Automated Detection of Perturbed Cardiac Physiology During Oral Food Allergen Challenge in Children

N. Twomey, A. Temko, *Senior Member, IEEE*, J. O'B. Hourihane, and W. P. Marnane

***Abstract*—This paper investigates the fully automated computer-based detection of allergic reaction in oral food challenges using pediatric ECG signals. Nonallergic background is modeled using a mixture of Gaussians during oral food challenges, and the model likelihoods are used to determine whether a subject is allergic to a food type. The system performance is assessed on the dataset of 24 children (15 allergic and 9 nonallergic) totaling 34 h of data. The proposed detector correctly classified all nonallergic subjects (100% *specificity*) and 12 allergic subjects (80% *sensitivity*) and is capable of detecting allergy on average 17 min earlier than trained clinicians during oral food challenges, the gold standard of allergy diagnosis. Inclusion of the developed allergy classification platform during oral food challenges recorded would result in a 30% reduction of doses administered to allergic subjects. The results of study introduce the possibility to halt challenges earlier which can safely advance the state of clinical art of allergy diagnosis by reducing the overall exposure to the allergens.**

***Index Terms*—Automated diagnosis, decision support, machine learning, novelty detection.**

I. Introduction

IT is estimated that 5% of infants under three years of age [1], and 3–4% of adults [2] worldwide have food allergies. The oral food challenge (OFC) is the definitive diagnostic test for food allergies [3] and involves the supervised and controlled ingestion of the implicated food allergen. OFCs are, by nature, stressful for the subject and the subject's family and have the inherent risk of provoking allergic reactions as required by the diagnostic task. Even when supervised by experienced staffs who are trained to recognize, prevent, and manage allergy as it manifests, 3–11% of OFCs end in anaphylaxis [4]—an acute and potentially fatal allergic reaction if untreated [5]. This paper aims at automatic detection of allergy which can reduce the consumption of allergens and improve patient safety.

Manuscript received November 1, 2012; revised April 29, 2013; September 12, 2013;

This work was supported by an IRCSET enterprise partnership with Intel, and by Science Foundation Ireland (SFI/07/SRC/I1169).

N. Twomey, A. Temko, and W. P. Marnane are with the Department of Electrical and Electronic Engineering, University College Cork, Cork, Ireland (e-mail: niallt@rennes.ucc.ie; aatemko@gmail.com; l.marnane@ucc.ie).

J. O'B. Hourihane is with the Department of Paediatrics and Child Health, University College Cork, Cork, Ireland (e-mail: J.Hourihane@ucc.ie).

Digital Object Identifier 10.1109/JBHI.2013.2290706

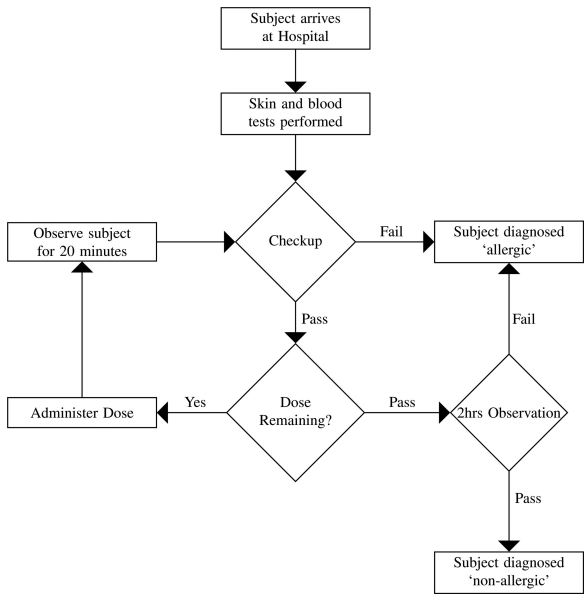

Fig. 1. OFC diagnosis flowchart.

OFCs are conducted in a hospital environment, and the route to a diagnosis of allergy during OFCs is outlined in Fig. 1. In this Figure, the term "fail" is used to signify that allergic reactions have occurred and the subject would be diagnosed allergic to the allergen, and "pass" is used to identify that no reaction has occurred. The subject (exclusively children in this study) arrives in the hospital with their parents. Skin and blood tests are performed on the subject [6] to give a baseline indication of the subject's susceptibility toward the potential allergen.

One age-appropriate portion of the suspect food type is divided into five subportions which double in size from 1/32 to 1/2 of a portion. The smallest subportion is first consumed by the subject and they are observed for 15–20 min by the allergist after. During this period, the allergist observes the subject for signs that an allergic reaction might be occurring. Symptoms of an allergic reaction can include stomach pain, hive outbreak, vomiting, swelling, and wheezing.

After this observation period, the subject is given a checkup by the allergist. The vital signs of the subject, such as heart rate, blood pressure, blood oxygen saturation level, and temperature, are recorded manually at this stage. If any typical symptoms of allergy are observed the test is concluded and the subject has been proven allergic to the tested food. If no

reaction has occurred the next subportion is consumed by the subject.

This process is repeated until all the portions have been consumed in totality. After this total consumption, the subject is observed for a 2-h time period, as delayed reactions may occur. The nature of reactions which require termination of OFCs need not be severe. For example, the swelling of a subject's lip during a peanut challenge is sufficient for allergy to be confirmed. Indeed, OFCs will be terminated at the first sign of allergy, as untreated reactions will leave subjects in discomfort and in danger of provoking a more serious reaction.

It can be seen that an OFC represents a challenging diagnostic task. The use of real-time classification of physiological signals during OFC has the ability to reduce the length of the test, the extent of the reaction, and improve the quality of life of the subjects [7]. The blood pressure and blood oxygen saturation levels are routinely collected as a feature of many medical challenge test involving incremental exposures. These measurements (known as "vital signs") do not change until very late in the event. It can be understood that the body's physiological response to stress is that adaptations will be made to ensure that the oxygen levels and blood pressure will be preserved at all costs. That is why, when they are seen to change, the patient is clearly decompensating. The electrocardiogram (ECG), however, is more sensitive to physiological stress. The measurement and quantification of ECG and heart rate variability (HRV) changes during OFCs could provide a major safety dividend in the field of *in vivo* dynamic allergy testing as subjects could be exposed to a lesser quantity of the allergen during their diagnoses [8].

It has been clinically shown that the change before a physical manifestation of the allergic reaction changes can be observed in characteristics of HRV [9]– [11], but obtaining this has not before been investigated in a fully automated manner. The ECG has also previously been shown to be successful for diagnosing cardiac disease [12], sleep apneas [13], and for recognizing valid arterial-blood-pressure pulses [14], and in this paper, we show it is also well suited for helping diagnose food allergy.

In this paper, the normal (i.e., nonallergic) HRV features are modeled with Gaussian mixture models (GMMs). This is a well-known density estimation technique which establishes a multidimensional representation of the underlying distribution of data. The likelihood that new features are normal is computed, with lower values of likelihood indicating that the features are less likely to belong to the background distribution. The boundaries of normality (i.e., the regions beyond which data can be considered abnormal) are established. When likelihoods are outside of these boundaries, the features which were obtained are novel with regard to the background data, and under these conditions allergy is classified.

The automated classification procedure, however, cannot replace allergists during OFCs, as they will always be required to administer doses of the allergens and recovery medicines throughout. Therefore, the design criteria for the allergy detector must complement the allergists during the diagnosis procedure, and in order to not introduce negative effects to the quality of life of the subjects, it is imperative that false positive classifications are minimized as much as possible.

TABLE I
CHARACTERISTICS OF THE SUBJECTS POPULATING THE ALLERGY DATABASE

Index	Gender	Age	Challenge length (minutes)	Allergen	Result
1	male	1.5 years	13	wheat	Allergic
2	male	6 years	95	peanut	
3	male	9 years	90	egg	
4	male	1 years	100	milk	
5	male	8 years	120	peanut	
6	female	9 years	33	peanut	
7	male	6 years	57	soy	
8	male	5 years	100	peanut	
9	female	8 years	50	egg (cake)	
10	male	3 years	82	milk	
11	female	6 years	85	peanut	
12	female	5 years	40	milk	
13	female	3 years	105	milk	
14	male	8 years	69	soy	
15	female	9 months	81	wheat	
16	male	6 years	91	egg	Non-allergic
17	male	10 years	125	egg (cake)	
18	female	4 years	125	soy	
19	male	6 years	130	peanut	
20	female	1.5 years	98	milk	
21	female	7 months	44	milk	
22	male	1 years	93	milk	
23	female	4 years	81	wheat	
24	male	2 years	58	peanut	

II. METHODS

A. Data Collection

Ethical approval was secured from the Clinical Research Ethics Committee of the Cork Teaching Hospitals to monitor subjects undergoing OFC. Informed parental approval was obtained to record the ECG of the subjects. The ECG electrodes were arranged in the Einthoven triangle configuration [15], and were plugged into the ECG daughterboard of a SHIMMER [16] device. The ECG data were streamed to a PC via Bluetooth connectivity in real time.

In total, 24 subjects (14 male, 10 female) were monitored. Fifteen of these subjects were diagnosed food allergic and nine subjects were diagnosed nonallergic by OFC. Table I tabulates the index, gender, age, challenge length, allergen, and result of the subjects in the database.

B. Allergy Detector

Fig. 2 illustrates the classification procedure of the developed allergy detector. From ECG data recorded during OFCs, QRS points are extracted by an automatic QRS detector. HRV features are extracted from these points. A GMM is then utilized to continuously estimate the likelihoods that the subject has no allergic symptoms during the OFC. After postprocessing the likelihoods, a decision was made as to whether the subject is allergic to a food type or not.

1) QRS Extraction: In this study, the QRS points of the ECG traces of the subjects in Table I were automatically extracted via a QRS detection algorithm based on the Hilbert transform [17]. This algorithm first filters and differentiates the ECG signal to reduce noise and the contribution of ECG baseline wander,

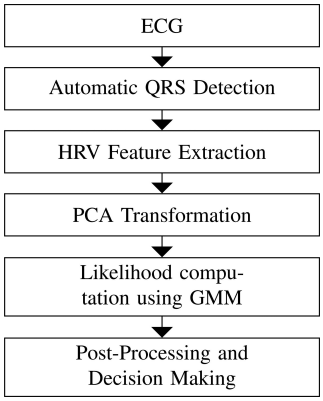

Fig. 2. Automatic allergy detection procedure.

TABLE II
FREQUENCY DOMAIN FREQUENCY BANDS

Type	Frequency range (Hz)	Origin
HF	0.15 – 0.4	Parasympathetic, respiratory sinus arrhythmia
LF	0.04 – 0.15	Sympathetic + parasympathetic
VLF	0.0033 – 0.04	Sympathetic, chemoreceptors, thermoregulation, endocrine

P- and T-waves in the signal. The Hilbert transform is then performed on this dataset and the envelope of this signal is computed. Dynamic thresholding is utilized on the envelope and QRS validation rules [18] are incorporated in order to increase the QRS detection accuracy.

2) Features: Eighteen features were extracted from the QRS points. These features were chosen due to their inclusion in the HRV task-force and by relevant HRV researchers [19]. The features were extracted every second from a sliding window (epoch) of 60 s. The features which were extracted were:

1) Mean heart rate (HR) ($n = 1$).
2) Standard deviation of HR ($n = 1$).
3) Coefficient of variation of HR ($n = 1$).
4) Root mean squared successive difference of HR ($n = 1$).
5) PNN25/PNN50 and NN25/NN50: Percentage and sum of successive QRS points that differ by more than 25 or 50ms. ($n = 4$).
6) Sequential trend analysis: Percentage of successively increasing and decreasing heart rates. ($n = 1$).
7) Poincaré: The length of the transverse and perpendicular axes of the Poincaré plot, and the cardiac sympathetic cardiac vagal indices of the HR. ($n = 4$).
8) Histogram: The ratio of the most frequent RR interval to the number of beats of the set of QRS points in the epoch. ($n = 1$).
9) Frequency domain: The normalized powers in the very-low, low and high frequency bands specified in Table II. The ratio between the low and high frequencies were also selected. ($n = 4$).

Four features were calculated from the frequency domain, which was computed by the Lomb Periodogram [20]. The Lomb Periodogram computes spectral power from nonperiodically sampled data, and is defined by

$$P_x(f) = \frac{1}{2\sigma_l^2}\left\{\frac{\left[\sum_{k=1}^{K}(x(t_k)-\overline{x})\cos(2\pi f(t_k-\tau))\right]^2}{\sum_{k=1}^{K}\cos^2(2\pi f(t_k-\tau))} + \frac{\left[\sum_{k=1}^{K}(x(t_k)-\overline{x})\sin(2\pi f(t_k-\tau))\right]^2}{\sum_{k=1}^{K}\sin^2(2\pi f(t_k-\tau))}\right\} \quad (1)$$

where time shift insensitivity factor, τ, is given by

$$\tau = \frac{1}{4\pi f}\arctan\left(\frac{\sum_{k=1}^{K}\sin(4\pi f t_k)}{\sum_{k=1}^{K}\cos(4\pi f t_k)}\right) \quad (2)$$

and where the power spectrum (P_x) of a signal x (of length K) is computed at a specified frequency, f. The standard deviation of the signal is indicated by σ_l. The indices t_n are the times at which the signal was sampled, and are not periodic.

3) Normalization: The pediatric subjects were analyzed in this paper, with widely varying baseline heart rate characteristics obtained. For example, Subject 7 (seven months) presented with an average heart rate of approximately 140 beats per minute (BPM), whereas Subject 17 (10 years) presented with approximately 90 BPM. This difference is reflected in the HRV features. Subject-adaptive normalization was performed to compensate for the variability of features across subjects. Specifically, the mean and standard deviation of the features were computed from the very start of the recording of every patient, which is guaranteed to represent the nonallergic state. The mean and standard deviation of the background features are used to normalize the remainder of OFC recordings.

4) Statistical Modeling Using GMMs: The OFC does not provide labels as to the current state of allergy throughout the OFC. Therefore, epoch-by-epoch annotation of allergic and nonallergic states cannot be resolved.

Thus, for each subject only the final result or label of the challenge is known. Additionally, the data recorded before the initial checkup, and therefore before administration of the first dose of the problem food, are available and is the only data that can be labeled nonreaction or background. These data (typically 10 min) provide the only temporal annotations for the OFC. With one class of annotated data (background class), a single-class classification platform can be utilized to determine the likelihood that the HRV features of the remainder of the OFC (after the consumption of the allergen) belongs to the background class.

When expert labels for all classes are available, multiclass classification algorithms can be employed. However, when only one class label is available, as is the case with allergy detection, it is only possible to adopt novelty detection procedures. Single-class classifiers and modeling tools have been successfully deployed by the machine learning community. Identification of seizure using one-class support vector machines [21] have been performed by Gardner *et al.* [22]. GMMs applied to novelty detection have been employed for epilepsy, state of anesthesia, and vigilance quantification applications [23]. GMMs are also widely used in the area of speech and audio processing for modeling the conditional distributions of auditory units (phonemes,

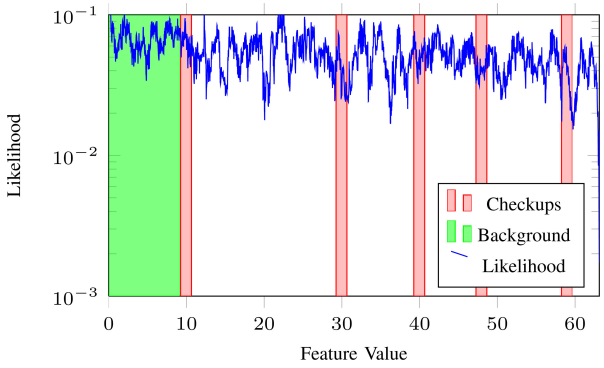

Fig. 3. A sample likelihood series obtained from Subject 23.

phones, etc.) [24]. A comprehensive review of statistical approaches for novelty detection was performed by Markou and Singh [25].

In this paper, allergy detection is performed using likelihoods computed via GMM models built upon the background class. For the GMM, the likelihood function is defined by

$$p\left(\boldsymbol{x}|\, w, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \sum_{i=1}^{M} w_i N\left(\boldsymbol{x}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right) \tag{3}$$

where M is the number of Gaussians (referred to as the GMM order), the weights w_i are nonnegative and sum to 1, and $N(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate Gaussian distribution

$$N\left(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \frac{1}{(2\pi)^{\frac{|\boldsymbol{x}|}{2}}\sqrt{|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right) \tag{4}$$

where μ is the mean vector and Σ is the covariance matrix. The parameters of the GMM (w_i, $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$) are obtained via the expectation–maximization algorithm [26].

Principal component analysis (PCA) is commonly used to preprocess the features before feeding them into the GMM [27]. It is used to reduce the dimensionality of the feature vector. More importantly, the resultant transformed features are decorrelated which allows the application of diagonal covariance matrices in GMM.

5) Postprocessing and Decision Making: A vector of HRV features is obtained from every epoch of ECG data, and these features are then employed by the GMM to obtain the likelihood that the epoch in time represents normal, background data. Lower likelihoods indicate that data are less likely to belong to the background class. Fig. 3 shows a sample likelihood data series for Subject 23. The green region highlights the background activity before administration of an allergen, and the red regions highlighted represent the times that checkups were performed by the allergists. The doses are administered at the end of every checkup after the subjects have been shown to remain stable after the previous dose and interdose observation period.

The criteria for novelty (and therefore allergy classifications) are chosen here to be that the likelihood must fall below a specific threshold (*th*) for a given duration (*d*). The mean (μ) and standard deviation (σ) of the background likelihood data are computed from the testing subject. This step effectively provides automatic adaptation of the generated subject-independent GMM model of background activity for the test subject. The threshold (*th*) is then defined by

$$th = \mu - n\sigma \tag{5}$$

where *n* is a factor which, when multiplied by σ, sets the threshold to a constant value under the mean of the background likelihood data. Large values of n require larger deviations from the background likelihood levels in order to surpass the threshold. The purpose of the duration parameter (d) is to reduce the effect of spurious signals which might fall under a given threshold but are not due to allergy but other cardiac or physical artefacts.

C. Leave-One-Out Performance Assessment

The data of the testing subjects are never available beforehand in the clinic. It is thus necessary to assess the performance of the developed allergy detector in a subject-independent manner. There are various performance assessment routines proposed in the literature [28] such as bootstrapping, split-sample, etc. Their effect on the reported performance for neonatal seizure detection has been compared in a previous study of our group [29]. The split-sample method where one fixed partition of the available data is allocated for training and the rest is used as a testing set has several major disadvantages. Such a division results in a potentially large bias. Overoptimistic or indeed overpessimistic results can be obtained depending on what seems an arbitrary partition of the data, i.e. a "good" or "bad" split.

Therefore, leave-one-out (LOO) is used here to assess the performance of the developed allergy detector. In this procedure, all but one subject are used for training and the remaining subject is used for testing. The process is repeated until every subject was tested, and the average performance is reported from these 24 splits. LOO is known to be an almost unbiased estimation of true generalization error [30], and for databases of the order of the allergy database here, LOO is the best routine to adopt.

D. Model Selection

In each of these 24 splits, nested cross-validation model selection on the training 23 subjects data was performed to choose suitable model parameters. Those include:

1) Percentage of variance retained by PCA for feature set reduction:
   The following set of values was searched: $\{90\%, 95\%, 99\%, 99.9\%, 100\%\}$.
2) The number of Gaussians in the GMM model:
   The following set of values was searched: $\{2, 4, 8, 16, 32, 64\}$.
3) The multiplicative factor (*n*) in decision making:
   The unique, integer rounded values logarithmically distributed between 1 and 200 were searched ($n = 38$).
4) The duration parameter (*d*) in decision making:
   The unique, integer rounded values logarithmically distributed between 1 and 200 were searched ($n = 38$).

In order to select a suitable set of n and d postprocessing parameters, a cost function must be defined which selects parameters based on the performance metrics discussed in the

next section. It has already been stated that it is imperative to not obtain false-positive classifications. With this consideration, parameters which achieve 100% *specificity* in the nested cross-validation are initially considered. From this reduced set, the parameters which lead to the highest *sensitivity* are selected. If there are more than a single set of parameters that satisfy these conditions, the parameters which achieve the maximum *time-gain* are finally chosen.

It can be seen that the model selection routine in our study is completely independent of the performance assessment routine and testing subject data were not seen or used at any time for any supervised tuning of system parameters.

E. Metrics

Sensitivity and *specificity* were computed in order to measure the performance of the automatic allergy detection framework. *Sensitivity* measures the true positive rate, whereas *specificity* measures the true negative rate.

Another metric which provides an insight into the algorithmic performance is *time-gain*. It measures the average difference in time between detections produced by allergists and the algorithm for subjects who were diagnosed allergic in their OFC. In effect, it demonstrates whether it is possible to terminate OFC earlier, and reduce the overall risk of anaphylaxis or other severe reactions. From the *time-gain* factor, the number of subportions which would retrospectively not have been administered can be calculated and this number is expressed as a percentage saving in relation to the total number of subportions administered to the allergic subjects.

III. Results and Discussion

A. Sensitivity and Specificity

The subject-independent allergy detection classification system designed here correctly classified all nonallergic subjects (100% *specificity*). The effect of being diagnosed allergic by clinicians results in that the subject will have to continuously avoid consuming and coming into contact with the food type they were tested against. This avoidance yields a definite and measurable effect on the quality of life of not only the subject but also their immediate family. The clinicians who will use this platform as a diagnostic decision support tool concluded that all our classification systems should minimize these violations on the quality of life. Therefore, obtaining specificity of 100% was deemed most critical feature of the platform. This constraint was enforced both in the parameter selection routine and the subject-independent performance assessment routine. In this context, the significance of the developed fully automated system is indicated by the fact that the classification accuracy of allergy detection can be equivalent (and sometimes superior, see next section) to the current means of allergy diagnosis.

Sensitivity is also a very important metric, and a value of 80% was obtained. This is also a significant result as it indicates that the majority of subjects can be classified accurately even under our imposed constraints.

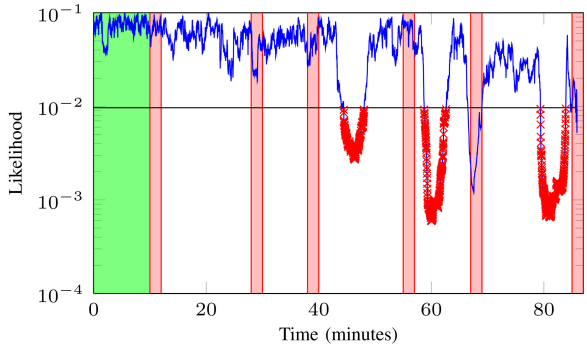

Fig. 4. Demonstration of *time-gain* (Subject 11).

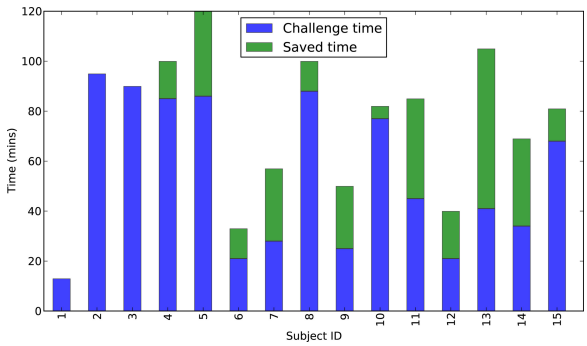

Fig. 5. Illustration of the time saved by the allergy detector.

B. Time Gain

Overall, from all the allergic subjects in the dataset an average *time gain* of 17 ± 19 min (with a median value of 12 min) is obtained, which increases to 22 ± 18 min when considering only the 12 subjects who were correctly classified by the developed detector.

Fig. 4 shows how the developed allergy detection platform identifies allergy before the reaction manifested. In this figure, the blue signal trace is the likelihood that the HRV features belong to the background class. The red rectangles are the times when the subject underwent a checkup by the allergist, the black horizontal line is the fixed threshold which is a function of subject independent n and d parameters, and subject adaptive μ and σ parameters computed on the background data are highlighted in blue. The red markings on the likelihood are the points which satisfied the allergic classification criteria. The test was terminated at the 85th min, and the subject was diagnosed by the allergist as being allergic to the allergen.

It can be seen from this example in that the subject is classified allergic by the system developed here approximately 40 min sooner than the challenge was concluded by the allergist.

Fig. 5 plots a bar chart of the challenge times and the *time gains*. In this figure, the sections of the bars which are filled in green indicate the *time gains* which were obtained for the allergic subjects. This figure shows how allergy can be classified early by the developed routine. The best *time gain* was obtained by Subject 13. Shortly following this subject's initial dose, the HRV features began to display signs of allergic signature. This

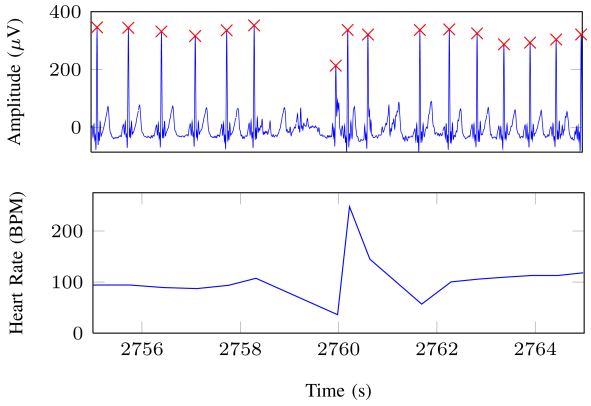

Fig. 6. Example of ectopia with Subject 20.

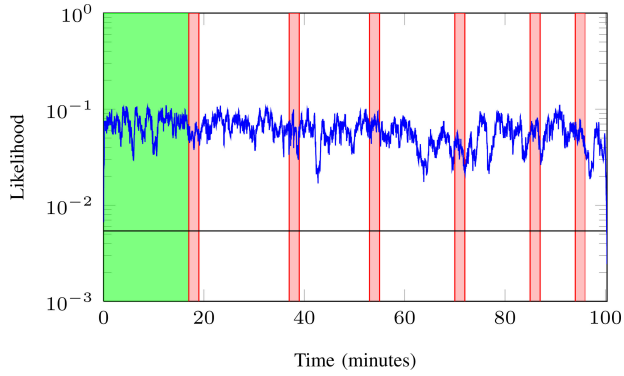

Fig. 7. Example of invariance of the likelihood of Subject 2.

continued for nearly 70 min until the test was finally concluded after 100 min.

In total, 108 subportions of the foods were administered to the subjects, and 41 of these were administered to subjects who were ultimately diagnosed allergic to the food they were tested against. It is worth noting that if the allergy detection framework presented here was used, $12/41$ of the subportions administered to the allergic subjects ($\approx 30\%$) need not have been consumed.

C. Robust Classification

It was observed that the system was robust toward ectopia, as in the case with Subject 20. Fig. 6 shows an example of the ectopic beats for this subject. The top of the figure plots Subject 20's beat-to-beat heart rate over a 10-s time period, and the bottom plot displays the ECG trace with the automatically extracted QRS points. The subject's resting HR can be seen to be ≈ 100 BPM, and rises to ≈ 200 BPM during ectopia between seconds 2745 and 2746. This results in five successive HR calculations of ≈ 107, 36, 247, 144, and 56.

The abnormal variance observed from Subject 20 (who was diagnosed nonallergic during their OFC and by the classification platform) reflects the challenging task of separating allergic and nonallergic subjects with HRV features. While it could be stated that the mean heart rate over an epoch may not significantly be affected by this, it must be noted that other features (for example, standard deviation, and frequency-domain features) will more strongly reflect these abnormalities, and it is principally these which show the robustness of the developed detector. Indeed, by viewing the top trace in Fig. 6, it can be seen that what was shown in the ECG was an example of the most varied RR intervals. At $t = \{2727, 2728, 2751, 2753, 2768, 2769, 2776, 2778\}$ intermediate deviations of over 25 BPM were also observed from the instantaneous heart rate from the mean values. The set of these will certainly contribute to a noticeable change to all of the features and the subsequent principal components computed in this epoch. This robustness is obtained by the combination of subject adaptive normalization and subject-independent parameter selection.

Furthermore, this classification routine has been shown to be robust toward highly disparate HRV baselines. However, for much older subjects (e.g., adults) it is believed that separate models trained on new data would be optimal.

D. Analysis of Errors

Subjects 1, 2, and 3 showed HRV features which did not vary due to their allergic reactions. As a result, allergy in these subjects was impossible to detect through statistical HRV feature analysis.

An example of the invariance of the likelihood is shown in Fig. 7 for Subject 2. Comparing this to the likelihood trace of Fig. 4, it can be seen that the noticeable and sustained departure from background likelihood levels that are present in Fig. 4 are not present in Fig. 7 for Subject 2. It can be seen here that statistical modeling of HRV features is not capable, in some cases, of detecting allergy, and that the likelihoods obtained can be no more varied than background levels. In fact, this and previous works [9] [10] have been the first to demonstrate this clinical discovery in an objective manner.

E. Consistency of Parameter Selection

Due to the LOO parameter selection routine, a different set of classification and postprocessing parameters are obtained for each subject. However, for each of the splits, consistent parameters were selected. For example, the same PCA and GMM parameters were chosen in 20 out of 24 cases. These parameters can be selected from a possible combination of 30 options. However, similar parameters are chosen 22 out of 24 instances, which presents remarkable consistency considering the search space here is of the order of 1500 possibilities (38^2). This consistency suggests that the signatures of allergy consistently identifiable from the HRV, even among different age groups.

IV. Limitations and Future Work

Although LOO is known to be a robust predictor of performance obtainable on future datasets, blind testing on an independent larger dataset with all parameters fixed would undoubtedly provide complementary information about the model behavior and in particular its robustness to system setup permutations. This pilot study results presented here are novel and consistent which justifies future research into this area, which will encompass a wider scale validation study.

This study has focused in obtaining classification which would be appropriate for use in clinical allergy tests and as a result obtaining 100% specificity was the most critical aspect. As a consequence of this restriction, it was not possible to obtain sensitivity of over 80%. This is, however, a specific constraint of our application, and it is possible to obtain higher sensitivities if lower specificities were tolerable.

Subsequent research in this area will target characterization of the HRV features during allergic events. It has not before been possible to separate the normal (i.e., background) HRV features from the allergic features as no temporal annotations were previously available. Since the developed system can provide signal segmentation of background and allergic data, further statistical inference can now be made about the system behavior and relevance of the chosen features.

The use of one-class support vector machines will be investigated to assess benefits of discriminative modeling for allergy detection. The applicability of model-adaptive techniques in contrast to postprocessing parameter adaption will also be investigated.

V. Conclusion

Food allergy reactions provoked by diagnostic OFC can be detected and separated from the nonallergy cases by means of statistical HRV analysis and classification.

It was shown that with the proposed allergy detection system, allergy can be detected on average 17 min sooner than the current state of clinical art. This time saving can be represented as a saving of $\approx 30\%$ of the doses administered to the subjects who were diagnosed allergic, and this introduces the possibility of reducing exposure to the allergen and administering antihistamines early which could reduce (and even eliminate) allergic reactions for some subjects.

The current state of clinical art of allergy diagnosis is the OFC, but by the methods outlined in this paper it can be seen that the state of allergy diagnosis can be advanced with machine-based analysis in a fully automated manner, and these benefits are pertinent for OFCs in Ireland and elsewhere.

Acknowledgment

The authors would like to thank D. Daly and C. Cullinane.

References

[1] H. Sampson, "Update on food allergy," *J. Allergy Clin. Immunol.*, vol. 113, no. 5, pp. 805–819, 2004.

[2] S. Sicherer and H. Sampson, "9. Food allergy.," *J. Allergy Clin. Immunol.*, vol. 117, no. 2, pp. 470–475, 2006.

[3] C. Bindslev-Jensen, B. Ballmer-Weber, U. Bengtsson, C. Blanco, C. Ebner, J. Hourihane, A. Knulst, D. Moneret-Vautrin, K. Nekam, B. Niggemann M. Osterballe, C. Ortolani, J. Ring, C. Schnopp, T. Werfel, European Academy of Allergology, and Clinical Immunology, "Standardization of food challenges in patients with immediate reactions to foods–position paper from the European Academy of Allergology and Clinical Immunology," *Allergy*, vol. 59, no. 7, pp. 690–697, 2004.

[4] N. Twomey, "Digital signal processing and artificial intelligence for the automated classification of food allergy," Ph.D. dissertation, Dept. Elect. Eng., Univ. College Cork, Cork, Ireland, 2013.

[5] M. Primeau, R. Kagan, L. Joseph, H. Lim, C. Dufresne, C. Duffy, D. Prhcal, and A. Clarke, "The psychological burden of peanut allergy as perceived by adults with peanut allergy and the parents of peanut-allergic children," *J. Clin. Exp. Allergy*, vol. 30, no. 8, pp. 1135–1143, 2000.

[6] J. Ménardo, J. Bousquet, M. Rodière, J. Astruc, and F. Michel, "Skin test reactivity in infancy," *J. Allergy Clin. Immunol.*, vol. 75, no. 6, pp. 646–651, 1985.

[7] J. Oude Elberink, J. de Monchy, S. van der Heide, G. Guyatt, and A. Dubois, "Venom immunotherapy improves health-related quality of life in patients allergic to yellow jacket venom," *J. Allergy Clin. Immunol.*, vol. 110, no. 1, pp. 174–182, 2002.

[8] N. Twomey, A. Temko, C. Cullinane, D. Daly, W. P. Marnane, and J. O. Hourihane, "Detection of heart rate variation could improve patient safety and diagnostic yield during oral food challenge," *EAACI*, 2013, to be published.

[9] N. Twomey, S. Faul, D. Daly, J. Hourihane, and W. Marnane, "Classification of biophysical changes during food allergy challenges," in *Proc. 3rd Int. Symp. Appl. Sci. Biomed. Commun. Technol.*, 2010, pp. 1–5.

[10] N. Twomey, A. Temko, J. Hourihane, and W. Marnane, "Allergy detection with statistical modelling of HRV-based non-reaction baseline features," in *Int. Symp. Appl. Sci. Biomed. Commun. Technol.*, 2011, pp. 134–138.

[11] R. Gutiérrez, J. Campos-Garcia, W. P. Marnane, N. Twomey, and A. Temko, "Real-time allergy detection," in *WISP*, pp. 21–26, 2013.

[12] J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, and A. Cheng, "A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 734–740, May 2010.

[13] A. Burgos, A. Goñi, A. Illarramendi, and J. Bermúdez, "Real-time detection of apneas on a PDA," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 4, pp. 995–1002, Jul. 2010.

[14] S. Asgari, M. Bergsneider, and X. Hu, "A robust approach toward recognizing valid arterial-blood-pressure pulses," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 1, pp. 166–172, Jan. 2010.

[15] F. Wilson, F. Johnston, and C. Kossmann, "The substitution of a tetrahedron for the Einthoven triangle," *Amer. Heart J.*, vol. 33, no. 5, pp. 594–603, 1947.

[16] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca, "SHIMMER–A wireless sensor platform for noninvasive biomedical research," *IEEE Sens. J.*, vol. 10, no. 9, pp. 1527–1534, Sep. 2010.

[17] D. Benitez, P. Gaydecki, A. Zaidi, and A. Fitzpatrick, "The use of the Hilbert transform in ECG signal analysis," *Comput. Biol. Med.*, vol. 31, no. 5, pp. 399–406, 2001.

[18] P. Hamilton, "Open source ECG analysis," in *Proc. Comput. Cardiol.*, 2002, pp. 101–104.

[19] R. Kleiger, P. Stein, and J. Bigger Jr., "Heart rate variability: Measurement and clinical utility," *Ann. Noninvas. Electrocardiol.*, vol. 10, no. 1, pp. 88–101, 2005.

[20] N. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophys. Space Sci.*, vol. 39, no. 2, pp. 447–462, 1976.

[21] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001.

[22] A. Gardner, A. Krieger, G. Vachtsevanos, and B. Litt, "One-class novelty detection for seizure analysis from intracranial EEG," *J. Mach. Learn. Res.*, vol. 7, no. 1, pp. 1025–1044, 2006.

[23] S. Roberts, "Extreme value statistics for novelty detection in biomedical data processing," *IEE Proc—Sci., Meas. Technol.*, vol. 147, no. 6, pp. 363–367, Nov. 2000.

[24] D. Reynolds, T. Quatieri, and R. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digit Signal Process*, vol. 10, no. 1, pp. 19–41, 2000.

[25] M. Markou and S. Singh, "Novelty detection: A review—Part 1: Statistical approaches," *Signal Process*, vol. 83, no. 12, pp. 2481–2497, 2003.

[26] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ, USA: Prentice-hall, 1993.

[27] E. Thomas, A. Temko, G. Lightbody, W. Marnane, and G. Boylan, "Gaussian mixture models for classification of neonatal seizures using EEG," *Physiol. Meas.*, vol. 31, no. 7, pp. 1047–1064, 2010.

[28] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell.*, 1995, vol. 2, pp. 1137–1145.

[29] E. Thomas, A. Temko, W. Marnane, G. Boylan, and G. Lightbody, "Discriminative and generative classification techniques applied to automated neonatal seizure detection," *IEEE J. Biomed. Health Inf.*, vol. 17, no. 2, pp. 297–304, Mar. 2013.

[30] V. Vapnik, *Estimation of Dependences Based on Empirical Data*. New York, NY, USA: Springer-Verlag, 2006.