Discriminating between V and N Beats from ECGs Introducing an Integrated Reduced Representation along with a Neural Network Classifier

Vaclav Chudacek[1], George Georgoulas[2], Michal Huptych[1], Chrysostomos Stylios[3], and Lenka Lhotska[1]

[1] Dept. of Cybernetics - Czech Technical University in Prague, Prague, Czech Republic
{chudacv,huptycm,lhotska}@fel.cvut.cz
[2] Dept of Computer Applications in Finance and Management, TEI of Ionian Islands, Lefkas, Greece
georgoul@teiion.gr
[3] Dept. of Informatics and Telecommunications Technology TEI of Epirus, Arta, Greece
stylios@teiep.gr

Abstract. The main objective of this paper is to investigate and propose a new approach to distinguish between two classes of beats from the ECG holter recordings - the premature ventricular beats (V) and the normal ones (N). The integrated methodology consists of a specific sequence: R-peak detection, feature extraction, Principal Component Analysis dimensionality reduction and classification with a neural classifier. ECG beats of holter recordings are described using means as simple as possible resulting in a description of the QRS complex by features derived mathematically from the signal using only R-peak detection. For this research work, normal (N) and ventricular (V) beats from the well known MIT-BIH database were used to test the proposed methodology. The results are promising paving the way for the more demanding multiclass classification problem.

Keywords: Holter monitoring, PCA, Neural Networks, MLP classification.

1 Introduction

Many different methods have been proposed to solve the crucial problem of long-term holter recordings evaluation, which could be transformed into the classification problem of discriminating between normal 'N' and a variety of other beats, mainly premature ventricular 'V' beats and supraventricular beats (S).

A lot of research effort has been put to investigate and propose methods to examine and classify the holter recordings based on beat-shape description

C. Alippi et al. (Eds.): ICANN 2009, Part II, LNCS 5769, pp. 485–494, 2009.

parameters [1], shape descriptive parameters transformed with the Karhunen-Loeve method [2], and hermite polynomials [3]. Some other research proposals use time-frequency features [4] and features obtained from heartbeat interval measurements [5,6] in order to identify cardiac arrhythmia.

For any classification problem, in order to compare different approaches, the setup of the experiments, where the type of the training and the selection of testing sets are defined, is of major importance. In most problems, there are two main setups to be considered: training based on a local learning set [7] and on a global learning set [8].

In holter monitoring, the main reason for using local learning is the fact that beats within one patient tend to look alike - but tend to differ widely among different patients - therefore using locally trained classifiers usually leads to better overall results in a personalized medical approach where the patient himself is its own control. In the case of global training the records used for training the classifier are distinct from those used for testing - warranting therefore better, if the data are representative, generality of the classifier.

In this paper we investigate a new configuration to deal solely with distinguishing between normal and ventricular beats [9] where global training fashion is considered; for our testing we use a reduced representation of the original beats coming from the MIT-BIH database [10] and a neural network classifier. The results of the proposed approach are then compared with other published methods using the metrics of sensitivity and specificity.

The rest of the paper is organized as follows: Section 2 presents the proposed methodology to handle the data and extract the specific set of features. Section 3 briefly describes the classification methods involved. In Section 4 the selection of the training and testing sets is described and the results are presented. Section 5 concludes the paper giving some directions for future research.

2 Handling of Data and Feature Extraction

2.1 Preprocessing

For this research work in order to prepare the Holter records, all data records were re-sampled to 500Hz from the original 360Hz. No filtering was performed on any of the signals.

The detection and localization of the R-peak is of paramount importance in the subsequent analysis. For the detection of R-peaks of our data set, the method proposed firstly by Christov [11] was applied. The feature set that is used here is based solely on the R-peak findings - we do not use any other measurement of beat's characteristic points. More specifically the maximum of the major R-peak is found and a window of 128 samples with R peak centered on position 64 is selected for further computations. Fig. 1 presents the result of the applied method for the two different classes under investigation. As it can be seen the "mean" N and V beats have quite a distinct morphology (even though some beats can deviate quite a lot from these shapes, constituting what is widely known as "outliers"). The extracted features are described in the next section.

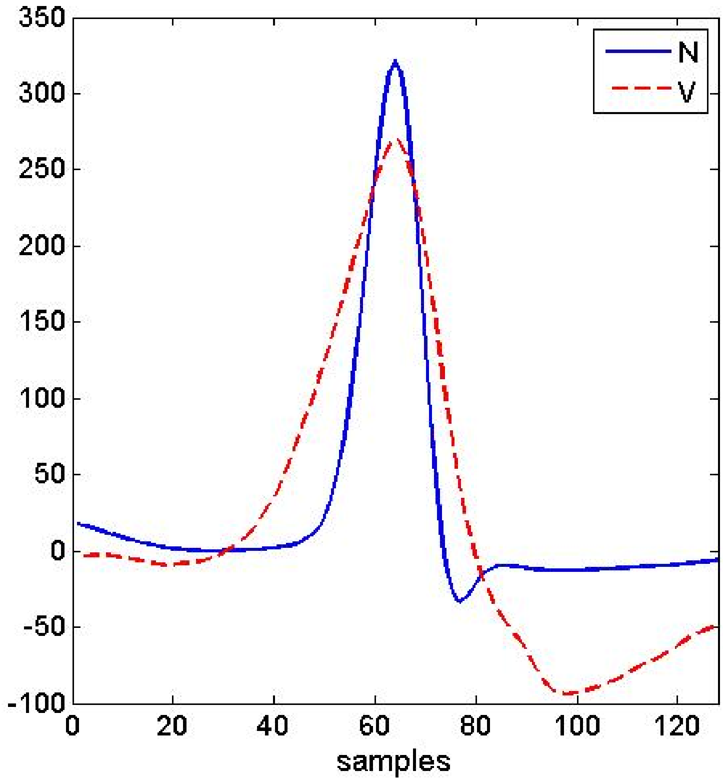

Fig. 1. The "mean" N and V beat waveforms as they were calculated using the MIT database

2.2 Feature Extraction and Selection

Usually feature sets for beat characterization use time intervals, amplitudes and their ratios based on the important points measured from the signal [9].

Here, we propose the use of a feature set consisting of features that are computed solely on the 128 samples around the R-peak. We propose to use features that were selected after visually inspecting the waveforms coming from different classes. These features as proven by the classification results can quantify the difference between the 2 classes.

More specifically the feature set consists of nine measures. Three of them are directly calculated from the truncated signal; namely, the minimum value of the second half of the signal (i.e. from sample 64 till the end of the signal) along with the location of the minimum, and the standard deviation (using all 128 samples).

The rest 6 features are calculated by processing the binary sequences that are constructed after thresholding the first order difference of the original signal and

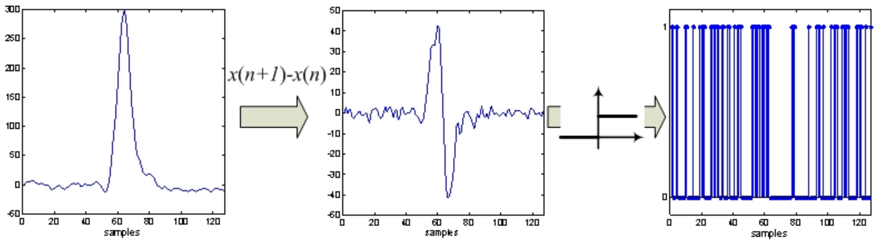

Fig. 2. Transformation of the original signal into a binary sequence

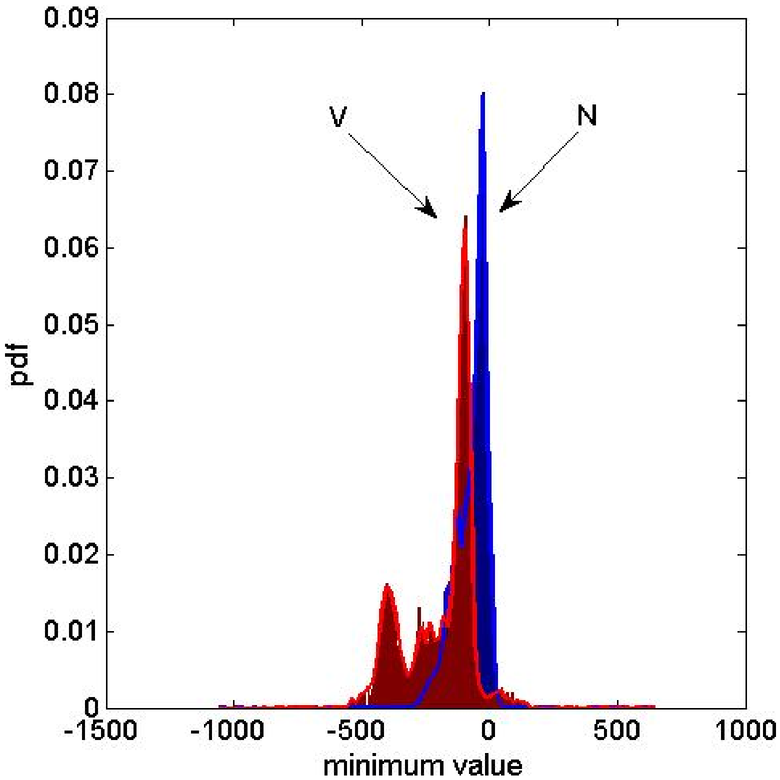

Fig. 3. Experimental "pdfs" of the minimum value for the N (blue) and V (red) class

the second order difference of the original signal. Fig. 2 depicts the transformation of the original signal into a binary sequence as described above.

For each one of the two aforementioned sequences the Shannon entropy is calculated (eq. 1). The two binary sequences are then combined creating a 4-level sequence (a two digit binary word can be described by one digit of an "alphabet" with base 4). The Shannon entropy of this sequence constitutes the third feature and the probabilities (ratios of occurrences) of the three out of the four levels (including the fourth would be redundant) completes the feature set.

$$Entropy = -\sum_{n=1}^{K} p_n \ln(p_n) \qquad (1)$$

Fig. 3 depicts the “pdfs” (histograms with a superimposed spline curve for illustration purposes) of the distribution of the minimum value of the waveform for the two classes. As it can be seen this feature captures the variation between the 2 classes. The rest of the features (not shown here) also have distributions that show the potential benefit of being used for this particular discrimination task.

The feature set is mainly based on the ability to find correctly the maxima of the major R-peak. Then all the features are computed based on the “truncated” signal itself without the need of any other measurements. Therefore this feature set could be a very useful model for classifying the data obtained by telemedicine application devices. In the proposed approach, we have no information from the depolarization phase of the beat cycle - since the behavior of the T-wave varies wildly in terms of shape and length and therefore it would be necessary to employ additional measurement of the end of T-wave.

It is also apparent that some of those features might be correlated. But, it is well known that when we use neural networks classifiers, it is beneficial to feed them with uncorrelated features and also to get rid of redundant information. Thus, in the proposed methodology a dimensionality reduction stage was included before the neural network stage.

2.3 Dimensionality Reduction

It is well known that in pattern recognition tasks, usually potential improvement (better generalization) can be achieved by using fewer features than those available [12]. Actually, literature proposes during the development of a classifier to extract several features, which may convey redundant information about the pattern-class of interest. Therefore, in the proposed approach we included a Principal Component Analysis (PCA) stage so that to un-correlate the originally extracted features using a linear transformation [12].

PCA, or Karhunen-Loeve transformation, is an approach to perform dimensionality reduction by linear combination of the original features in such a way that preserves as much of the relevant information as possible [12,13]. This method computes eigenvalues of the correlation matrix of the input data vector and then projects the data orthogonally onto the subspace spanned by the eigenvectors (principal components) corresponding to the dominant eigenvalues. Even if the whole set of the eigenvectors is retained, this may also lead to an improvement of the classification performance, because the new set has features that are uncorrelated and this, in general, improves the classification capabilities of a classifier.

3 Neural Network Classification

Artificial Neural Networks (ANNs) are increasingly and successfully used in classification problems. They are structures composed of many simple processing elements, that operate in parallel and whose main function is determined by the network's structure, the strength of their connection and the processing carried out by the processing elements (artificial neurons). They are capable of finding commonalities in a set of seemingly unrelated data and for this reason are used in a growing number of classification tasks.

Among the numerous ANN paradigms encountered in the literature [12], the Multi-layer Perceptron (MLP) is the most widely used in the field of pattern recognition [12,13,14]. Training of an MLP is often formulated as the minimization of an error function, such as the total mean square error between the actual output and the desired output summed over all available data. While the sum-of-squares error function is appropriate for regression, for classification problems it is often advantageous [14,15] to optimize the network using the cross entropy error function (eq. 2), i.e. optimizing the network to represent the posterior probabilities of each class [12,13].

$$E = -\sum_{n=1}^{N} \sum_{k=1}^{c} \{t_k^n \ln y_k^n + (1 - t_k^n) \ln(1 - y_k^n)\} \tag{2}$$

where N is the number of training samples and c the number of classes, $t_k^n \in \{0, 1\}$ is a binary class label, (k=1,...,c) of the n^{th} data sample and y_k^n is the actual output of the k^{th} neuron of the ANN, when the n^{th} data sample is presented at its input.

For this case, we use the logistic activation function for the hidden layer units and the softmax (eq. 3) activation function for the output nodes [13,14].

$$y_j = \frac{\exp(a_j)}{\sum_i \exp(a_i)} \tag{3}$$

where a_i is the intermediate linear output of an artificial neuron.

The above configuration has proven to be more appropriate for classification purposes with many successful implementations [13,14]. Therefore, in this research work the above formulation has been adopted.

4 Experimental Results

For evaluation of the proposed approach, we used the commonly used MIT-BIH database [10]. There are two ways of training the classifiers with this database.

The first one is to use local training - using vertical division of the database. That means that, usually, the beginnings of each of the recordings from the database are used for training and remaining parts of each of the recordings are used as a testing set. Although this type of training brings usually results close to absolute sensitivity and specificity as it is often encountered in the literature e.g.

[1,7], it is very controversial from the point of view that any practical application would require additional annotation of at least a short part of each patient's recordings. On the other hand global training implies that the records used for training the classifier are distinct from those used for testing. This means that no additional annotation is needed and the classifier can be directly used on any new patient.

After considering the above mentioned advantages and disadvantages the results reported in the next section are based on the global classification approach using 44 of the MIT recordings and employing the leave one out technique. In other words each time 43 recordings were used for training the classifier and 1 for testing.

Since one of the classes is heavily underrepresented in the given dataset (this is not a flaw of the data, it is "just the way things are"), this makes training of the MLP problematic. This means that we are running the risk to build a classifier heavily biased to classify everything as N class. Different approaches have been proposed in order to alleviate this problem. In our case we downsampled the N class (only during the creation of the train set) taking one every 14 samples. By doing so we have "pushed" the MLP to better learn the V class since in the problem at hand having a high sensitivity is a bit more significant than having a very high specificity (a very high specificity is achieved in almost all similar studies as reported in the conclusion section).

As mentioned in Section 2, after the feature extraction stage, we have included a dimensionality reduction stage based on PCA. In PCA, selecting the number of the retained Principal Components constitutes another design parameter and more than one "criteria" can be found in order to guide the selection process [15]. However, usually the selection is based on a trial and error approach. In our case through an initial experimentation phase using a simple classifier we found out that five to seven Principal Components yield similar results. As a result we selected to retain six of them. Among the different configurations of the MLP (10, 15, 20 and 25 neurons in the hidden layer were tested using a small subset of the dataset in a few preliminary runs without however a thorough search into the parameter search) the one with 20 neurons yields slightly better results.

The overall procedure is depicted in Fig. 4. In total the classifier (6-20-2) managed to classify correctly 58651 out of the 67264 N beats and 5535 out of the 5997 V beats resulting in sensitivity equal to 92.30% and specificity equal to 87.20%. The results are summarized in Table 1 for each one of the 44 recordings.

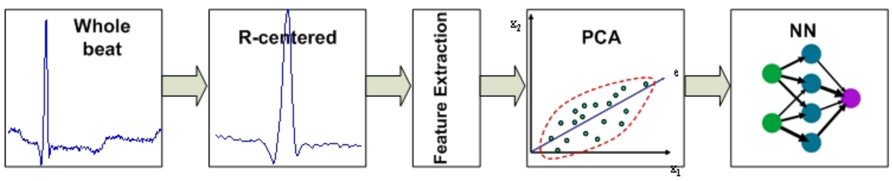

Fig. 4. Overall procedure

Table 1. Classification results for all 44 recordings

Record Number	# N beats	# V beats	Correctly classified N beats	Correctly classified V beats
100	2234	1	2234 (100%)	1 (100%)
101	1855	0	1841 (99.25%)	0 (-)
103	2077	0	2070 (99.66%)	0 (-)
105	2521	41	2190 (86.87%)	38 (92.68%)
106	1504	518	1501 (99.80%)	441 (85.14%)
108	1735	14	860 (49.57%)	6 (42.86%)
109	0	38	0 (-)	37 (97.37%)
111	0	1	0 (-)	0 (0%)
112	2532	0	2388 (94.31%)	0 (-)
113	1784	0	1781 (99.83%)	0 (-)
114	1815	43	1644 (90.58%)	39 (90.70%)
115	1948	0	1935 (99.33%)	0 (-)
116	763	41	148 (19.40%)	41 (100%)
117	1529	0	1224 (80.05%)	0 (-)
118	0	2	0 (-)	1 (50%)
119	1539	443	1538 (99.94%)	440 (99.32%)
121	1856	1	1849 (99.62%)	1 (100%)
122	2471	0	2462 (99.64%)	0 (-)
123	1510	3	1462 (96.82%)	3 (100%)
124	0	47	0 (-)	42 (89.36%)
200	1479	724	1294 (87.49%)	701 (96.82%)
201	588	4	585 (99.49%)	0
202	2056	19	2024 (98.44%)	7 (36.84%)
203	2519	396	1093 (43.39%)	328 (82.83%)
205	448	10	448 (100%)	10 (100%)
207	0	104	0 (-)	80 (76.92%)
208	1584	990	1556 (98.23%)	943 (95.25%)
209	2616	1	2582 (98.70%)	1 (100%)
210	319	24	297 (93.10%)	14 (58.33%)
212	920	0	895 (97.28%)	0 (-)
213	2636	220	2078 (78.83%)	220 (100%)
214	0	18	0	18 (100%)
215	3191	164	2005 (62.83%)	143 (87.20%)
219	2010	64	1677 (83.43%)	58 (90.63%)
220	1949	0	1670 (85.69%)	0
221	2026	396	2015 (99.46%)	388 (97.98%)
222	2057	0	1962 (95.38%)	0
223	2024	473	1942 (95.95%)	367 (77.59%)
228	1684	361	1571 (93.29%)	351 (97.23%)
230	2250	1	912 (40.53%)	1 (100%)
231	314	2	314 (100%)	1 (50%)
232	0	0	0 (-)	0 (-)
233	2226	830	1915 (86.02%)	813 (97.95%)
234	2695	3	2689 (99.78%)	1 (33.33%)

5 Conclusions

It is essential to compare the proposed integrated methodology with the work of other researchers but it is important to bear in mind two distinguishing points where this work is unique. First of all, there are, at least according to the best of our knowledge, no recent works dealing with **global** classification of ECG signals. And most important in this work, we introduced and used only **mathematically obtained features** derived from the ECG signal utilizing only the detected R-peak.

However, there exist research works dealing with global classifiers, using the MIT-BIH database to distinguish 'N' and 'V' beats, usually with slight modifications in the way of obtaining the global training/testing for each one of them. Hu and his coworkers [8] achieved global accuracy of 62.2% for distinguishing 'N' and 'V' beats. The sensitivity and specificity achieved in [7] is about 80%. Jekova et al [17] reports sensitivity 78.79% and specificity of 80.61% on the global training set when distinguishing also right and left bundle branch blocks. Lower numbers but on a more difficult task are reported in [1] with 86.7% specificity and 67.3% sensitivity for V beats when classifying holter beats into five classes on the MIT database.

There are also works trying to distinguish between N and V beats using simple features derived from one-lead signal where only R-peaks were computed. In [18] Tsipouras et al. have used HRV for classification obtaining sensitivity 87.27% and specificity of 94.77% on the MIT database - but they did not use global training. In [19] four descriptive parameters were used for beat classification but the experiments were performed on the selected signals only, with unspecified training routine.

To sum up our results are at all times at least as good as and in some occasions better than those reported in the literature. The prime novelty of this work is the proposal of a new combination of features for the discrimination of "V" and "N" beats. A neural network classifier has been employed using the cross entropy error function which usually performs better for classification problems. The results are very promising and in the next phase of our research we will test the usefulness of our approach on the more demanding problem of distinguishing between five beat categories. Towards this path we will also experiment with more advanced methods for the construction of our classifier (i.e. an incremental building of the hidden layer) since the discrimination of five classes increases the need for a more customized classifier. Moreover more elaborated techniques for handling imbalanced data sets might be needed (i.e. Synthetic Minority Oversampling Technique (SMOTE) [20]). Finally, we will test our method using the AHA database which will allow for the generality of our approach to be examined.

References

1. Chazal, P., O'Dwyer, O., Reilly, R.B.: Automatic Classification of Heartbeats Using ECG Morp. Heartbeat Interval Features. IEEE Trans. Biom. Eng. 51(7), 1196–1206 (2004)

2. Cuesta-Frau, D., Perez-Cortes, J.C., Andreu-Garcia, G.: Clustering of electrocardiograph signals in computer-aided Holter analysis. Computer methods and programs in Biomedicine 72, 179–196 (2003)
3. Moody, G., Mark, R.: QRS morphology representation and noise estimation using the Karhunen-Loeve transform. Comput. Cardiol. 16, 269–272 (1989)
4. Lagerholm, M., Peterson, C., Braccini, G., Edenbrandt, L., Sornmo, L.: Clustering ECG complexes using hermite functions and Self-organizing maps. IEEE Trans. Biomed. Eng. 47, 838–848 (2000)
5. Christov, I., Herrero, G.G., Krasteva, V., Jekova, I., Gotchev, A., Egiazarian, K.: Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification. Medical Engineering & Physics 28, 876–887 (2006)
6. Tsipouras, M.G., Voglis, C., Lagaris, I.E., Fotiadis, D.I.: Cardiag Arrhytmia Classification Using Support Vector Machines. In: 3rd European Medical and Biological Engineering Conference (2005)
7. Bortolan, G., Jekova, I., Christov, I.: Comparison of Four Methods for Premature Ventricular Contraction and Normal Beat Clustering. Computers in Cardiology 32, 921–924 (2005)
8. Hu, Y.H., Palreddy, S., Tompkins, W.J.: A patient-adaptable ECG beat classifier using a mixture of experts approach. IEEE Trans. Biomed. Eng. 44, 891–900 (1997)
9. Chudáček, V., Lhotská, L., Stylios, C., Georgoulas, G.: Comparison of Methods for Premature Ventricular Beat Detection. In: ITAB Conference. IEEE, Piscataway (2006)
10. Goldberger, L., Amaral, L., Glass, L., Hausdorf, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23), e215–e220
11. Christov, I.: Real time electrocardiogram QRS detection using combined adaptive threshold. Biomed. Eng., `http://www.biomedical-engineering-online.com/content/3/1/28`
12. Haykin, S.: Neural Networks: A Comprehensive Foundation., 2nd edn. Prentice Hall, Englewood Cliffs (1999)
13. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, New York (1995)
14. Dunne, R.A.: A statistical Approach to Neural Networks for Pattern Recognition. John Wiley & Sons, New Jersey (2006)
15. Berthold, M., Hand, D.J.: Intelligent Data Analysis. Springer, Heidelberg (2003)
16. Simard, P.Y., Steinkraus, D., Platt, J.C.: Best practices for convolutional neural networks applied to visual document analysis. In: International Conference on Document Analysis and Recogntion (ICDAR), pp. 958–962 (2003)
17. Jekova, I., Bortolan, G., Christov, I.: Assessment and comparison of different methods for heartbeat classification. Medical Engineering & Physics 30(2), 248–257 (2008)
18. Tsipouras, M.G., Fotiadis, D.I., Sideris, D.: An Arrhythmia Classification System Based on the RR-Interval Signal. J. Artificial Intelligence in Medicine 33, 237–250 (2004)
19. Augustyniak, P.: The Use of Shape Factors for Heart Beats Classification in Holter Recordings. In: Conference on Computers in Medicine, pp. 47–52 (1997)
20. Chawla, N.V., Hall, L., Kegelmeyer, W.: SMOTE: Synthetic Minority Over-Sampling Technique. J. Artif. Intell. Res. 16, 3241–3357 (2002)