Allergy detection with statistical modelling of HRV-based non-reaction baseline features*

Niall Twomey
Department of Electrical and Electronic Engineering
University College Cork, Ireland
niallt@rennes.ucc.ie

Andrey Temko
Department of Electrical and Electronic Engineering
University College Cork, Ireland
andreyt@rennes.ucc.ie

Jonathan O'B. Hourihane
Department of Paediatrics and Child Health
University College Cork, Ireland
j.hourihane@ucc.ie

William P. Marnane
Department of Electrical and Electronic Engineering
University College Cork, Ireland
l.marnane@ucc.ie

ABSTRACT

This paper investigates the automated classification of oral food challenges ('allergy tests'). The electrocardiograms (ECG) of the subjects being tested for allergies were recorded via a wireless mote, and the QRS complexes were manually annotated and 18 features were extracted from the signals. Principal component analysis was used for feature decorelation and dimensionality reduction and diagonal covariance Gaussian mixture models were used to model non-reaction baseline patient condition. The generated subject independent log likelihood plots were used to separate allergic reaction by means of subject adaptive thresholding.The platform resulted in 87% accuracy of classification with 100% specificity. The algorithm presented can detect allergy up to 30 minutes sooner than the current state of the clinical art allergy detection ($7\text{minutes} \pm 9$).

1. INTRODUCTION

20,000 children in Ireland are allergic to a food type, and the incidences of allergy are increasing. The oral food challenge is the definitive diagnostic test for food allergies and involves the supervised and controlled ingestion of a potential medical poison by a subject. They are, by nature, stressful for the subject and the subject's family and have the inherent risk of allergic reaction. Even when supervised by experienced staff, trained to recognise and prevent severe reactions before they manifest, $3-11\%$ of oral food challenges end in anaphylaxis [6], an acute and potentially fatal allergic reaction if untreated.

During an oral food challenge one portion of the suspect food type is divided in five sub–portions. The smallest portion is the first consumed by the subject. Once consumed they are then observed for 20 minutes. If the onset of a reaction is suspected the vitals of the subject are recorded—heart rate, blood pressure, blood oxygen saturation level and temperature. Stomach pain, hive outbreak, vomiting, swelling and wheezing can be symptoms of an allergic reaction. If any of these symptoms are observed, or if the subject's vital signs have changed significantly since they were last recorded, the test is concluded and the subject has been proven allergic to the tested food and has failed the test. If no reaction has occurred in the 20 minutes the next largest sub–portion is consumed by the subject. This process is repeated until the portion has been fully ingested. If no reaction to the food type occurs the subject passed the oral food challenge.

Only one medical clinic in Ireland performs oral food challenges. As a result subjects must wait up to 18 months for verification of allergy. This reduces the quality of life as the family of the subject living with with then unverified fears of the danger of contact with a possible allergen. Clinical experience has shown that for a subset of the subjects tested heart rate variability (HRV) features tend to change before a physical manifestation of the allergic reaction is observed.

A classification framework based on heart rate variability features is presented. A wireless device is attached to the subjects undergoing the oral food challenges 10 minutes before administration of the first dose of the problem food. This device streams ECG to a nearby computer, and the signal trace is stored in a database. 18 HRV features are extracted and Gaussian mixture models (GMM) are employed for modelling of non-reactive basilene condition data. The lack of detailed time annotation represents the main chal-

ISABEL '11, October 26-29, Barcelona, Spain

lenge of the task and prohibits the application of two-class classifiers. In fact, only the first 10 minutes of data before consumption of the first food portion can securely represent the non-reaction condition. Likelihoods are then processed by subject adaptive thresholding to automatically separate 'pass' and 'fail' subjects.

The use of real–time classification of physiological signals during oral food challenges has the ability to reduce the length of the test and the extent of the reaction. If adopted, a platform such as this could reduce the risk of severe reactions (anaphylaxis) and encourage more hospitals to offer food challenge clinics, thereby reducing the waiting list for the procedure.

2. PROCEDURE

2.1 Data collection

Ethical approval was secured from the Clinical Research Ethics Committee of the Cork Teaching Hospitals to monitor the subjects during the oral food challenge. The SHIMMER [3] (Sensing Health with Intelligence, Modularity, Mobility, and Experimental Reusability) wireless mote was used for data collection. It utilises an ECG daughterboard and Bluetooth connectivity to stream ECG data to a PC in real time. Informed parental approval was obtained to record the ECG of a subject on a subject by subject basis.

After written parental and child consent was obtained, a SHIMMER is attached to a subject via a custom made neoprene strap. The electrocardiogram is recorded with three electrodes (arranged in the Einthoven triangle configuration [5]). The leads plug into the SHIMMER via an ECG daughterboard, and live ECG signals are transmitted via a Bluetooth connection to a nearby computer where it's stored and added to a database.

28 subjects (15 male, 13 female) were monitored and their ECG recorded. Of these the ECG trace of four subjects was corrupted by ECG artefacts to the point where consistent manual and automatic QRS extraction was impossible. Fifteen of the remaining subjects failed the oral food challenge (they were diagnosed as being allergic to the food type once the test was completed) and nine subjects passed the food challenge. The four subjects whose ECG was corrupted passed oral food challenge. The twenty four subjects whose ECG was annotated are henceforth referred to as the 'allergy database'.

Table 1 tabulates the characteristics of the subjects that populate the allergy database. They are categorised according to the result of the food challenge and randomly arranged (subjects 1—15 failed and subjects 16—24 passed the oral food challenge).

34 hours of the allergy database ECG data were fully and manually annotated. The QRS points extracted from the N=24 subjects are referred to as the 'QRS database' henceforth.

2.2 Performance assessment and metrics

It is necessary to develop a subject–independent allergy detection platform so as not to bias the results towards a specific subject and leave-one-out cross validation was used to provide this. With this procedure a test subject is selected from Table 1 and the remaining training subjects data are used for classification model generation. Leave-one-out is an almost ubniased estimation of true generalisation error

Table 1: Characteristics of the subjects populating the allergy database

Index	Gender	Age	Allergen	Result
1	male	1.5 years	wheat	
2	male	6 years	peanut	
3	male	9 years	egg	
4	male	1 years	milk	
5	male	8 years	peanut	
6	female	9 years	peanut	
7	male	6 years	soy	
8	male	5 years	peanut	FAIL
9	female	8 years	egg (cake)	
10	male	3 years	milk	
11	female	6 years	peanut	
12	female	5 years	milk	
13	female	3 years	milk	
14	male	8 years	soy	
15	female	9 months	wheat	
16	male	6 years	egg	
17	male	10 years	egg (cake)	
18	female	4 years	soy	
19	male	6 years	peanut	
20	female	1.5 years	milk	PASS
21	female	7 months	milk	
22	male	1 years	milk	
23	female	4 years	wheat	
24	male	2 years	peanut	

Sensitivity, specificity and accuracy figures were computed in order to gauge the performance of the automatic allergy detection framework.

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{2}$$

$$\text{accuracy} = \frac{\text{number of correct decisions}}{\text{total number of subjects}} \tag{3}$$

The true positives (TP) are the subjects who failed the test which were correctly classified as 'fail,' false negatives (FN) are the subjects who failed the challenge but whose reaction was not detected by the framework presented. True negatives (TN) are the subjects who failed the challenge and were correctly classified as having failed, while the false positives (FP) are subjects who have passed the challenge but were flagged 'fail' by the classification system [4].

2.3 Features

The following features were extracted from the manually QRS database:

- Time-domain features:
  - Mean heart rate
  - Standard deviation of heart rate
  - Coefficient of variation of heart rate
  - RMSSD of heart rate
    Root mean squared successive difference.
  - PNN25/PNN50 of heart rate
    Percentage of successive QRS points that differ by more than 25ms or 50ms.

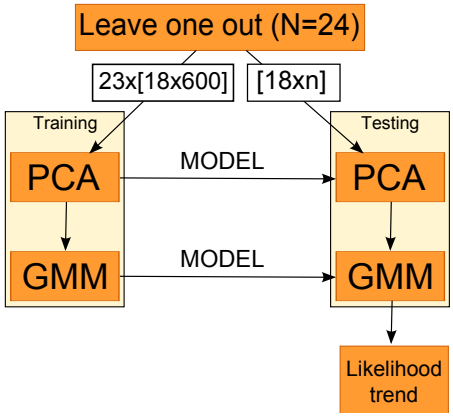


Figure 1: Independent performance assessment

  - Sequential trend analysis
  - Poincaré CSI/CVI The cardiac sympathetic index and the cardiac vagal index.

- Frequency-domain (computed via Lomb-periodogram[1]):
  - Very low frequency power
  - Low frequency power
  - High frequency power
  - High to low power ratio

These features were chosen due to their inclusion in the heart rate variability standards taskforce [2] and topical heart rate variability literature. The features were extracted over sixty second epochs with one second increments (98% overlap).

2.4 Statistical modelling

When a reaction is detected by a trained allergist antihistamines can be administrated. Upon administration the histamines in the bloodstream reduce and the heart rate variability features will return to their normal range. For classification purposes it is not possible to annotate both 'normal' and 'reaction' training data as the amount of the 'reaction' data will always kept to a minimum, and because it is not possible to determine when a reaction begins. Thus the first ten minutes of the challenge (referred to as baseline and background interchangeably) is the only annotated data available.

Gaussian mixture models were employed to compute the likelihood that the data from the rest of the challenge is generated by the baseline model. To falicilate the estimation problem, PCA was first used to decorrelate and reduce the dimension of features maintaining 95% of the information. GMM with diagonal covariance is subsequently used.

Figure 1 illustrates the procedure used to train the subject independent GMM classifier incorporating PCA with diagonal covariance. A leave-one-out framework is introduced where PCA and GMM models are computed for the i^{th} subject based off the background data of the remaining subjects, $1 \geq i \leq N$, where N is the number of subjects in the QRS database.

The i^{th} subject is selected and their feature vectors are stored in memory (data is of the dimensions $[18 \times n]$, where n is the length of the subject's feature vector). For the remaining subjects, the feature vectors pertaining to the baseline HRV characterisation (first ten minutes) are loaded. This data is of dimensions $(N-1) \times [18 \times (600+m)]$, where m represents the initial checkup time. This is included in the training data because the classification system must be aware of normal variations of the heart rate. This data then trains a GMM model consisting of four Gaussians after the PCA stage.

The feature vectors of the i^{th} subject then undergo PCA and GMM stages, based on the models generated by the leave-one-out stage, resulting in a non–reaction log likelihood data series—the smaller the likelihood the less similar to the background data is. This leave-one-out stage was repeated for every subject resulting in 24 log likelihood series.

2.5 Post processing

Subject adaptive post-processing was employed in order to classify the likelihoods as 'fail' or 'pass'. The first 600 likelihood points of each subject will represent the likelihood that the subject is not reacting to an allergen, as the allergen has not yet been administered (features were extracted from epochs of one minute in length with one second increments). It is with the mean and standard deviation of the normalisation data that the patient adaptive thresholding is performed.

The likelihood vector is of dimensions $[1 \times n]$, where n is the length of the the i^{th} subject's feature vector. A likelihood is available for each element of the vector, and the lower the likelihood is the less likely it is that it belongs to the background dataset.

The criteria for classifying a subject as having failed a challenge is that the likelihood vector for the i^{th} subject must remain below a constant threshold, th_i, for a duration of d_i. The inclusion of the d_i parameter facilitates rejection of short–term spikes in the likelihood series (the variation of the data series is $\sigma = 130$ with frequent spikes present). The values of th_i and d_i were computed by an internal leave-one-out stage

2.6 Model selection with internal leave-one-out

The th_i and d_i parameters were computed in a subject independent leave-one-out search that maximises the accuracy of detection for the i^{th} subject based on the results of the other subjects. They were swept over a finite range ($0 \leq \text{th}_i \geq 50$, $0 \leq \text{d}_i \geq 100$s) for each allergy likelihood series of the remaining subjects. For each point of the optimisation search, a subject was classed as 'fail' if the likelihood vector fell below th_i threshold for a duration of d_i. This automatically computed result is then compared to the reference label (the diagnosis of the allergist).

With this performed for all the remaining subjects overall sensitivity (1), specificity (2) and accuracy (3) can be computed. The effect of th_i and d_i on the accuracy of allergy detection in the internal LOO can be visualised in Figure 2. The specific values of th_i and d_i which maximise the accuracy are chosen for accessing the likelihood plots of the i^{th}

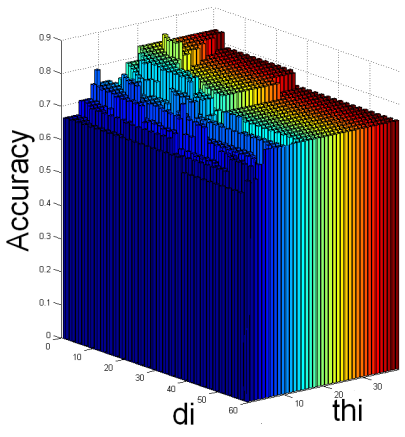


Figure 2: Effect of th_i and d_i on accuracy of allergy detection

subject.

3. RESULTS AND DISCUSSION

The allergy detection platform described detected allergy with 87.5% accuracy with 100% specificity. Interestingly, it was also observed that the detection algorithm would have classed subjects as 'fail' an average of 7 minutes sooner than the allergists supervising the challenges, see Table 2.

Four sample log likelihood plots are shown in Figure 3 based off the optimised parameters of both of the leave-one-out stages. The blue trace is the computed log likelihood, the green highlighted regions are where the subject was undergoing a checkup (these regions are omitted from analysis), and the black horizontal line are the subject's optimised thresholds. The red dots in subfigure 3 (b) represent points which have been flagged 'fail' by the classification process.

Figure 3 (a) shows a subject whose heart rate variability features did not change during the course of their challenge even though they failed the food challenge. It is common for the heart rates to change during the checkups that the subjects undergo. This subject is a member of a subset of subjects whose heart rate variability features—and therefore likelihood—do not vary from their baseline parameters to a degree sufficient enough to classify them as 'fail' with the classification system presented. It is not possible to classify these subjects based on the framework proposed without further information.

Figure 3 (b) shows a subject whose likelihood data series does change during the course of the oral food challenge. The subject also failed the challenge, and was diagnosed as being allergic to the food type they were tested against. This subject is party to a subset of subjects whose features changed prior to the onset of a reaction. The likelihood points that are considered as 'fail' are flagged with red circles. One point (shown within the green circle at 71 minutes) did not fall under the threshold for a sufficiently long period of time to be flagged as 'fail.'

It can be seen in Figure 3 (b) that the trend of the likelihood

Table 2: Time-gain of allergy detection platform

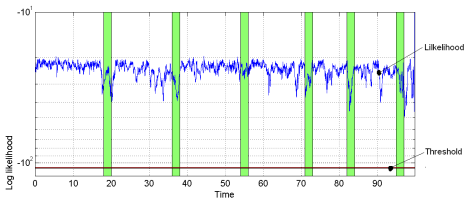
Subject	Time gain
1	no detection
2	no detection
3	10 minutes
4	20 minutes
5	no detection
6	6 minutes
7	5 minutes
8	3 minutes
9	8 minutes
10	no detection
11	30 minutes
12	20 minutes
13	no detection
14	no detection
15	5 minutes

plot falls below the fail criteria 30–minutes prior to the test being halted. With the allergy detection system proposed, the oral food challenge could have been halted at this time. This early detection was seen for all the subjects who failed the challenge, the 'pass' subjects were not included. Table 2 summarises the gain in time for the system. Allergy was detected an average of 7 ± 9 minutes sooner than the current state for subjects whose allergic reaction was detected. The significance of early detection is that rescue medicines can be administered sooner, thereby reducing the effect of the allergic reaction.

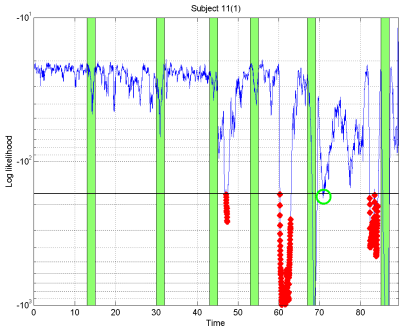
Figure 3 (c) and (d) show the likelihood plots of two subjects who passed the oral food challenge. Each plot is on the same scale. Subject 22 (c) presented with a likelihood chart which varied significantly ($\mu = -33, \sigma = 13$). The reason for the comparatively large variation is due to the fact that the subject when presented with the problem food gagged on it and was unable to tolerate consumption. In the absence of objective signs of an allergic reaction, this response was considered psychogenic, and the the challenge was considered negative. The variation of the signal was rejected by the allergy detection system.

Subject 20 also passed the test, but presented with the largest likelihood changes over the allergy database. These were registered when the subject underwent their periodic checkups and occurred because the subject became very agitated when the clinical staff approached. The agitation progressed to the point where the SHIMMER was removed in an attempt to minimise stresses imposed on the subject. However, as the classification system is trained on the features extracted during checkups, the large variation did not result in false positive classifications.

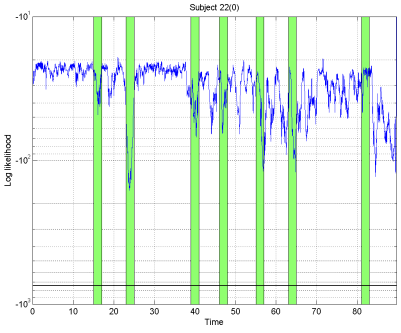
The optimised th_i and d_i resulted in 100% specificity. From a subjective human point of view, the 100% specificity suggests that conficence may be placed in the 'fail' classifications of the classification platform. Maintaining 100% specificity is the most important constraint to preserve. HRV is the only feature computed in diagnoses by this system, but allergists have more to hand (e.g. mood, blood pressure, etc). Allergy does not always present with changes in HRV features (as demonstrated in Figure 3 (a)) so it is not always possible to detect allergy. Therefore it is foreknown that allergy might not be detected with HRV features by even a perfect platform, so 100% sensitivity is unattainable.



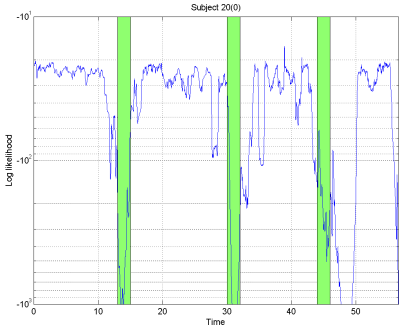
(a) Subject 2, fail, no reaction detected



(b) subject 11, fail, reaction detected



(c) Subject 22, pass, no reaction detected



(d) Subject 20, pass, no reaction detected

Figure 3: Reaction loglikelihood plots for subjects 2 (a), 11 (b), 22 (c) and 20 (d)

Furthermore, if confidence is given to the results of the classification, false positives will result in unrequired and undesired effects on the quality of life of the subject and their family. It is important that highest specificity is maintained in order to give confidence to the results of the allergy detection platform that is designed.

Three subjects who passed the oral food challenge presented with abnormal ECG traces. Subject 20 has gastrointestinal reaction when consuming new or unliked foodtypes. Subject 22 became agitated during the challenge, resulting in large changes in likelihoods. Subject 21 presented with frequent single isolated ectopic beats during the oral food challenges (the heart rate could rise to 200 beats per minute for a single beat and then return to its normal rate). These irregularities can induce large changes in likelihood series. However, the classification resulted in 100% specificity and proves a robust means of classification.

4. CONCLUSION

The results presented here show that allergy can be detected with classification of heart rate variability features with 100% specificity.

When the system described is incorporated into a real time allergy detection platform the food challenges could be halted up to 30 minutes sooner than the current state of the clinical art. Confidence can be attributed to the results of the allergy classification due to the fact that 100% specificity was obtained with subject independent classification and optimised subject independent parameter selection for subject adaptive post processing. In an allergy detection environment allergists have access to more signals (mood, blood pressure etc) and will be able to use these to assist in the diagnosis and the platform presented could be another that might assist during the food challenges as a decision support tool.

A robust classification framework is demonstrated which attains 100% specificity. This specificity was maintained even with three subjects whose heart rate variability features were highly varied. There is great clinical potential to use these findings to make oral food challenge safer and therefore more widely available. If validated in this setting this finding could be applied in other areas of dynamic interventional testing in many branches of medicine.

5. REFERENCES

[1] N. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science*, 39(2):447–462, 1976.

[2] M. Malik. Heart rate variability. *Annals of Noninvasive Electrocardiology*, 1(2):151–181, 1996.

[3] SHIMMER-research. Shimmer homepage. http://www.shimmer-research.com/, 1 2010.

[4] V. Vapnik. *Estimation of dependences based on empirical data*. Springer-Verlag New York Inc, 2006.

[5] F. Wilson, F. Johnston, and C. Kossmann. The substitution of a tetrahedron for the einthoven triangle* 1. *American Heart Journal*, 33(5):594–603, 1947.

[6] Y. Yanishevsky and J. O. Hourihane. Differences in treatment of food challengeŰinduced reactions reflect physicians' protocols more than reaction severity. *The Journal of allergy and clinical immunology*, 126:182, 2010. 10.1016/j.jaci.2010.03.038.