The impact of Negative to Positive Training Dataset.pdf

by

Journal of Physics: Conference Series

PAPER • OPEN ACCESS

The impact of Negative to Positive Training Dataset Ratio on Atrial Fibrillation Classification Machine Learning Algorithms Performance

To cite this article: Firdaus *et al* 2020 *J. Phys.: Conf. Ser.* **1500** 012131

FIRST 2019 IOP Publishing
Journal of Physics: Conference Series **1500** (2020) 012131 doi:10.1088/1742-6596/1500/1/012131

The impact of Negative to Positive Training Dataset Ratio on Atrial Fibrillation Classification Machine Learning Algorithms Performance

Firdaus[a], Andre Herviant Juliano[a], Naufal Rachmatullah[a], Sarifah Putri Rafflesia[b], Dinna Yunika Hardiyanti[b], Ahmad Zarkasi[a], Ferlita Pratiwi Arisanti[a], Siti Nurmaini[a]*

[a]Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indralaya, Indonesia.
[b]Computer Science Faculty, Universitas Sriwijaya, Indralaya, Indonesia

siti_nurmaini@unsri.ac.id

Abstract. With the few numbers of cardiologists in Indonesia who not evenly distributed, especially in rural areas, there has been a lot of smart telehealth specifically developed for heart monitoring using ECG. Many techniques have been developed to improve the accuracy of this device by using datasets that are mostly imbalanced, more positive data than negative. This paper presents the comparison of negative to positive training dataset ratio on atrial fibrillation classification machine learning algorithms performance. An AliveCor ECG recording dataset is train with deep neural networks, support vector machine and logistic regression as classifier with three different ratios, 1:1, 1:5 to 1:All. Results show an increase in classifier performance along with the increasing number of negative data.

1. Introduction

Electrocardiography plays a very important role in the medical field, because it functions to evaluate electrical activity and conditions in the human heart. The results of the evaluation will be in the form of a graph or signal that represents the human heart rate per unit of time, better known as an Electrocardiogram (ECG) [1]. Evaluating ECG signals is important, because a normal or regular heart rate is very important for pumping blood flow to all parts of the body[2], [3].

Electrocardiography can be regarded as the most common technique or method for diagnosing arrhythmias, because all activities related to electricity to the heart can be detected using an ECG signal. One of the most common types of arrhythmias, namely Atrial Fibrillation (AF). AF usually occurs because the heart beats too fast or too slow, and atrial activity is irregular and out of sync. Besides there is also no P wave on the ECG signal. If it is not immediately treated, AF can cause new problems, because about 1-2% of the population with AF has a stroke, one of which is AF. There are 2 approaches used to detect AF, the first approach is more focused on the cause of the absence of P waves or disturbances in the P wave. While the second approach is more focused on heart rate by detecting QRS waves [4], [5].

Recently there have been many techniques applied to classify cardiac arrhythmias; Deep Belief Network (DBN))[6], Support Vector Machines (SVM))[7], k -Nearest Neighbor (kNN) and Neural Network (NN)) [8], Lead Convolutional Neural Network (LCNN) [9], Convolutional Neural Network (CNN) [10], a combination of Variational Mode Decomposition (VMD) and Deep Belief Network

Journal of Physics: Conference Series **1500** (2020) 012131 doi:10.1088/1742-6596/1500/1/012131

(DBN) [11], as well as a combination of Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM)) [12] which produce varying degrees of accuracy.

There are two standard datasets used in the AF classification research, MIT-BIH Atrial Fibrillation Database[13][14][15][16][17] dan Physionet Atrial Fibrillation Database [13][18][19][20][4]. Two of them are imbalanced. There is another imbalanced dataset used on ECG classification research[21].This paper show how imbalanced AF dataset affect AF classification machine learning algorithms performance.

2. Methods

In this paper, we compare classifiers performance based on 3 ratios of negative and positive data, 1:1, 1:5, dan 1:R with the following stages of scenarios;

2.1. Dataset

The ECG dataset used for the Atrial Fibrillation classification is obtained from physionet.org[22]. The raw data obtained consists of 2 files, files with mat and hea extension. Mat files contain ECG signal data, and hea files contain information about ECG signals. This dataset is recorded using AliveCor by holding electrodes on each hand and stored in a frequency of 300 Hz and 16 bits. With an ECG signal sampling frequency of 300 Hz, the ECG signal length is equal to 300 nodes in every second.

The dataset consists of 8528 ECG leads 1 (LA-RA) signal recordings. ECG signal recordings have different signal length ranges, from the shortest 9 seconds to 60 seconds for the longest, with an average signal length of 30 seconds (Table 1). The dataset consists of 5076 normal ECG signal data, 758 atrial fibrillation, 279 noisy and 2415 other rhythms ECG signal data. This paper only uses 2 classes, AF and normal. Normal class is represented by 0 (5076 rows of data) and AF is represented by 1 (758 rows of data). So the total data used is 5834 data lines.

Table 1. Dataset Description

Class	Record	Length (nodes)				
		Mean	SD	Max	Median	Min
normal	5076	9570	3000	18300	9000	2700
AF	758	9480	3750	18000	9000	3000
other rhythm	2415	10230	3540	18270	9000	2730
noisy	279	8130	2700	18000	9000	3060
total	8528	9750	3270	18300	9000	2700

2.2. Data Pre-processing

Data Pre-processing applied to datasets are signal normalization and signal denoising. Signal normalization is used to overcome the range of amplitude in different signals. The amplitude of the normalized signal will be in the range 0 and 1. The signal normalization process uses bound normalization (equation 1) from the WFDB package.

$$F(x) = x * coef - \left(\left(\frac{\max(x)-(\max(x)-\min(x))}{2}\right) * coef\right) + mid \tag{1}$$

$$coef = \frac{(ub-lb)}{(\max(x)-\min(x))} \tag{2}$$

$$mid = \frac{ub-(ub-lb)}{2} \tag{3}$$

By normalizing the signal, the normalized signal will have an amplitude range in the range 0 and 1, with a lower limit of 0 and an upper limit of 1. The normalization process also uses a sampling

Journal of Physics: Conference Series **1500** (2020) 012131 doi:10.1088/1742-6596/1500/1/012131

frequency of signals of 300 Hz. This normalization method does not change the morphology of the signal.

The second pre-processing is signal denoising. In this paper we used discrete wavelet transform (DWT) to denoising the signals, because DWT is very efficient in terms of analysis and signal denoising. DWT is used to analyze signals by breaking the signals at different resolutions. We applied 8 levels DWT. Where the largest frequency sequence starts from the first to 8^{th} level. We used Symlets sym5 as the wavelet function to reduce or eliminate noise in the signal, while for thresholding we used Soft Thresholding with the universal threshold (equation 4) [23].

$$thresholding\ (t) = \sigma\sqrt{2\log N} \tag{4}$$

Where N is the length of the ECG signal, while σ is the standard deviation of noise (equation 5).

$$\sigma = \left(median\left|cD_j\right|/0.6457\right) \tag{5}$$

Whereas, cDj is obtained using equation 6.

$$c\widehat{D}j = \begin{cases} sign(cDj)(|cDj| - t), & |cDj|\ \geq t \\ 0, & |cDj|\ \leq t \end{cases} \tag{6}$$

Signals denoising is used to detect R peak. Detection of R peak uses the xqrs detection function of t WFDB package, where the sampling frequency becomes one of the parameters or elements used to obtain the R peak. After the R peak of all signals is obtained.

The next step is to calculate the RR interval based on the R peak. Just as in the signal normalization and R peak processes, to calculate this RR interval also uses the sampling frequency as one of the parameters. The RR interval has the main requirement, which must have at least 2 R peaks, so for signals that only produce 1 R peak it will not be able to produce RR intervals. Therefore, samples that cannot produce the interval RR will be filled with 0 (Zero) in order to pass the next stage.

2.3. Feature Extraction

A feature that will be processed into the classifier, namely Heart Rate Variability (HRV) features[24]. The feature is obtained by using the hrvanalysis package. The interval RR is used as an input. Because there are several NaN results from the RR interval. Then NaN is replaced by number 0 (Zero). The feature used in the classifier is in the Time Domain Features section; the mean of RR-intervals (mean_nni), the standard deviation of the time interval between successive normal heart beats(sdnn), the standard deviation of differences between adjacent RR-intervals (sdsd), the square root of the mean of the sum of the squares of differences between adjacent NN-intervals (rmssd), median Absolute values of the successive differences between the RR-intervals (median_nni), Number of interval differences of successive RR-intervals greater than 50 ms (nni_50), the proportion derived by dividing nni_50 (pnni_50), number of interval differences of successive RR-intervals greater than 20 ms(nni_20), the proportion derived by dividing nni_20 (The number of interval differences of successive RR-intervals greater than 20 ms (pnni_20), difference between the maximum and minimum nn_interval (range_nni), coefficient of variation of successive differences equal to the rmssd divided by mean_nni(cvsd), coefficient of variation equal to the ratio of sdnn divided by mean_nni (cvnni), the mean Heart Rate (mean_hr), max heart rate (max_hr), min heart rate (min_hr), standard deviation of heart rate (std_hr).

2.4. Classification

Before the classification process, the data is divided into 80% and 20% for each ratio. Three classifiers are prepared for 3 data ratios; Deep Neural Network (DNN), Support Vector Machine (SVM), and Logistic Regression.

2.4.1. Deep Neural Network. In the DNN scenario, the best architecture used is composed of 7 layers consisting of 1 input layer, 5 hidden layers, and 1 output layer. Where the input layer consists of

16 nodes which are the result of feature extraction from the HRV analysis, while the 5 hidden layers each use 100 nodes, and the output layer consists of 1 nodes whose contents are 0 or 1. With the activation function of each layer using Rectified Linear Units (ReLU) and Sigmoid for the output layer. The configuration of this DNN model uses Adam optimizers with a learning rate of 0.0001, and the loss function used is Crossentropy Binary. The model trained uses a batch size of 16 with an epoch of 100.

2.4.2. Support Vector Machine. SVM is chosen because this method can handle high dimensional data [25]. In the SVM scenario, the values for the parameters used are obtained from the results of automatic tuning based on the data that will be used for training, resulting in the best parameter values. The kernel used is the Radial Basis Function (RBF). The values for the C parameter provided during SVM tuning include 0.001, 0.01, 0.1, 1, 10 and the values for the gamma parameters provided, including 0.001, 0.01, 0.1, and 1. The process of tuning values for C and gamma parameters uses package GridSearchCV. The results of the tuning process produce C and gamma values of 10 and 1.

2.4.3. Logistic Regression. In the LR scenario, the value for the parameter used is the default value of the package. The value for the penalty parameter is l2, the dual formulation is set to False, the tolerance value is 0.0001.

2.5. Model evaluation
This evaluation is carried out to determine the accuracy and accuracy of the models that have been made in conducting the classification. The accuracy and accuracy of the model in conducting classification can be calculated using confusion matrix, where in the confusion matrix consists of True Negative (TN), False Negative (FN), False Positive (FP), and True Positive (TP). The four terms will be used to calculate the performance of the model that has been made[26].

3. Results, Analysis and Discussions
In the DNN scenario with 1:1 ratio, the accuracy obtained for training and testing is 97.17% and 97.36%, training and testing loss are 8.3% and 7.9% respectively. While for the DNN scenario with a 1: 5 ratio, the accuracy of training and testing are 97.99% and 95.71% respectively, training and testing loss are 12.75% and 5.69% respectively. And in the DNN scenario for 1: R ratio, the accuracy of training and testing are 97.81% and 97.60%, training and testing loss are 5.63% and 7.89%. Performance measurement of the DNN model on testing data shown in figure 1.

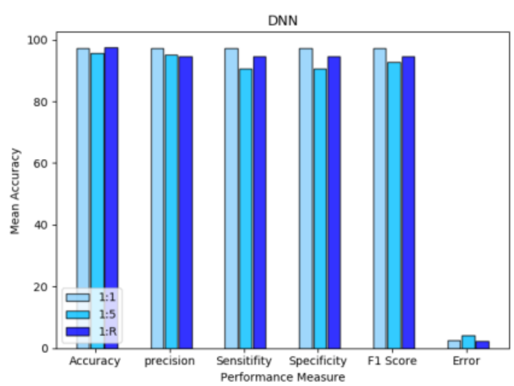


Figure 1. Performance measurement of the DNN model on testing data

In the SVM scenario for 1:1 ratio, the accuracy obtained for training and testing are 96.61% and 97.69% respectively. Whereas for 1:5 ratio, the accuracy obtained for training and testing are 97.58%

Journal of Physics: Conference Series **1500** (2020) 012131 doi:10.1088/1742-6596/1500/1/012131

and 96.37% respectively. And for 1: R ratio, the accuracy of training and testing obtained are 97.53% and 97.68%. figure 2 show the performance measurement of SVM model on testing data.

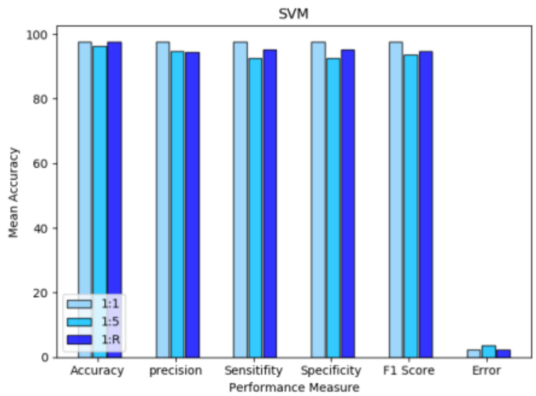


Figure 2. Performance measurement of the SVM model on testing data

In the LR scenario for of 1: 1 ratio, the accuracy obtained for training and testing are 94.14% and 97.03%, respectively. Whereas for 1: 5 ratio, the accuracy obtained for training and testing are 74.77% and 71.47% respectively. And for 1: R, ratio, the accuracy of training and testing obtained are 70.45% and 69.93%. Figure 3 shows performance measurement of LR Model on testing data.

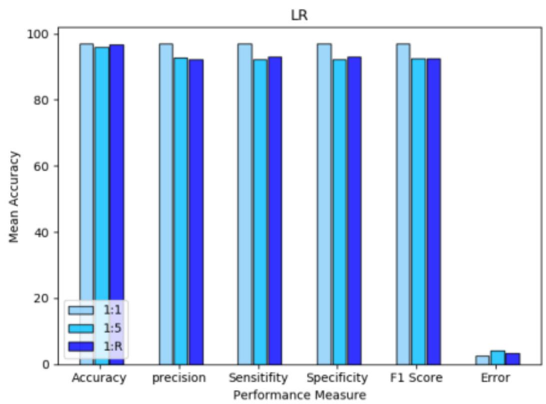
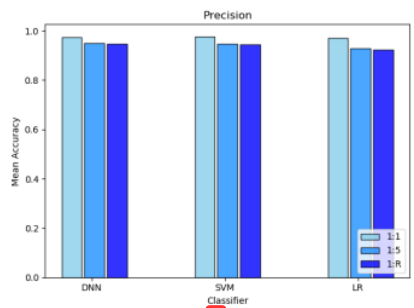


Figure 3. Performance measurement of the LR model on testing data

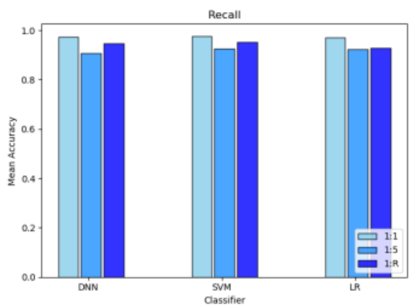
Based on the accuracy of the training and testing data obtained from the DNN, SVM, and LR models, it shown that the data ratio affects the accuracy of the DNN and LR classifiers, but not for SVM. For the DNN classifier accuracy of training and testing data, the 1: 5 ratio is affected by imbalanced data. Meanwhile, the accuracy of training and testing of the data obtained from the LR classifier shows that the more imbalance of a data, the worse performance will be.

Similar to the performance measurement obtained from the confusion matrix for testing data, a good performance value is only found in the ratio 1: 1 for each classifier (figure 4). It means the imbalance of data is very affected classifier performance.

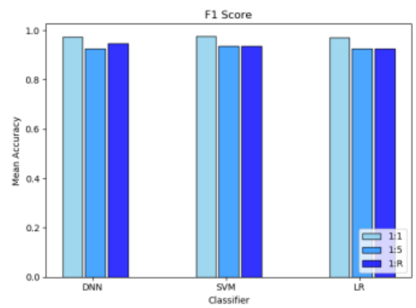
Journal of Physics: Conference Series **1500** (2020) 012131 doi:10.1088/1742-6596/1500/1/012131



(a)



(b)



(c)

Figure 4. Comparison of performance measurement, (a) precision, (b) recall, dan (c) F1 Score, of DNN, SVM dan LR classification model against negative to positive training dataset ratio on atrial fibrillation classification

4. Conclusion

This paper tested to show how negative to positive training dataset ratio on atrial fibrillation classification affect the classifier performance. The result show an increase in all classifiers (DNN, SVM, and Logistic Regression) performance along with the increasing number of negative data.

5. References

[1] S. Savalia and V. Emamian, "Cardiac Arrhythmia Classification by Multi-Layer Perceptron and Convolution Neural Networks," 2018.

[2] S. Nurmaini, R. U. P, M. N. R, and A. Gani, "Cardiac Arrhythmias Classification Using Deep Neural Networks and Principle Component Analysis Algorithm," *Int. J. Adv. Soft Compu. Appl.*, vol. 10, no. 2, 2018.

[3] P. Warrick and M. Nabhan Homsi, "Cardiac Arrhythmia Detection from ECG Combining Convolutional and Long Short-Term Memory Networks," vol. 44, pp. 1–4, 2017.

[4] S. Ghiasi, M. Abdollahpur, N. Madani, K. Kiani, and A. Ghaffari, "Atrial Fibrillation Detection Using Feature Based Algorithm and Deep Convolutional Neural Network," vol. 44, pp. 1–4, 2017.

[5] Y. Hagiwara *et al.*, "Computer-aided diagnosis of atrial fibrillation based on ECG Signals: A review," *Inf. Sci. (Ny).*, vol. 467, pp. 99–114, 2018.

[6] B. Taji, A. D. C. Chan, and S. Shirmohammadi, "False Alarm Reduction in Atrial Fibrillation

Detection Using Deep Belief Networks," *Ieee Trans. Instrum. Meas.*, vol. 67, no. 5, pp. 1–8, 2018.
[7] E. Rivera, Daniel & Veiga, Cesar & Rodriguez-Andina, J.J. & Farina, Jose & Garcia, "Using Support Vector Machines for Atrial Fibrillation Screening," *Int. Symp. Ind. Electron.*, pp. 2056–2060, 2017.
[8] K. Lee, H. O. Choi, S. D. Min, J. Lee, B. B. Gupta, and Y. Nam, "A Comparative Evaluation of Atrial Fibrillation Detection Methods in Koreans Based on Optical Recordings Using a Smartphone," *IEEE Access*, vol. 5, no. c, pp. 11437–11443, 2017.
[9] L. Jin and J. Dong, "Normal Versus Abnormal ECG Classification by the Aid of Deep Learning," *Artif. Intell. - Emerg. Trends Appl.*, 2018.
[10] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Inf. Sci. (Ny).*, vol. 405, pp. 81–90, 2017.
[11] R. Tripathy, M. R. Arrieta Paternina, J. G. Arrieta, P. Pattanaik, and G. € Sanatorio, "Automated Detection of Atrial Fibrillation Ecg Signals Using Two Stage Vmd and Atrial Fibrillation Diagnosis Index," *J. Mech. Med. Biol.*, vol. 17, no. 7, pp. 1–20, 2017.
[12] M. Zihlmann, D. Perekrestenko, M. Tschannen, and E. T. H. Zurich, "Convolutional Recurrent Neural Networks for Electrocardiogram Classification," vol. 44, pp. 1–4, 2017.
[13] X. Zhou, X. Zhu, K. Nakamura, and M. Noro, "Atrial Fibrillation Detection Using Convolutional Neural Networks," *2018 9th Int. Conf. Aware. Sci. Technol.*, pp. 84–89, 2018.
[14] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," *Comput. Biol. Med.*, vol. 93, pp. 84–92, 2018.
[15] O. Faust, A. Shen, M. Kareem, T. Ru, H. Fujita, and U. R. Acharya, "Automated detection of atrial fi brillation using long short-term memory network with RR interval signals," no. May, 2018.
[16] N. Henzel, J. Wr, and K. Horoba, "Atrial Fibrillation Episodes Detection Based on Classification of Heart Rate Derived Features," 2017.
[17] R. He, K. Wang, N. Zhao, Y. Liu, Y. Yuan, and Q. Li, "Automatic Detection of Atrial Fibrillation Based on Continuous Wavelet Transform and 2D Convolutional Neural Networks," vol. 9, no. August, pp. 1–11, 2018.
[18] T. Teijeiro, C. A. García, D. Castro, and P. Félix, "Arrhythmia Classification from the Abductive Interpretation of Short Single-Lead ECG Records," vol. 44, pp. 1–4, 2017.
[19] M. Zabihi, A. B. Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, and M. Gabbouj, "Detection of Atrial Fibrillation in ECG Hand-held Devices Using a Random Forest Classifier," vol. 44, pp. 3–6, 2017.
[20] S. Hong *et al.*, "ENCASE : an ENsemble ClASsifiEr for ECG Classification Using Expert Features and Deep Neural Networks," vol. 44, pp. 2–5, 2017.
[21] A. Darmawahyuni, S. Nurmaini, W. Caesarendra, V. Bhayyu, M. N. Rachmatullah, and others, "Deep Learning with a Recurrent Network Structure in the Sequence Modeling of Imbalanced Data for ECG-Rhythm Classifier," *Algorithms*, vol. 12, no. 6, p. 118, 2019.
[22] G. D. Clifford *et al.*, "AF classification from a short single lead ECG recording: The Physionet Computing in Cardiology Challenge 2017," *Proc. Comput. Cardiol.*, vol. 44, p. 1, 2017.
[23] H. Lin, S. Liang, Y. Ho, Y. Lin, and H. Ma, "Discrete-wavelet-transform-based noise removal and feature extraction for ECG signals," vol. 35, pp. 351–361, 2014.
[24] S. Datta *et al.*, "Identifying Normal , AF and other Abnormal ECG Rhythms using a Cascaded Binary Classifier," vol. 44, pp. 2–5, 2017.
[25] R. Smíšek *et al.*, "SVM Based ECG Classification Using Rhythm and Morphology Features , Cluster Analysis and Multilevel Noise Estimation Institute of Scientific Instruments of the Czech Academy of Sciences , Brno , Czech Republic University Hospital Brno – Children ' s Medic," vol. 44, pp. 1–4, 2017.
[26] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for

classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.

Acknowledgments
This article's publication are supported by Universitas Sriwjaya which providing financial research Hibah Penelitian Dosen Muda SATEKs 2019 contract No. 0146.209/UN9/SB3.LP2M.PT /2019 and The Kemenristek Dikti Indonesia under the Basic Research Fund Number. 096/SP2H/LT/DRPM/2019 and Universitas Sriwijaya, Indonesia under Hibah Unggulan Profesi Fund 2019.

The impact of Negative to Positive Training Dataset.pdf

ORIGINALITY REPORT

90%	24%	90%	16%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Firdaus, Andre Herviant Juliano, Naufal Rachmatullah, Sarifah Putri Rafflesia et al. "The impact of Negative to Positive Training Dataset Ratio on Atrial Fibrillation Classification Machine Learning Algorithms Performance", Journal of Physics: Conference Series, 2020 Publication	86%
2	Submitted to University of Surrey Student Paper	4%
3	iopscience.iop.org Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		