Data Augmentation for Discrimination of Atrial Flutter Mechanism Using 12-Lead Surface Electrocardiogram

Muhammad Usman Gul, Kushsairy Kadir, Muhammad Haziq Kamarul Azman

Universiti Kuala Lumpur, British Malaysian Institute, Gombak, Malaysia

Abstract

In the previous study, the atrial flutter mechanism (i.e., Focal or Macroreentrant) was differentiated from the standard 12-lead ECG by the variability of the cycle length of visible successive P-waves (between the R-R waves). This study aims to help researchers reduce imbalances through two different techniques, especially in atrial flutter. Besides, early detection of the AFL mechanism can increase the efficacy of invasive elimination. Forty-eight patients were undergone endoscopic catheter ablation for the identifications of the AFL mechanism. Accessed data only 5 *Focal AFL against* 41 *Macroreentrant AFL. Two different techniques, SMOTE and Smoothed-Bootstrap, have been used to augmented and re-balance the dataset. Furthermore, three different techniques, the Goodness-of-Fit test (Chi-Square), the Root Mean Square Error between synthetic and original performance (accuracy, specificity, sensitivity), and the descriptive statistical test, have been used to validate the augmented dataset. The proposed model has been extracted several features derived from statistical analysis of the intervals of successive atrial rhythm to discriminate the AFL mechanism. The performance has been evaluated by three linear classifiers Linear Discriminant Analysis (LDA), Logistic Regression (LOG), and Support Vector Machine (SVM). The synthetic data generated by Smoothed-Bootstrap has been much closer to the original dataset and relatively better than SMOTE technique. The LOG classifier achieved its average performance with accuracy, specificity, sensitivity,* 71.08%, 77.13%, *and* 65.12%, *respectively. The variability in cycle length of consecutive P-waves from the surface ECG has differentiated the Focal AFL from Macrorrentrant AFL. Smoothed-Bootstrap is a suitable technique in AFL cases to minimize the imbalance issue.*

1. Introduction

Atrial flutter (AFL) is a type of Supraventricular tachycardia in which electrical signals frequently transmit through various physiological pathways other than sinus rhythm [1]. It is associated with a variety of risks and complications, such as stroke and heart attacks. The mechanism of AFL has been categorized into two groups, Focal (F-AFL) and Macroreentrant (M-AFL), based on its diameter of activation spreading. Within atria activities, electrical propagation of less than 2 cm in diameter is considered F-AFL and more than that as an M-AFL mechanism. Discrimination of F-AFL from M-AFL through non-invasive techniques (i.e., 12-lead surface ECG) can improve the efficacy of its existing catheter ablation treatment. The AFL mechanism can be differentiated using machine learning classifiers in the proposed features to identify variations in the surface ECG [2]. However, a large number of data is required for the use of classifiers in machine learning. Unfortunately, the AFL-related data and its labeled mechanism are not available in the public database [3]. Furthermore, our database of AFL recordings present high prevalence of M-AFL as compared to F-AFL. This will inevitably lead to bias when using machine learning techniques for classification. These types of issues can be addressed using data augmentation techniques to minimize imbalances. This paper has two purposes: the first is to compare two data augmentation techniques to determine which technique is best for AFL cases. The second is to improve the F-AFL discrimination from the M-AFL by extracting the features from 12-lead surface ECG. In the previous study, It had been hypothesized that the variation of two visible consecutive atrial activities (non-overlap, two and more than two P-waves) within R-R waves could discriminate the AFL mechanism [2]. This research aims to improve the performance of previous research by minimizing the bias in the original dataset. The performance of three classifiers LDA, LOG, and SVM has been evaluated concerning maximum accuracy, sensitivity, and specificity.

2. Material and Method

2.1. Data Description and Prepossessing

A total of 46 ECG records has been selected after elimination of un-labeled AFL mechanism and non-consecutive ECG (which have block ratio less than $3:1$) out of 61

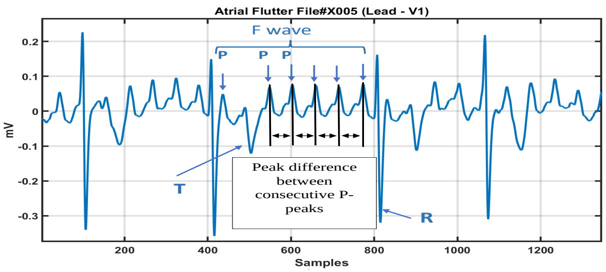

Figure 1. Proposed measures of peak-to-peak interval between consecutive P-waves of extract F-waves in AFL

ECG records. All are 1-minutes long recording and has been acquired at 2000 Hz sampling rates. In detail, the ratio of selected data is $41 : 5$, M-AFL and F-AFL ratio respectively. Furthermore, a 50 Hz power line filter and bandpass (Chebychev-II) filter with $0.5 - 60$ Hz cutoff frequencies have been used for noise elimination.

2.2. Consecutive P-P Interval

It is not the only way to differentiate Focal from Macroreentrant by observing isoelectric baseline. At times, F-AFL does not exhibit isoelectric intervals where the atrial conduction velocity is high-speed, and conduction times through the atria are slow in comparison to the tachycardia rate [4, 5]. On the other hand, certain M-AFL shows isoelectric intervals when conduction happens through a narrow isthmus that does not produce enough voltage to induce a surface ECG deflection [6]. Macroreentrant AFL's loop is usually very stable, which accounts for the regularity of the ECG patterns. However, since the propagation direction is not guaranteed to be identical at each beat in F-AFL, we should expect substantial variability in the ECG pattern. Therefore, it is hypothesized that the mechanism of AFL can be differentiated from 12-lead surface ECG based on intervals between two & more than two continuous atrial activities within R-R intervals, which is shown in Fig-1. These atrial activities have been detected by the Generalized Likelihood Ratio Test (GLRT). There are 444 intervals of 5 F-AFL ECGs and 2546 intervals of M-AFL ECGs in the original dataset. These intervals have been used in the data augmentation techniques to balance data in ECG numbers.

2.3. Data-Augmentation Techniques

Numerous strategies for compensating for the imbalance have been suggested [7], including resampling the dataset. Either reduce the sample size of the majority class or increase the sample size of the minority class. In this paper, a comparative study between two data-augmentation techniques, namely Synthetic Minority Oversampling Techniques (SMOTE) [8] and Smoothed Bootstrap [9], has been proposed and determined which one is the better technique for data-augmentation in the AFL case. In These augmentation techniques, the consecutive p-wave intervals have been used from the original ECG as a input data to generate synthetic ECG intervals (augmented-dataset).

2.4. Validation of Data-Augmentation Techniques

Three different kinds of methods have been proposed to validate the augmented data with their original dataset properties.

1. Statistical Descriptive of Dataset
   This method has compared the statistical properties of both original and synthetic datasets and observed how much synthetic dataset has a difference in percentage from their original dataset properties. The difference should be as low as possible to accept that the augmented data has correctly generated like their original one. The mean, variance, and skewness have been taken into account in the comparison method.
2. Goodness-of-Fit Test
   A typical test for the goodness-of-fit test is a Chi-square test. Chi-square determines if a relationship exists between categorical data. This test has been used to compare the ratio or frequency between the original and synthetic datasets. With this method's help, it has been verified that whether the null hypothesis is accepted with their original data set or not. The accepted null hypothesis has further validated that both original and synthetic datasets are similar to each other. Whereas, significance level has been used as default as $p < 0.05$.
3. Root-Mean-Square-Error (RMSE)
   In the third and last test for validation of augmented data has been used RMSE. This method has been used after evaluating the classifiers' performance (accuracy, specificity, and sensitivity). Moreover, synthetic data's performance must have the same performance as the original data to validate it. Two dataset were used at the classifiers' input, the original dataset (5 Focal & 41 Macro ECGs) and the synthetic dataset (5 synthetic Focal & 41 original Macro ECG). However, the previous two methods have only original and synthetic Focal ECG as the input dataset. The RMSE has been measured between the original and synthetic dataset's performance, as shown in Fig. 3. Moreover, its measurement must be minimum value to validate the dataset.

2.5. Feature Extraction

The features extracted from successive intervals of atrial activities have been divided into four statistical descriptive

Table 1. Feature List

Statistical Descriptive	Variable
Central Tendency	Mean
	Median
	Mode
Dispersion	Standard Deviation
	Variance
Shape	Skewness
	Kurtosis
Length (Intervals)	Minimum
	Maximum
	Sum of all

analysis parts. Their names are Central tendency, dispersion, shape, and length of data. Central tendency gives information about the center of the data such as Mean (average), Median (middle value), and Mode (most frequent value). Dispersion of data has two variable names standard deviation and variance. Further, the shape of data has two variables: skewness for symmetric and kurtosis for outliers of the data. In contrast, the data length has minimum interval, maximum interval, and total length of all consecutive intervals. In short, a total of ten features have been extracted from the calculated intervals of proposed successive intervals of atrial activities as shown in Table 1.

2.6. Classification

Classifiers are used to train the data in order to discriminate against the AFL mechanism. The output of these classifiers is directly related to the dataset's attributes. Three distinct types of classifiers were evaluated for their reliability. The linear discriminant analysis (LDA), the logistic regression analysis (LOG), and the linear support vector machine (SVM) are the terms used to describe these classifiers. The previous article described the learning setup. To summarize, an exhaustive wrapper evaluation of all possible combinations of features was performed, and classifier performance indicators were calculated for each combination. This was done for all three classifiers and all two datasets [2].

3. Results and Discussion

The accessed data set is highly imbalanced. Its performance has been obtained from the original data set is average accuracy, specificity, and sensitivity are $90.48, 16.75, 99.47$ LDA, $90.96, 16.87, 100$ LOG, and $89.13, 0, 100$ SVM in percentage, respectively. The maximum performance of the LOG classifier at original dataset has been shown in Fig. 2. It has been observed that the specificity is poor with less than 17% in all classifiers as compare to other performances. The main reason is biased

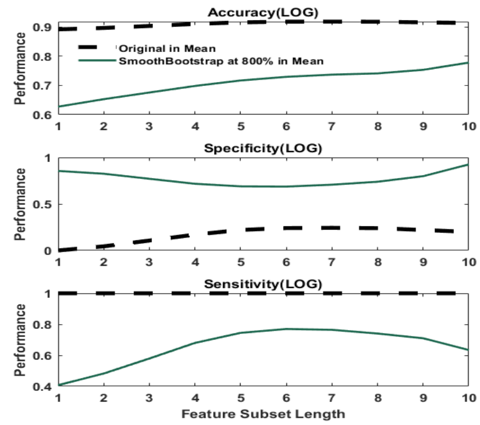

Figure 2. Performance of the LOG classifier at the original and the 800% SmoothedBootstrap dataset

in the original dataset because the classifiers do not have adequate ECGs of the minority F-AFL class for proper training. Therefore, the SMOTE and Smoothed-Bootstrap techniques have been used for data augmentation in the minority class only. Initially, five synthetic F-AFL ECGs were generated based on five original Focal minority ECGs by two proposed techniques. Then it has been verified with the available original minority dataset by three different methods. Three main parameters (Mean, Variance, and Skewness) have been chosen from statistical properties that reflect the dataset's behavior. It has been observed that Smoothed-Bootstrap has minor percentage error in the Variance and Skewness parameter, whereas SMOTE has only better in the Mean parameter. The result has been shown in Table-2. For further validation, the Chi-square test has also been performed to identify that the data of Smoothed-Bootstrap has accepted the null hypothesis with a p-value of 0.1934, which means its data set is similar to the original minority data set. Whereas SMOTE has rejected the null hypothesis with a p-value of less than 0.005, its data set is not similar to the original minority data set. This test has been used to compare the ratio or frequency between the original and synthetic datasets with $6\ msec$. Moreover, the minimum RMSE value has also been found in the Smoothed-Bootstrap, as shown in Fig 3. In detail, the average RMSE of LOG has of 0.0055 in the accuracy performance of the smoothed bootstrap. Whereas the SMOTE has 0.0060, which is higher, Smoothed-Bootstrap is selected as better techniques than Smote. The other specificity and sensitivity performances of the same LOG classifier have approximately equal to each other with 0.0545 and 0.0001 average RMSE, respectively. In summary, overall, three tests have been performed for data validation, and Smoothed-Bootstrap has been found prominent in two tests; however, in the descriptive statistical compar-

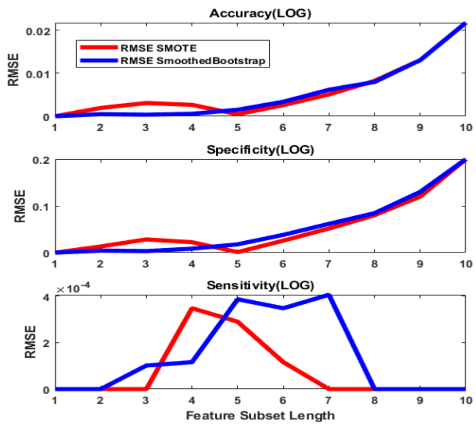

Figure 3. RMSE of augmented techniques (SMOTE, Smoothed-Bootstrap) between the original dataset for LOG classifier

Table 2. Statistical Descriptive Analysis of Minority Dataset

	Mean		Variance		Skewness	
	Difference from the Original in Percentage					
	SMOTE	Smoothed	SMOTE	Smoothed	SMOTE	Smoothed
Synthetic	-0.170	0.486	28.191	-2.195	22.1289	3.7883

ison, it is partially prominent. So, it has been concluded that the Smoothed-Bootstrap is the best technique to utilize for data augmentation to balance the minority class with their majority class. Finally, the Smoothed-Bootstrap technique has been preferred to augment the minority F-AFL dataset with 800% to balance with the majority M-AFL. The performance of such a balanced dataset from three classifiers has been shown in Table-3. In the data validation, comparing the original dataset with its copy synthetic dataset, the Smoothed-Bootstrap is a good technique compared to SMOTE. Therefore, the performance has only focused on the Smoothed-Bootstrap. However, SMOTE has a high result but has also higher RMSE while validating, and has more error or variation in the 800% dataset. Furthermore, it has been observed that the LOG has maximum performance with $71.08\%, 77.13\%, 65.12\%$ accuracy, specificity, and sensitivity, respectively. According to performance, it has been validated that the proposed hypothesis of consecutive P-P wave intervals is a valuable method for discriminate the F-AFL from M-AFL.

Table 3. Mean values of classifier's performance after 100 iterations (Balanced dataset 40:41 Focal-Maro ECG ratio) at 800% Augmentation

Techniques	Performance	LDA	LOG	SVM
	Accuracy	68.28	73.58	70.28
SMOTE	**Specificity**	70.61	85.27	73.96
	Sensitivity	66.31	62.42	66.67
Smoothed-	**Accuracy**	62.93	**71.08**	67.67
	Specificity	59.47	**77.13**	68.27
Bootstrap	**Sensitivity**	66.23	**65.12**	67.60

4. Conclusion

This study draws two conclusions: first, the smoothed-Bootstrap technique is the most effective technique for balancing the majority class's data set. Second, the variability in the cycle length of consecutive P-waves from the surface ECG has discriminated the F-AFL from M-AFL, supporting the proposed hypothesis. The 12-lead surface ECG has been utilized to distinguish the mechanism of AFL.

Acknowledgements

We thank Dr. D. G. Latcu from Centre Hospitalier Princesse Grace, Monaco for having provided us with the data for this study.

References

[1] Cosío FG, Pastor A, Núñez A, Magalhaes AP, Awamleh P. Atrial flutter: an update. Revista Espanola de Cardiologia English Edition 2006;59(8):816–831.

[2] Gul MU, Kadir K, Azman MHK. Novel classifiers to differentiate focal and macroreentrant atrial flutter using 12-lead surface electrocardiogram. In 2020 Computing in Cardiology. 2020; 1–4.

[3] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. circulation 2000;101(23):e215–e220.

[4] Shah D. Ecg manifestations of left atrial flutter. Current opinion in cardiology 2009;24(1):35–41.

[5] Pascale P, Roten L, Shah AJ, Scherr D, Komatsu Y, Ramoul K, Daly M, Denis A, Derval N, Sacher F, et al. Useful electrocardiographic features to help identify the mechanism of atrial tachycardia occurring after persistent atrial fibrillation ablation. JACC Clinical Electrophysiology 2018;4(1):33–45.

[6] Markowitz SM, Thomas G, Liu CF, Cheung JW, Ip JE, Lerman BB. Atrial tachycardias and atypical atrial flutters: mechanisms and approaches to ablation. Arrhythmia electrophysiology review 2019;8(2):131.

[7] Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. Rusboost: Improving classification performance when training data is skewed. In 2008 19th international conference on pattern recognition. IEEE, 2008; 1–4.

[8] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research 2002;16:321–357.

[9] Silverman B, Young G. The bootstrap: to smooth or not to smooth? Biometrika 1987;74(3):469–479.