Comparison of Different Classifiers for Early Meal Detection Using Abdominal Sounds

Muhammad Asaad Cheema, Salman Ijaz Siddiqui, Pierluigi Salvo Rossi
Department of Electronic Systems, Norwegian University of Science and Technology
7491 Trondheim, Norway
Email: {asaad.cheema,salman.siddiqui}@ntnu.no, salvorossi@ieee.org

2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM) | 978-1-6654-0633-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SAM53842.2022.9827841

***Abstract*—One of the challenges for the diabetic patients is to regulate the amount of glucose in the blood. Early and reliable meal detection represents one relevant issue to develop more effective treatments. This paper presents a comparison of different classifiers for early meal detection using abdominal sounds. The data presented in the paper is obtained from two different equipment and the classifiers are trained and tested on twelve recordings. The results show that neural networks and convolutional neural networks provide better average detection time (2.875 min and 2.791 min, respectively) than alternative methods recently proposed, and no false positives are observed during testing. Early and reliable meal detection eases the mental burden of the diabetic patients from documenting every meal in the controller and also reduces the risk of hypoglycemia.**

***Index Terms*—Acoustic sensors, machine learning, meal detection.**

I. Introduction

In subjects with Type 1 Diabetes Mellitus (T1DM), the pancreas produces little or no insulin. In order to regulate the blood glucose level (BGL) for TIDM patients, the insulin is infused externally. These infusions need to be administered closely which is a major concern for people with TIDM [1]. One of the actively researched solutions for this is the development of an automated system to regulate the BGL [2]. Continuous glucose monitoring (CGM) systems are used nowadays, where a controller administers the insulin infusions based on the glucose amount observed in the subcutaneous (SC) tissue. Due to slow absorption of glucose in the SC tissues, the CGM systems incur a delay of around 40 min in detecting the meal [3]. In addition to that, the current systems relies on the patients to input the meal or some information about the meal content. The patients very often forget about it and then the controller struggles to maintain the glucose levels. Hence, there is a need of a robust system which can detect the meal quickly and reliably. This will ease the mental burden of the diabetic patients. The reliability of the meal detection is also important because false meal detection may lead to a hypoglycemia.

There is a growing interest in using different sensing modalities in addition to SC glucose sensing. The focus of this paper is to use audible sounds from the gastronomical tract for early meal detection. The first attempt of integrating the sound based sensing to artificial pancreas is presented in [4]. Another sound based meal detection system is proposed in [5], which reported the meal detection time of around 10 minutes and a true positive rate of 0.5%. In this paper, three machine learning classifiers (support vector machine (SVM), neural network (NN) and convolutional neural network (CNN)) are trained and tested on the abdominal sounds to improve the detection time as well as the reliability of the meal detection. The paper is organized as follows: Section II describes the hardware and how it was used for collecting the acoustic data; Section III presents the data pre-processing for feature extraction from the raw acoustic signals and the classifiers used for meal detection; Section IV illustrates the results; and finally Section V concludes the paper.

[0]This work was partially supported by the Research Council of Norway under the project ML4ITS within the IKTPLUSS framework and under grant no. 294828 through the Centre for Digital Life Norway (digitallifenorway.org).

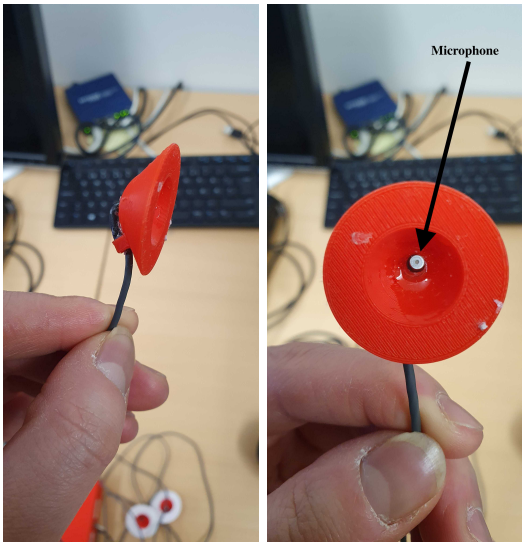


Fig. 1: Cup-shaped microphones.

II. Experimental Setup

A. Hardware

The data presented in the paper were collected from two set of stethoscopes.

The first one is the Electronic Stethoscope Model 3200 produced by 3M Littmann, which supports single channel data acquisition. The sampling frequency of this device is 4 kHz. It was applied to the lower abdomen of the subject under the umbilicus using medical tape.

Fig. 2: Complete acquisition system. The red box is the connection box for powering the microphones. The black box is the sound card for digitizing the sound signal and a laptop for saving the data.

The second one is developed by SINTEF Digital AS in Trondheim, Norway. A knowles electret condenser microphone (product number FG-23329-P07) is fitted inside a 3D-printed cup-shaped plastic cover. The hardware supports simultaneous multichannel data acquisition up to 4 channels. The sampling frequency of the microphones is 48 kHz, thus it can record variety of body sounds. Two microphones are applied on the left and right of the abdomen under the umbilicus using double sided tape.

Fig. 1 shows the cup-shaped cover and the microphone and Fig. 2 shows the complete setup of the data acquisition system, which includes the (red) connection box for powering the microphones, a sound card, and a laptop.

Five subjects were available for various types of recording, and twelve recordings are considered in this work: four were collected using Littmann stethoscope (all from a single subject) and eight using 3D-printed stethoscope.

B. Data Acquisition

A slightly different protocol was followed for data acquisition with each equipment.

In the case of Littmann stethoscope, the subject remained seated during each recording. For all the recordings, the subject fasted at least 10 hours before the recording session. Each recording lasted a total of approximately 60 minutes. Four out of five meals started 15 minutes after the start of the recording, while the last one started after 21 minutes of recording. In the case of 3D-printed stethoscope, the data were collected on four subjects. The subjects fasted at least 3 hours before the experiments and the data are collected in sitting position. In all recordings, the subject started eating 15 minutes after the start of the recording. After finishing the meal, the subject was sitting still and silently for additional forty five minutes to capture the postprandial bowel sounds.

III. Data Processing

Data-driven binary-classification methods applied to features extracted from the raw acoustic signals have been used to detect the meal.

A. Feature Extraction

The first step is to pre-process the dataset to prepare it as input to the classifiers for the training purposes. In the pre-processing phase, 30 minutes of each recording are used (15 minutes before the meal start and 15 minutes after it), and then acoustic signals are split into consecutive segments of 10 seconds with 50% overlap. Each segment is processed to produce 41 features, More specifically, *feature 1* is the total power in the range $[0, 2000]$ Hz, *feature 2* to *feature 21* represent the power in each 100-Hz-band from 0 to 2000 Hz and *feature 22* to *feature 41* represent the ratio of the powers in the respective band and the total power. Features are extracted and collected in the $S \times N$ feature matrix

$$\boldsymbol{X} = \begin{bmatrix} X_1(1) & X_2(1) & \cdots & X_N(1) \\ X_1(2) & X_2(2) & \cdots & X_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ X_1(S) & X_2(S) & \cdots & X_N(S) \end{bmatrix} , \tag{1}$$

where $X_n(s)$ represents the nth feature extracted from the sth segment, while S and N denote the numbers of segments and features, respectively.

The response vector to the feature matrix, i.e. the corresponding label information, has entries with zeros and ones where 0 (resp. 1) represents absence (resp. presence) of meal. Based on the built feature matrices, entries corresponding to the the first 15 minutes are 0s and entries corresponding to last 15 minutes are 1s, e.g.

$$\boldsymbol{y} = [0, \ldots, 0, 1, \ldots, 1]^T \ . \tag{2}$$

B. Classification

Twelve recordings are used for training, validation and testing purposes: eight are used for training and validation, while four are used for testing. Leave one out cross-validation (LOCV) is performed on the eight recording. Picking one meal out of eight gives eight combinations/folds to train and validate the classification models.

Three classification techniques are considered (SVM, NN, and CNN) for providing a soft decision on each segment, namely $d_s \in [0, 1]$ is the output of the classifier related to the sth segment. The detection time for every test meal is also observed. The binary output from the classifier is filtered according to an Exponentially Weighted Moving Average (EWMA) for a final (hopefully more reliable) decision. EWMA [6] is a sequential change detection procedure that exploits past observations and applied to reduce the number of false alarms. EWMA relies on the following equation:

$$z_s = \alpha d_s + (1 - \alpha) z_{s-1} \ , \tag{3}$$

where α is a parameter determining a tradeoff between current and past values from the classifier. The output (z_s) of the EWMA filtering is then converted to a final binary decision based on a threshold mechanism.

1) Support Vector Machine: A radial-basis-function kernel is being used with tolerance = 0.001 for training purposes. Unlike the general case of SVM (which gives output in the form of 0 and 1), probabilities of meal or no meal are taken out of the trained model to fed the EWMA equation.

2) Neural Network: The NN-based classifier is used with an input layer with 41 nodes, with 2 hidden layers made of 200 and 100 nodes, respectively, followed by a dropout layer and a single-node output. Leaky Re-Lu activation function is used between layers except for the output layer which uses the sigmoid as activation function. The sigmoid provides output values between 0 and 1. The overall structural details of the NN used are:

- Fully connected layer with output size of 200, Batch normalization, Leaky Re-Lu activation function;
- Fully connected layer with output size of 100, Batch normalization, Leaky Re-Lu activation function, Dropout layer;
- Fully connected layer with output size of 1, Sigmoid activation function.

The means square error is used as the objective function for training the NN and the stochastic gradient descent is used with a batch size of 16 and a learning rate of 10^{-4}.

3) Convolutional Neural Network: The CNN-based classifier is built with one 1D-convolutional layer for information extraction and two fully connected layers for classification purpose. Leaky Re-Lu activation function is used after each convolution layer and fully connected layer. Just like for the NN, the sigmoid has been used to produce output values between 0 and 1. Dropout layer has been used just before the output layer to avoid over-fitting. The overall structural details of the CNN used are:

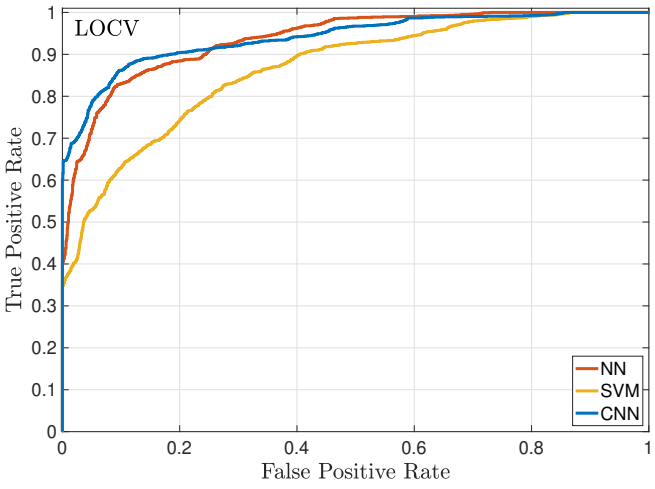
- 1d Conv with 32 channels and kernel size of 5, Batch normalization, Leaky Re-Lu activation function;
- Average pooling with window of size=2 and stride=2;
- Fully connected layer with output size of 50, Batch normalization, Leaky Re-Lu activation function, Dropout layer;
- Fully connected layer with output size of One, Sigmoid activation function.

The mean square error is used as the objective function for training the CNN and the stochastic gradient descent with the batch size of 16 and learning rate of 10^{-4}.

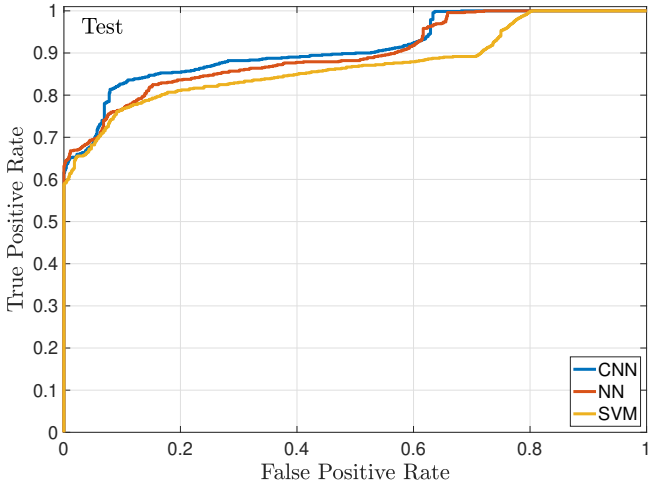
C. Performance Metrics

Performance of the different considered techniques is assessed and compared in terms of probability of detection (true positive rate), probability of false alarm (false positive rate), and detection time. Detection time is obtained by calculating the difference in the meal start between the response vector given by Eq. (2) and the predicted vector from the trained model. The meal start is defined as the time instant when the label vector transit from 0 to 1.

A false alarm is considered if a meal is predicted from the trained model, while the actual labels from the response vector show no meal for some input.



(a) LOCV.



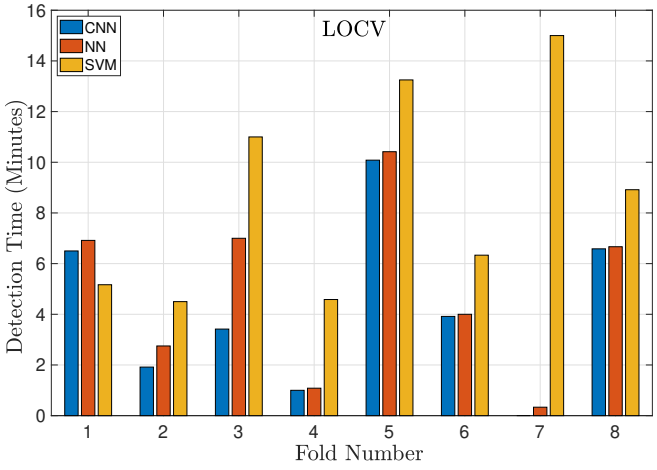
(b) Test meals.

Fig. 3: ROC curves.

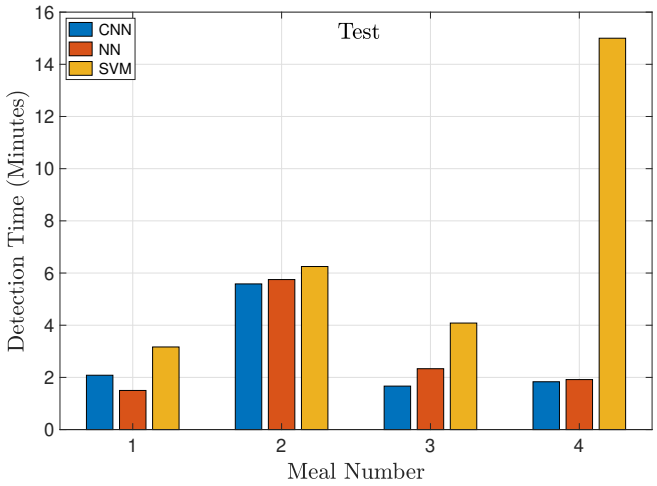
IV. Numerical Results

The classifiers are implemented in python using Pytorch and Scikit-learn packages. The value of the parameter α in eq. (3) is set to 0.05. The Receiver Operating Characteristic (ROC) curve for LOCV is obtained by averaging the validation performance achieved on each fold. ROC curves are computing based on binary quantization of the soft values provided by the EWMA post-processing. Moreover, if a classifier is unable to detect the meal, we put a penalty of 15 minutes to its detection time. In this analysis, the focus is on reducing the number of false alarms at the expense of misdetection or higher detection time. This is because the false alarm leads to serious condition like hypoglycemia thus having higher cost compared to other events.

Fig. 3 shows the ROC for each classifier. In the case of LOCV, the CNN-based classifier is performing better than the other two type of classifiers. For small false positive rates,



(a) LOCV.



(b) Test meals.

Fig. 4: Meal detection time.

CNN-, NN- and SVM-based classifiers have true positive rates close to 63%, 40% and 37%, respectively. For larger false positive rates larger, the increase of the corresponding true positive rate is more significant for CNN- and NN-based classifiers then for the SVM-based classifier. A similar trend can also be seen when considering the performance on the test set, although the gaps between all 3 types of classifiers is reduced.

Fig. 4 shows the detection time for validation and test meals. The height of each bar represents the detection time in minutes and different colors represent each classifier. The fold number is on the x-axis. The detection time refers to models operating with minimum false alarm rate (via proper selection of the threshold value). During LOCV, the models are saved for each fold. For testing, one model is selected such that the detection time is approximately equal to the average detection time of all the models. The selected model is tested on four meals and the detection time is calculated for each meal.

TABLE I: Meal Detection Result Summary

Classifiers	SVM	NN	CNN
No. of Training Meals (Each Fold)	7	7	7
No. of Validating Meals (Each Fold)	1	1	1
No. of Testing Meals	4	4	4
No. of False Alarms (Test)	0	0	0
No. of False Alarms (LOCV)	0	0	0
Avg. Detection Time in Min (Test)	7.125	2.875	2.791
Avg. Detection Time in Min (LOCV)	8.593	4.895	4.177

Apparently, the detection time is much higher in the case of SVM-based classifiers than NN- and CNN-based classifiers, except for fold number 1. The detection time of the SVM-based classifier is 15 minutes in two cases: one during the testing and one during the training. These cases correspond to events in which the classifiers completely missed the meal. The average meal detection times for the SVM-, NN- and CNN-based classifiers during LOCV are 8.593 min, 4.895 min and 4.177 min, respectively, while 7.125 min, 2.875 min and 2.791 min, respectively, during testing.

Table I summarises the results from the performed experiments and corresponding analysis. CNN-based classifiers have the best performance both in terms of ROC and in terms of detection time. NN-based classifiers perform very close to the CNN, while SVM-based classifiers have much higher detection times and worse ROC. The threshold and the corresponding operating point on the ROC have been selected in order to achieve null false alarm rate in the training phase. Trained classifiers have kept a null false alarm rate on the test set as well. The results show a significant improvement as compared to [5] by reducing the detection time to half and lesser probability of false alarm.

V. Conclusion

The paper compares the meal detection performance of different classifiers using abdominal sounds. The data presented in the paper consists of 12 recordings from 5 subjects. The data are acquired using two set of stethoscopes. The recordings are split in 8 and 4 sets used for training/validation and testing, respectively. LOCV is performed for training and validation and the optimal model is selected for testing. The average meal detection times for SVM-, NN- and CNN-based classifiers during LOCV are 8.593 min, 4.895 min and 4.177 min, respectively, while 7.125 min, 2.875 min and 2.791 min, respectively, during testing. The results show that CNN- and NN-based classifiers exploit the relationship between features and the meal intake more effectively, thus exhibiting reduced detection time and improve reliability for meal detection.

References

[1] Sverre Christian Christiansen, Anders Lyngvi Fougner, Øyvind Stavdahl, Konstanze Kölle, Reinold Ellingsen, and Sven Magnus Carlsen. A review of the current challenges associated with the development of an artificial pancreas by a double subcutaneous approach. *Diabetes Therapy*, 8(3):489–506, 2017.

[2] Thomas Peyser, Eyal Dassau, Marc Breton, and Jay S Skyler. The artificial pancreas: current status and future prospects in the management of diabetes. *Annals of the New York Academy of Sciences*, 1311(1):102–123, 2014.

[3] Sediqeh Samadi, Mudassir Rashid, Kamuran Turksoy, Jianyuan Feng, Iman Hajizadeh, Nicole Hobbs, Caterina Lazaro, Mert Sevil, Elizabeth Littlejohn, and Ali Cinar. Automatic detection and estimation of unannounced meals for multivariable artificial pancreas system. *Diabetes technology & therapeutics*, 20(3):235–246, 2018.

[4] Khandaker A Al Mamun and Nicole McFarlane. Integrated real time bowel sound detector for artificial pancreas systems. *Sensing and biosensing research*, 7:84–89, 2016.

[5] Konstanze Kölle, Anders Lyngvi Fougner, Reinold Ellingsen, Sven Magnus Carlsen, and Øyvind Stavdahl. Feasibility of early meal detection based on abdominal sound. *IEEE Journal of Translational Engineering in Health and Medicine*, 7:1–12, 2019.

[6] Liyan Xie, Shaofeng Zou, Yao Xie, and Venugopal V Veeravalli. Sequential (quickest) change detection: Classical results and new directions. *IEEE Journal on Selected Areas in Information Theory*, 2(2):494–514, 2021.