2017 IEEE 56th Annual Conference on Decision and Control (CDC)
December 12-15, 2017, Melbourne, Australia

Ventricular Defibrillation: Classification with G.E.M. and a Roadmap for Future Investigations

Fabio Baronio* Manuela Baronio† Marco C. Campi* Algo Carè* Simone Garatti** Giovanna Perone‡

Abstract— **Developing an ability to classify ventricular fibrillation (VF) into cases where restoration of an organized electrical activity (ROEA) is achieved after the application of a defibrillatory shock, and telling these cases apart from cases where such a restoration does not happen, is of paramount importance to guide first-aid therapy in patients in cardiac arrest. Indeed, VF is a medical emergency of enormous proportions and it is one of the first causes of sudden death in a large range of population's age. In this article, we address this problem in the light of recent achievements in the field of machine learning and present results with the use of a new machine called GEM (Guaranteed Error Machine) applied to a group of patients with out-of-hospital cardiac arrest. While our results indicate that this methodology is promising, it remains a fact that this study is still at the outset, and by this article we also want to make the current state of the art available with the use of GEM to others and indicate what we believe are the research priorities for the near future. This is done in the belief that this important medical endeavor is better addressed by the cooperation of various teams, possibly carrying complementary expertise.**

I. Ventricular fibrillation: the need for a classifier to support therapeutic decisions

Numerous studies in humans indicate that electrocardiographic (ECG) tracings during ventricular fibrillation (VF) carry important information on the cardiac condition and that this can be used during a cardiac arrest for the purpose of predicting the outcome of a defibrillatory shock and to guide a therapy. Various measures like amplitude, power spectrum, or nonlinear statistical indexes have been considered as means to extract information from the ECG tracing.

In more detail, it has been shown that VF-ECG waveforms exhibit a decreasing organization as time goes by after a cardiac arrest. This fact has been described as early as in 1930, [27], and quantitative ECG studies of this phenomenon have become widespread over the last two decades, partly fostered by the availability of modern instruments able to perform real-time signal analysis, see e.g. [3] and the references therein. These studies have provided quantitative confirmation of the common clinical experience that the VF waveform structure fades away with increasing ischemia duration, and these modifications are associated with a decreased likelihood of successful defibrillation and resuscitation.

On a different count, European guidelines indicate that more or less prolonged VF calls for different treatments, [25]. For example, brief (i.e. less than 2-3 minutes) VF is quite reversible with rescue shocks, and rapid defibrillation promotes survival after brief VF arrest, [13]. Instead, rescue shocks applied to VF of more than 4 or 5 minutes' duration rarely result in spontaneous circulation, and often fail to promote a restoration of organized electrical activity (ROEA). In these cases, studies in humans indicate that reperfusion prior to rescue shocks improve defibrillation success and survival, [28]. The quest for effective methods of analysis of VF-ECG waveforms to direct therapy towards immediate defibrillation or reperfusion followed by defibrillation is openly endorsed by the last European guidelines [25] with the following words:

> "If optimal defibrillation waveforms and the optimal timing of shock delivery can be determined in prospective studies, it should be possible to prevent the delivery of unsuccessful high energy shocks and minimise myocardial injury. This technology is under active development and investigation but current sensitivity and specificity is insufficient to enable introduction of VF waveform analysis into clinical practice."

As it has been mentioned, various measures have been used to extract information about the VF duration, and more in general about the cardiac condition, and the advantages of each measure have been reported under different circumstances, [1]. They can be grouped in three classes: amplitude-based, frequency-based, and nonlinear statistical measures, [2], [12], [26], [19], [21], [23], [24], [9]. Compared with traditional human-based clinical assessments, all these measures have the advantage of being quantitative. Amplitude and frequency measures of the ECG tracings have been mainly used in isolation, that is one at a time, to predict the likelihood of success of a defibrillation. Combinations of these measures have been so far employed in few studies

F. Baronio was supported by H2020 MSCA RISE Cardially 691051. The work of M. C. Campi and A. Carè work was partly supported by the H&W program of the University of Brescia under the project "Classificazione della fibrillazione ventricolare a supporto della decisione terapeutica - CLAFITE". A. Carè was supported by an "Alain Bensoussan Fellowship" from the European Research Consortium for Informatics and Mathematics (ERCIM).

*F. Baronio, M.C. Campi and A. Carè are with the Department of Information Engineering, University of Brescia, Via Branze 38, 25123 Brescia, Italy; (emails: {fabio.baronio,marco.campi,algo.care}@unibs.it)

†M. Baronio is with Fondazione Poliambulanza, Brescia, Italy; (email: manuela.baronio@poliambulanza.it)

**S. Garatti is with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy; (email: sgaratti@elet.polimi.it)

‡ G. Perone is with Spedali Civili, Brescia, Italy; (email: joperone@gmail.com)

only, [20], [12], [22], [18], [24]. This paper has a twofold objective. **(a)** We introduce a new approach based on a multi-dimensional pattern classification machine called guaranteed error machine (GEM) based on the work [4]. GEM allows one to combine in a coordinated manner measures of various type and a study presented here conducted on 170 patients with out-of-hospital cardiac arrest treated by the emergency medical services in Brescia, Italy, shows the promising features of this method. Due to page limits, only a limited number of experimental results are presented in the paper, but we invite interested readers to contact us if they want to receive other results. **(b)** On the other hand, this study provides only partial answers and this approach calls for more analyses, testings and extensions that we believe cannot be conducted by a small research group alone and it instead calls for the cooperation of various teams in machine learning worldwide. Hence, through this article we also want to share with others the expertise we have acquired through our research in this field and the second part of this paper is devoted towards pointing out what we see now to be a roadmap for future investigation. We would be happy for anyone interested in this topic to contact us for more exchange.

II. The Guaranteed Error Machine - GEM

A. Fundamentals on classification

To start with, we feel advisable to spend a few words on the problem of classification to give readers who do not have a background in this field the opportunity to go through this manuscript without having to consult specific manuals and textbooks. Given a member of a population, suppose that we can measure a set of its characteristics (also called patterns or explanatory variables) that we want to use to predict, or guess, its nature. In the problem of defibrillation, the member is a patient in cardiac arrest, the set of characteristics comprises various measures extracted from a ECG tracing, and the nature is the capability of the patient's body to positively react to a defibrillatory shock to obtain an ROEA. To set the notation, in the sequel $\mathbf{x}$ is the set of characteristics, normally real or integer values organized in a vector, and y is the nature of the member, which, in binary classification, is a number in the set $\{0,1\}$, where 0 and 1 are two alternative options. To a given $\mathbf{x}$, there can be associated both $y=0$ and $y=1$ as two members of the population carrying the same characteristics can have different nature.[1] Since the value y is not a priori measurable, it is guessed by means of a classifier, that is a function $\hat{y}(\mathbf{x})$, which is constructed from previously observed cases. The goal is of course that the classifier errs as rarely as possible on new instances, in our case patients.

In statistical classification, it is assumed that $\mathbf{x}$ values occur according to a probability. In the defibrillation problem, this corresponds to say that certain amplitude and frequency measures may occur more frequently than others, and their frequency of occurrence is described by a probability. However, the distribution of this probability is not known to the user. The quality of a classifier is measured by the probability of error $PE(\hat{y}) := \mu(\hat{y}(\mathbf{x}) \neq y(\mathbf{x}))$, where μ accounts for the probability of seeing an $\mathbf{x}$ plus the probability with which y distributes over 0 and 1 given $\mathbf{x}$. In the defibrillation application, this is the probability that a patient is met for which the classifier fails to correctly guess whether the defibrillation operation will or will not be successful.

[1] Given $\mathbf{x}$, y is under normal circumstances not totally determined. This is because y may depend on other variables than those contained in $\mathbf{x}$.

B. The GEM classifier

While we refer the interested reader to the article [4] for a comprehensive presentation, we feel advisable to recall here some fundamental features of the Guaranteed Error Machine (GEM) that are relevant to the present paper.

GEM is an algorithm to construct classifiers $\hat{y}(\cdot)$ that has a built-in mechanism to precisely keep control on $PE(\hat{y})$. This is achieved by adopting a ternary output, $\{0,1,\text{unknown}\}$, so that the classifier can abstain from expressing a judgment in doubtful cases. This is not different from the behavior of an expert who, when asked to answer a difficult question, would appeal to the right of saying "I do not know" to keep the chance of error under a given threshold.

A GEM classifier is constructed from a training sequence of N observations, i.e. previously seen cases, $\{(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_N,y_N)\}$, where each observation is assumed to be extracted independently of the others from the population. The complexity of a GEM classifier can be tuned by the user by selecting the value of a parameter $k<N$, and GEM is formed by various regions patched together in the $\mathbf{x}$ domain, where each region has associated an output value $\hat{y}(\mathbf{x})$, either 0 or 1 or "unknown". The classifier is constructed with the attempt to reduce as much as possible the size of the region where "unknown" is issued while maintaining control on the level of error, as indicated by the value k chosen by the user. A large value of k leads to classifiers that more often return a 0 or 1 value, but these classifiers misclassify more frequently, whereas smaller values of k correspond to more risk-averse classifiers paying emphasis on reducing the probability of misclassification, but also returning "unknown" with higher probability. For a detailed presentation of the probabilistic properties of GEM the reader is referred to [4], while here we recall the most relevant result for the present study: independently of μ, GEM constructs classifiers that return on average a wrong prediction, i.e. $\hat{y}(\mathbf{x}) \neq y(\mathbf{x})$, with a guaranteed probability that does not exceed k/N. For example, with $N=1000$ observations and $k=50$, the average probability of error is no more than 95%. Here, "average" refers to the training sequence $\{(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_N,y_N)\}$ and to the new case that one is considering, i.e. $(\mathbf{x},y(\mathbf{x}))$. As shown in [4], the result k/N is close to the real value, and it is exact under known

circumstances, and therefore it gives valuable quantitative indication of reliability[2]; moreover, the fact that the result is independent of μ is of great importance in applications where the value of μ is normally unknown or at best only partly known.

III. Application of GEM to the Classification of Ventricular Fibrillation

In this section we provide results obtained when GEM has been applied to to the problem of classifying patients in a condition of ventricular fibrillation. These studies indicate the potentials of this method for this problem. Thus far, however, only the first steps have been traveled, and in subsequent sections we share with the reader what we believe can/must be the next steps along this exploration in the hope to stimulate collaboration and research activity performed by others.

A. The data in the study

The data are part of an observational prospective study of 170 patients with out-of-hospital cardiac arrest treated by the emergency medical services in Brescia, Italy. The patients were treated according to the European cardiopulmonary resuscitation (CPR) guidelines, [25]. Advanced cardiac life support management, ECG and relevant information were documented using Heartstart 3000 defibrillators (Laerdal Medical, Stavanger, Norway).

ECG post shock tracings were systematically evaluated by two independent medical divisions. We considered the first shock applied to a patient. The shock outcome was labeled depending on the predominant rhythm in the 10 seconds after the shock in three categories: i) persistent VF or ventricular tachycardia, ii) pulseless electrical activity or asystole, iii) supraventricular rhythm with a pulse. The shock was regarded as successful, that is, it corresponded to a ROEA, in case (iii), which occurred in 14 patients. All other 156 cases were grouped together as non-ROEA.

The prediction of defibrillation success is performed in two stages. First, the ECG tracings are characterized through various amplitude and frequency measures. Second, different combinations of these measures are considered and used in the GEM multidimensional classifier proposed in [4]. In the next Subsection III-B, we describe the experimental set-up, followed in Subsection III-C by a presentation of the obtained results.

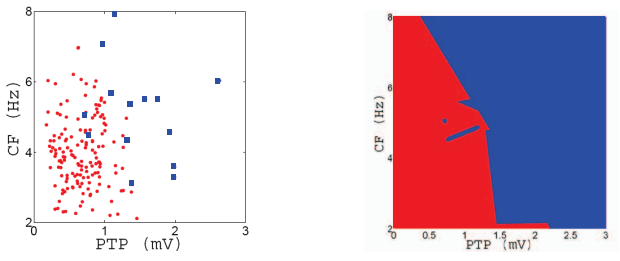
B. Experimental set-up

In line with previous studies, see in particular [2], we have focused our attention to ECG amplitude measures in the time domain and measures in the frequency domain. Precisely, peak-to-peak (PTP), maximum amplitude (Amax), minimum amplitude (Amin), wave amplitude (WA), and root mean square (RMS) are the amplitude measures, see [2], while dominant frequency (DF), centroid frequency (CF), edge frequency (EF), and amplitude spectral area (AMSA) are the frequency measures, see [2], [12], [26], [23], [24]. All these measures are calculated from a 4 second segment of ECG tracing recorded before the first shock was applied. Thus, each patient is characterized with nine measures.

The 170 patients were split into training and validation sets, [11], according to a leave-one-out cross-validation scheme, [14]. This scheme involves using a single observation from the original sample as the validation data, and the remaining observations as the training data ($N = 169$). This is repeated 170 times, such that each observation in the sample is used once as the validation data. Splitting data in training and validation is useful to estimate the ability of GEM to perform correctly on yet unseen patients. An ample investigation of the most suitable measures to be used for classification purposes was conducted. Precisely, patterns $\mathbf{x}$ with two or three measures selected from the set of nine time and frequency measures described above were considered. In this study, parameter k was kept fixed at the value $k = 20$ throughout, so that the theoretical bound on the probability of error was $k/N = 11.8\%$.

C. Results

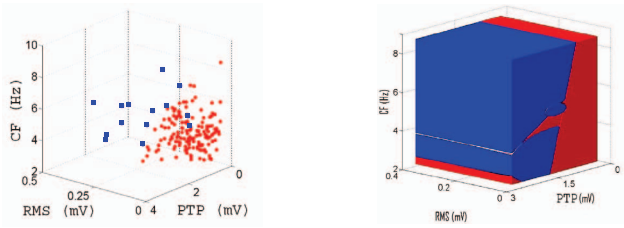
In all experiments of this section we have used the GEM algorithm given in Section 2 of [4]. We present results obtained by using a few combinations of the nine features, those that gave the most promising results. Figure 1a shows as an example $\mathbf{x}$ = (PTP, CF) measures in a bi-dimensional view for patients undergoing successful ($y = 1$, blue squares) and unsuccessful ($y = 0$, red circles) defibrillation attempts. PTP and CF have little correlation, as reported in [18], so that these measures are expected to carry complementary information. Figure 1b shows the classification map obtained using GEM for these measures. Table I reports the performances, obtained through the cross-validation scheme described above, of various bi-dimensional GEM classifiers. In the table, *Errors %* is simply the ratio between the number of errors in cross-validation over the total number of cases, i.e. 170, and *Doubts %* is the ratio between the number of abstentions over the total number of cases.



(a) (b)

Fig. 1: (a): 2D data. (b): Classifier constructed from 2D-data. All data in the training set are correctly classified since GEM does not allow misclassification of seen cases, [4], see also the discussion in Section IV-B.

[2]The results in [4] stemmed from the theory of the scenario approach, see e.g. [15], [6], [7] for some more recent results in the scenario approach.



(a) (b)

Fig. 2: (a): 3D data. (b): Classifier constructed from 3D-data.

	Errors %	*Doubts %*
AMSA-PTP	11.8	0
AMSA-WA	10.7	3
AMSA-CF	11.8	0
AMSA-DF	10.1	0
AMSA-Amax	11.2	0
PTP- CF	13.6	3
PTP-EF	7.7	8.9
RMS-CF	13	3
PTP- DF	10.1	3.6

TABLE I: Classification with 2 features. "Errors"=total number of errors; "Doubts" = abstentions.

Further, Figure 2a shows as an example in a three-dimensional view $\mathbf{x}$ = (RMS, PTP, CF), and the classifier obtained with these measures by GEM is displayed in Figure 2b. This figure gives the external view of the three-dimensional map; the nontrivial internal view can be obtained thorough different bi-dimensional cross-sections of the map. These views can be of interest for medical doctors who co-operate in the study. As is clear, however, the use of visual information loses effect as one moves up to classifiers based on an increasing number of features, so that one has rather to rely on quantitative evaluations. Table II reports the performances of a collection of three-dimensional GEM classifiers. The bi-dimensional GEM technique shows that combinations

	Errors %	*Doubts %*
RMS-PTP-CF	10.7	0
RMS-Amin-CF	7.7	3
WA-Amin-EF	11.8	4.1
PTP-AMSA-DF	8.9	0
RMS-WA-AMSA	7.1	1.8
RMS-AMSA-CF	10.1	1.2
WA-AMSA-DF	9.5	1.8
WA-Amin-DF	10.2	0.6

TABLE II: Classification with 3 features.

of different amplitude and frequency measures give different predictive performances. The three-dimensional GEM analysis indicates that the combination of three-different measures may lead to better predictions than with the bi-dimensional GEM analysis in terms of errors + doubts. This suggests a positive answer to the vexed question "whether combining multiple ECG features can improve the capability of defibrillation outcome prediction in comparison to single feature analysis", [16], [17]. A more decisive argument to settle the question could come from the potential that is offered by GEM for rigorous and automatic feature selection, as discussed in the next Section IV-A. The observed errors are in line with the theoretical counterpart which, as already seen, is $k/N = 11.8\%$. On the other hand, an aspect that is important in the VF application is that not all errors have the same meaning. Precisely, judging that an individual's body is ready to positively react to a defibrillation shock while it is not results in a loss of time (that used to give the ineffective shock) plus possibly in a further damage to the cardiac muscle due to the electrical shock, an effect that is however controversial and no significant literature is available at present on this aspect. On the other hand, making the opposite mistake of judging that the body is not ready to receive the shock while a shock would instead result in an ROEA is more severe. In the literature when referring to these two types of errors one speaks of specificity and sensitivity. Precisely, specificity is given by formula (number of negative (no ROEA) that have been classified as negative)/(total number of negative) while sensitivity is given by (number of positive (ROEA) that have been classified as positive)/(total number of positive). In our study we achieved a sensitivity that went up to 64.3% with *RMS-PTP-CF* but this is not sufficient. Achieving higher level of sensitivity than 64.3% is hard with a total theoretical level of error of 11.8% and, to increase sensitivity, a compromise is needed in the specificity value which will have to be decreased so that the total probability of error will be higher than 11.8%. How this process can be governed is under investigation and this aspect is further discussed in Section IV-B, while we here notice that the mechanisms that guarantee error levels present in GEM represent a starting point of interest in this study.

IV. Future Directions: A Roadmap of Investigation

The results presented in the previous section points to illustrating that GEM offers an opportunity to coordinate various measures so as to obtain new classes of classifiers that come with precise guarantees of performance. On the other hand, the research in this field is still at the outset, and in this section we present what we believe are the directions of investigation for the future.

A. GEM with complete classification

As we have seen, the GEM classifier gives an output which can be 0, 1 or "unknown", where the latter means that the classifier abstains from expressing a judgement. While this opportunity is of interest in many applications - and this is why "unknown" was included in the outputs of GEM which was conceived as a general-purpose method - its value in the VF classification problem is doubtful. Indeed, expressing a judgement seems a mandatory requirement in this application because a VF is a need-to-act condition where lacking to act means sure death of the patient. An interesting theoretical extension over the results the classical GEM are based upon that can help overcome this difficulty came out recently in the paper [5]. In this

paper it is theorized that one can perform optimization with many variables (GEM is constructed through optimization and the equivalent of the number of variables in the GEM application is the parameter k, see [4]) and wait until optimization is completed to evaluate the complexity of the solution. It is a fact shown in [5] that a-posteriori finding that the complexity of the solution is k (that is the same maximum complexity as when the variables are k) does not lead exactly to the same guarantees of performance as when k is the number of optimization variables. However, the difference between the two situations is quite small. This theoretical achievement, for which the reader is directed to [5], sets a new opportunity within the context of constructing classifiers along the lines posed in [4]: instead of fixing in advance k, one continues to patch the $\mathbf{x}$ domain with new regions until the whole $\mathbf{x}$ domain is completely covered. This means that the classifier will return 0 or 1 in all possible conditions. Upon termination, one evaluates the complexity of the classifier and from that draws a conclusion for its ability of correctly predicting new cases by applying the results from [5]. In this set-up, the result k/N is no longer valid and is substituted by an a-posteriori evaluation. While this appear to be a very reasonable approach, setting the details may call for the coordinated effort of various groups: (i) what is the best way to construct regions in the $\mathbf{x}$ domain? while one approach is proposed in [4], this approach is in no way optimal, and other constructions can be envisaged possibly driven by experimental studies; (ii) presently, the extension of the regions in the $\mathbf{x}$ domain are optimized through the consideration of quantities only indirectly related to the extension of the region. Is it possible to relate directly optimization to the extension of the region so that upon termination with the approach described in this section the complexity of the solution is kept small? Moreover, is the volume a suitable reference measure, or should one rather consider coverage of empirical points as a reference since this somehow reflects the density of the population? Answering these questions calls for extensions of the presently available algorithm.

One further aspect which is worth investigating along the road traced in this section is the possibility of including many features (i.e. ECG measures) in the classifier and let the classifier decide which features, or combination of them, are better used. This possibility is provided by the fact that including many features together in traditional approaches leads immediately to too many degrees of freedom, so that one loses a grasp on the generalization ability of the method. On the other hand, the new philosophy of waiting and judging introduced in [5] offers a new opportunity, that of verifying the complexity a-posteriori. Hence, one can use many features, and yet the classifier can turn out to be simple by perhaps exploiting combinations of the variables without introducing excessive intricacy and complication in the classifier. This study is believed to be of great importance towards the automatic selection not only of the classifier given the features, which is the main goal of traditional learning algorithms, but also the automatic selection of the relevant combinations of features to obtain a quality classifier.

B. Unbalanced Probability of Error

As previously mentioned, one aspect that is relevant to the VF application is that not all errors have the same importance. Hence, one important extension of the method would be that of introducing a way to systematically keep control on the two types of error, those related to specificity and sensitivity.

We here envisage two lines of research: (i) when GEM constructs regions to be patched together in the $\mathbf{x}$ domain, more flexibility can be allowed when constructing regions corresponding to a potential error which carries less severe consequences, and viceversa when constructing regions associated to more severe mistakes. By looking at the construction of a GEM classifier in [4], one can see that various levels of flexibility are introduced there for the purpose of meeting exactly complexity k at the end of the process. A similar scheme with various levels of flexibility can possibly be introduced in the present context so as to unbalance the probability of mistake of type 1 and type 2; (ii) GEM does not allow for misclassified points in the training set. This means that, by construction, all the observations that are used to build the classifier are correctly classified. This seems to be rigid as compared to other machine learning methods, where misclassification is allowed in relation to outlier points (see, for example, the concept of "soft margin" in Support Vector Machines, [10]). In GEM, one might introduce selective schemes to allow for a larger misclassification in relation to constructions leading to a type of error that is less important. At a theoretical level this set-up is at present completely unexplored. Finally, understanding how sensitivity and specificity can be kept under control would provide new solutions for building balanced classifiers from datasets where different categories are not equally represented, a task that nowadays is often carried out by resorting to under-sampling and over-sampling techniques, [8].

C. Discriminating Uncertain Cases

GEM constructs a classifier by expanding regions in the $\mathbf{x}$ domain as much as possible until a further expansion leads to a misclassification of some of the observations, refer to [4] for details. As a consequence, the boundary of the regions touch observed points and no margin is kept so that an arbitrarily small shift of the regions would lead to misclassification of some observations. One might conceive methods to introduce a margin by which generalization properties might be boosted. One approach which is worth investigating is the following. GEM contains some arbitrariness in its initialization: the first region that is being constructed is centered in an observation which can be selected by the user. Say that various initializations are attempted, leading to different GEM classifiers. It is expected that these classifiers will agree over large portions of the $\mathbf{x}$ domain, but they

may disagree over some restricted parts. Intuitively, these parts are less guaranteed to lead to correct classification, and therefore one might conceive of being more cautious in the use of GEM on these parts, e.g. by leaving some margin so that 0 is attributed to $\mathbf{x}$ values near 0 observations and viceversa for 1 values. While this approach calls for extra investigation, the fact that the parts of agreement are more guaranteed can be easily proven at a theoretical level, and we here sketch the argument to show this. For simplicity, refer to the case of two GEM classifiers, say GEM-1 and GEM-2, that have the same probability of mistake, call it PE, and let $P_d = Prob(\text{GEM-1 disagree with GEM-2})$. For the sake of the argument, consider a scheme where a classifier $\overline{\text{GEM}}$ says 0 where GEM-1 and GEM-2 agree on 0 and $\overline{\text{GEM}}$ says 1 where GEM-1 and GEM-2 agree on 1, while a coin is flipped to make a decision if GEM-1 and GEM-2 disagree. Notice that $\overline{\text{GEM}}$ can also be redefined as one where a coin is flipped in all situations, as flipping a coin in case of agreement is immaterial, so that the probability of error of $\overline{\text{GEM}}$ is still PE. Clearly $Prob(\text{mistake}|\text{disagreement occurs}) = 0.5$. Now let $PE|A$ be the conditional probability of error given that GEM-1 and GEM-2 are in agreement, that is, $PE|A = Prob(\text{mistake}|\text{agreement occurs})$. We have, $PE = P_d \cdot 0.5 + (1-P_d) \cdot PE|A$, from which $PE|A = (PE - P_d \cdot 0.5)/(1-P_d)$. This latter quantity is always less than PE when $PE < 0.5$ (that is, the classifiers do better than a blind prediction based on coin tossing), showing that over the regions where an agreement occurs the probability of error is smaller. While this property is intriguing, at present it is not clear how regions of disagreement should be dealt with so that the total probability of error is decreased.

V. Conclusions

In this paper we have introduced a new classification scheme for the problem of judging whether or not conditions of VF are ready to generate an ROEA after an electrical shock is applied. The main feature of the proposed scheme is that it can accommodate the systematic use of many ECG features. Moreover, the method comes accompanied by precise theoretical guarantees. The results of a study conducted on 170 patients with out-of-hospital cardiac arrest have shown the potentials of the method. On the other hand, many problems remain open and the last part of the paper has traced the directions that we consider more important for future investigations.

References

[1] A. Amann, K. Rheinberger, and U. Achleitner. Algorithms to analyze ventricular fibrillation signals. *Current Opinion in Critical Care*, 7(3):152–156, 2001.

[2] C. G. Brown and R. Dzwonczyk. Signal analysis of the human electrocardiogram during ventricular fibrillation: frequency and amplitude parameters as predictors of successful countershock. *Annals of Emergency Medicine*, 27(2):184–188, 1996.

[3] C. W. Callaway and J. J. Menegazzi. Waveform analysis of ventricular fibrillation to predict defibrillation. *Current Opinion in Critical Care*, 11(3):192–199, 2005.

[4] M. C. Campi. Classification with guaranteed probability of error. *Machine Learning*, 80(1):63–84, 2010.

[5] M. C. Campi and S. Garatti. Wait-and-judge scenario optimization. *Mathematical Programming*, doi:10.1007/s10107-016-1056-9, 2016.

[6] M.C. Campi and A. Carè. Random convex programs with L1-regularization: sparsity and generalization. *SIAM Journal on Control and Optimization*, 51(5):3532–3557, 2013.

[7] A. Carè, S. Garatti, and M.C. Campi. Scenario min-max optimization and the risk of empirical costs. *SIAM Journal on Optimization*, 25(4):2061–2080, 2015.

[8] N. V. Chawla, K. W. Bowyer, L.O. Hall, and W. P. Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[9] B. Chicote, U. Irusta, R. Alcaraz, J. J. Rieta, E. Aramendi, I. Isasi, et al. Application of entropy-based features to predict defibrillation outcome in cardiac arrest. *Entropy*, 18(9):313, 2016.

[10] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[11] T. Eftestøl, H. Losert, J. Kramer-Johansen, L. Wik, F. Sterz, and P. A. Steen. Independent evaluation of a defibrillation outcome predictor for out-of-hospital cardiac arrested patients. *Resuscitation*, 67(1):55–61, 2005.

[12] T. Eftestøl, K. Sunde, S. O. Aase, J. H. Husøy, and P. A. Steen. Predicting outcome of defibrillation by spectral characterization and nonparametric classification of ventricular fibrillation in patients with out-of-hospital cardiac arrest. *Circulation*, 102(13):1523–1529, 2000.

[13] M. S. Eisenberg, M. K. Copass, A. P. Hallstrom, B. Blake, L. Bergner, F. A. Short, and L. A. Cobb. Treatment of out-of-hospital cardiac arrests with rapid defibrillation by emergency medical technicians. *New England Journal of Medicine*, 302(25):1379–1383, 1980.

[14] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics. Springer, Berlin, 2001.

[15] S. Garatti and M.C. Campi. Modulating robustness in control design: Principles and algorithms. *Control Systems, IEEE*, 33(2):36–51, 2013.

[16] M. He, B. Chen, Y. Gong, K. Wang, and Y. Li. Prediction of defibrillation outcome by ventricular fibrillation waveform analysis: a clinical review. *Journal of Clinical & Experimental Cardiology*, S10, 2013.

[17] M. He, Y. Gong, Y. Li, T. Mauri, F. Fumagalli, M. Bozzola, et al. Combining multiple ECG features does not improve prediction of defibrillation outcome compared to single features in a large population of out-of-hospital cardiac arrests. *Critical Care*, 19(1):425, 2015.

[18] I. Jekova, F. Mougeolle, and A. Valance. Defibrillation shock success estimation by a set of six parameters derived from the electrocardiogram. *Physiological Measurement*, 25(5):1179, 2004.

[19] J. J. Menegazzi, C. W. Callaway, L. D. Sherman, D. P. Hostler, H. E. Wang, K. C. Fertig, and E. S. Logue. Ventricular fibrillation scaling exponent can guide timing of defibrillation and other therapies. *Circulation*, 109(7):926–931, 2004.

[20] K. G. Monsieurs, H. De Cauwer, F. L. Wuyts, and L. L. Bossaert. A rule for early outcome classification of out-of-hospital cardiac arrest patients presenting with ventricular fibrillation. *Resuscitation*, 36(1):37–44, 1998.

[21] A. Neurauter, T. Eftestøl, J. Kramer-Johansen, B. S. Abella, V. Wenzel, K. H. Lindner, et al. Improving countershock success prediction during cardiopulmonary resuscitation using ventricular fibrillation features from higher ECG frequency bands. *Resuscitation*, 79(3):453–459, 2008.

[22] M. Podbregar, M. Kovačič, A. Podbregar-Marš, and M. Brezocnik. Predicting defibrillation success by genetic programming in patients with out-of-hospital cardiac arrest. *Resuscitation*, 57(2):153–159, 2003.

[23] G. Ristagno, Y. Li, F. Fumagalli, A. Finzi, and W. Quan. Amplitude spectrum area to guide resuscitation. A retrospective analysis during out-of-hospital cardiopulmonary resuscitation in 609 patients with ventricular fibrillation cardiac arrest. *Resuscitation*, 84(12):1697–1703, 2013.

[24] G. Ristagno, T. Mauri, G. Cesana, L. Yongqin, A. Finzi, F. Fumagalli, et al. Amplitude spectrum area to guide defibrillation: A conclusive validation in 1,617 ventricular fibrillation patients. *Circulation*, 130(Suppl 2):A311–A311, 2014.

[25] J. Soar, J. P. Nolan, B. W. Böttiger, G. D. Perkins, C. Lott, P. Carli, et al. European resuscitation council guidelines for resuscitation 2015, Section 3. Adult advanced life support. *Resuscitation*, 95:100–147, 2015.

[26] H. U. Strohmenger, T. Eftestol, K. Sunde, V. Wenzel, M. Mair, H. Ulmer, et al. The predictive value of ventricular fibrillation electrocardiogram signal frequency and amplitude variables in patients with out-of-hospital cardiac arrest. *Anesthesia & Analgesia*, 93(6):1428–1433, 2001.

[27] C. J. Wiggers. Studies of ventricular fibrillation caused by electric shock: II. Cinematographic and electrocardiographic observations of the natural process in the dog's heart. Its inhibition by potassium and the revival of coordinated beats by calcium. *American Heart Journal*, 5(3):351–365, 1930.

[28] L. Wik, T. B. Hansen, F. Fylling, T. Steen, P. Vaagenes, B. H. Auestad, and P. A. Steen. Delaying defibrillation to give basic cardiopulmonary resuscitation to patients with out-of-hospital ventricular fibrillation: a randomized trial. *JAMA*, 289(11):1389–1395, 2003.