finally, the RF has been trained with 50 bags of

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on May 11,2020 at 14:04:07 UTC from IEEE Xplore. Restrictions apply.

TABLE I
VALIDATION RESULTS.

	Lab data			Real-life data		
Classifier	Precision [%]	Recall [%]	F1-score	Precision [%]	Recall [%]	F1-score
RF	86.6 ± 12.0	85.3 ± 13.0	0.86	74.6 ± 19.1	64.0 ± 27.3	0.69
KNN	84.6 ± 13.7	86.2 ± 18.8	0.85	75.1 ± 19.8	69.8 ± 24.5	0.72
SVM - linear	87.8 ± 15.8	85.7 ± 19.3	0.87	56.1 ± 27.3	35.1 ± 29.1	0.43
SVM - gaussian	85.7 ± 15.2	86.0 ± 18.0	0.86	81.2 ± 18.7	58.4 ± 26.4	0.68
SVM - quadratic	87.8 ± 15.0	84.3 ± 16.6	0.86	82.7 ± 15.3	65.9 ± 28.7	0.73

trees. Each classifier performs a binary classification. The chewing activity is considered the positive class, while non-chewing is negative.

Training and validation of each classifier type is done using the leave-one-person-out method. Data of one person is used to evaluate a model that is trained with the data of all other persons. This is done once for each subject in the dataset. Average precision and recall is calculated over all iterations. A leave-one-person-out validation is done for each dataset individually, data is not shared between the lab and real-life dataset.

IV. RESULTS

Table I shows the results of the leave-one-person-out validation for each classifier type and dataset. The table contains the average precision and average recall together with their respective standard deviations. From the precision and recall, F1-score is calculated as a measure to quantify the best performing classifier altogether. F1-score is defined as $2 \cdot \frac{precision \cdot recall}{precision + recall}$. The quadratic kernel SVM has the highest F1-score on the real-life dataset of 0.73 and achieves an overall detection accuracy, precision and recall of $89.2\,\%$, $82.7\,\%$ and $65.9\,\%$ respectively.

Figure 3 shows the effect of the window length on the average precision and recall of the SVM classifier validated using the real-life dataset. While a maximum precision presents itself at a window length of $19\,\mathrm{s}$, a window length of $15\,\mathrm{s}$ results in the highest combination of precision and recall with only a marginal decrease in precision.

Finally, Figure 4 shows the precision-recall curves for the three different classifier types validated using the real-life dataset.

V. DISCUSSION

The KNN and quadratic kernel SVM classifier have almost the same F1-scores and both perform well on the real-life dataset, each achieving the highest recall and precision respectively. When choosing the best classifier, a trade-off between precision and recall will have to be made. The highest F1 score of 0.73, which is the highest equally weighed combination of precision and recall, is achieved by the SVM with quadratic kernel. As seen in Figure 4, however, a different operating point can result in another classifier achieving a higher performance, depending on the chosen trade-off between precision and recall. Around recall values of $78\,\%$, for example, the Gaussian kernel outperforms

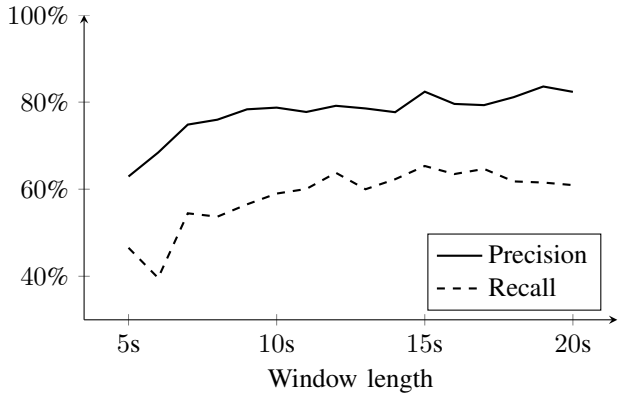


Fig. 3. Effect of the window length on precision and recall for the SVM classifier with quadratic kernel and real-life dataset.

the quadratic kernel for the real-life dataset, although the difference is marginal.

As seen in Table I, there is little to no variation between classifier types when using the lab dataset. When data is recorded in a controlled environment, epochs of chewing and non-chewing typically don't contain any auxiliary movements that can skew performance of the features and as such, different classifier types perform equally well. As expected, performance drops when transitioning to a real-life environment. Auxiliary movements performed while both chewing and non-chewing will have a negative impact on classifier performance. Since subjects were often recorded in pairs of two, talking is one of the auxiliary activities predominately present in the data.

A window length of $5\,\mathrm{s}$ was used in [5], while a maximum performance with a window length of $15\,\mathrm{s}$ is obtained using the approach presented in this work, as shown in Figure 3. This can be attributed to the fact that elderly adults typically take more time to chew than younger adults.

User acceptance of the prototype capture system was high. None of the participants had a problem wearing a sensor on their glasses, nor did anyone experience discomfort due to the light nature of the sensor. Furthermore, a focus group consisting of 6 elderly adults was organized to evaluate user perception and expectations of the proposed capture system. The majority of subjects in both the focus group and participant group would prefer a more integrated and invisible system. Possible form factors include a clip-on system or a type of sleeve that slides over the frame, which

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on May 11,2020 at 14:04:07 UTC from IEEE Xplore. Restrictions apply.

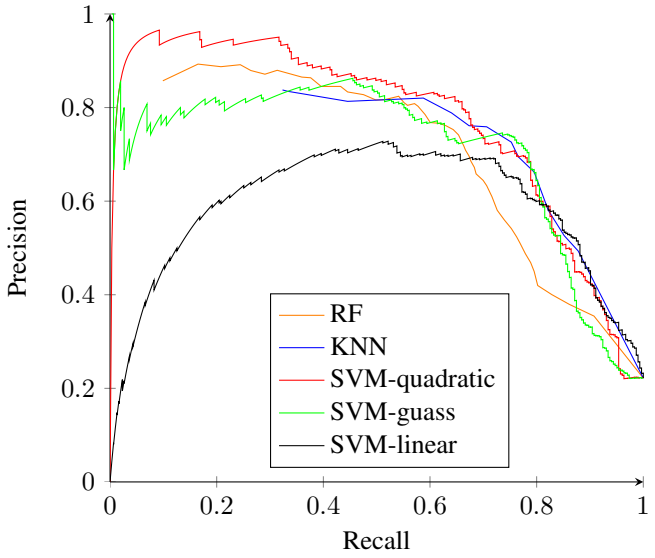


Fig. 4. Precision-recall curves for each of the evaluated classifiers validated using the real-life dataset.

would be virtually invisible and light enough not to be noticeable by the wearer. While not all people wear their glasses continuously, our system does provide an alternative to other, more stigmatizing, capture methods.

The approach presented in this work outperforms that of our previous work [5], [6]. For the lab dataset, we previously obtained an average precision and recall of 75.7 % and 80.0 % respectively. For the real-life dataset, we previously obtained an average precision and recall of 61.4 % and 56.0 % respectively. This work improves on those performance scores by combining features from both works.

VI. Related Work

Fontana and Sazonov [8] demonstrated a strain sensor taped to the jaw to detect epochs of chewing activity with an average accuracy of 90.52 %. In [9], the jaw motion sensor is further supplemented with an accelerometer worn on the wrist and one worn on a lanyard around the neck to detect overall periods of food intake based on chewing motion and body movements. This system achieves an average detection accuracy of 89.8 % and a precision and recall of 89.1 % and 90.4 % respectively. While both systems outperform ours in terms of overall detection accuracy, our system utilizes only a single accelerometer that can easily be integrated into the elderly's routine and doesn't require any sensors taped to the body. Passler et al. [10] show an audio based system using a microphone integrated in a hearing aid capable of detecting chewing activity with a precision and recall of 91.3 % and 81.8 % respectively. Their work, however, is difficult to compare to ours, as they also include data from young adults and data recording took place in a controlled environment where participants were not allowed to talk during meals. Lopez-Meyer et al. [11] follow a similar approach using a throat microphone. They do not report precision and recall scores, but are able to achieve a detection accuracy of 94 %. Their study, however, did not include elderly adults and rely on an obtrusive throat microphone. The aforementioned works all target healthy young to middle-aged adults or rely on stigmatizing or uncomfortable wearables. Our system is more suited for use by elderly adults. There is, however, still room for improvement. First and foremost, our feature set is relatively small. Performance may be improved by selecting more or better features. Furthermore, the influence of body movements on the detection accuracy may be reduced by supplementing the system with a wrist mounted IMU.

VII. Conclusion

This paper presents a method to detect chewing motion using a glasses mounted accelerometer. By combining features from previous work, we are able to improve detection accuracy and achieve an average accuracy, precision and recall of 89.2 %, 82.7 % and 65.9 % respectively. The proposed system is comfortable to wear, requires little adaptation to the user's routine and is non-stigmatizing, making it more suitable for use by elderly adults.

Acknowledgment

The authors would like to thank the staff and residents of care facility *WZC Edouard Remy* in Leuven for their support and participation in the data collection process.

References

[1] L. Donini, P. Scardella, L. Piombo, B. Neri, R. Asprino, A. Proietti, S. Carcaterra, E. Cava, S. Cataldi, D. Cucinotta, G. Di Bella, M. Barbagallo, and A. Morrone, "Malnutrition in elderly: Social and economic determintants," *The Journal of Nutrition, Health & Aging*, vol. 17, pp. 9–15, 2013.

[2] P. R. Borum, "Disease-related malnutrition: An evidence-based approach to treatment edited by rebecca j stratton, ceri j green, and marinos elia, 2003, 824 pages, hardcover, 175. cabi publishing, wallingford, united kingdom." 2004.

[3] M. L. Joseph and A. Carriquiry, "A measurement error approach to assess the association between dietary diversity, nutrient intake, and mean probability of adequacy," *The Journal of nutrition*, vol. 140, no. 11, pp. 2094S–2101S, 2010.

[4] L. Burke, M. Warziski, T. Starrett, J. Choo, E. Music, S. Sereika, S. Stark, and M. Sevick, "Self-monitoring dietary intake: Current and future practices," *Journal of Renal Nutrition*, pp. 281–290, 2005.

[5] G. Mertes, H. Hallez, T. Croonenborghs, and B. 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*, 2015.

[6] G. Mertes, T. Croonenborghs, B. Vanrumste, and H. Hallez, "Towards detection of chewing motion in the elderly using a glasses mounted accelerometer," in *Internatinal Conference on Biomedical and Health Informatics*, 2017.

[7] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226–1238, 2005.

[8] J. Fontana and E. Sazonov, "A robust classification scheme for detection of food intake through non-invasive monitoring of chewing," in *34th Annual International Conference of the IEEE EMBS*, 2012.

[9] J. Fontana, M. Farooq, and E. Sazonov, "Automatic ingestion monitor: A novel wearable device for monitoring of ingestive behavior," *IEEE Transactions on Biomedical Engineering*, pp. 1772–1779, 2014.

[10] S. Passler and W.-J. Fischer, "Food intake activity detection using a wearable microphone system," in *Intelligent Environments (IE), 2011 7th International Conference on*. IEEE, 2011, pp. 298–301.

[11] P. Lopez-Meyer, S. Schuckers, O. Makeyev, and E. Sazonov, "Detection of periods of food intake using support vector machines," in *32nd Annual International Conference of the IEEE EMBS*, 2010.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on May 11,2020 at 14:04:07 UTC from IEEE Xplore. Restrictions apply.