Abnormal Heartbeat Detection Based On ECG Using LSTM and Machine Learning Approach

MD Meher Hassan Chowdhury

Abstract

The heart is the most critical organ of human body. Any irregularity of the heart beat will cause abnormal heart rhythm. Clinical analysis of Electrocardiogram (ECG) signals is used to detect the abnormality but also it requires expertise to classify normal and other types of heart beats. This paper proposes a deep learning approach LSTM Auto-encoder along with Machine learning approaches (KNN, SVM) to classify abnormalities in heart beat.

Keywords— Abnormal Heartbeat Detection; LSTM Auto-encoder; Deep Learning; Electrocardiogram; Heartbeat Classification.

1 Introduction

A pulse happens when the ventricle and atrium depolarization and re polarization rhythmically. ECG is an easily used medical strategy that records during a cardiac activity. ECG can be recorded intrusive and noninvasively. A heartbeat involves a progressive period of systole and diastole. Each healthy pulse includes a P wave, QRS wave, and T wave that represents repolarization and depolarization of the atriums and ventricles within the heart. The characteristics of each wave, such as time of the event, beat adequacy, and frequency, carry different data about the pulse. In a healthy person, the pulse is called sinus cadence and happens an average of 60 to 100 beats per diminutive [1]. Abnormal heart beats may happen due to different neurotic and physiological variables and cause heartbeats to move away from sinus rhythms. So, it can be characterized as any abnormalities within the cardiac cycle [2]. Abnormal heartbeats provide important data around heart muscle diseases. It may cause brokenness of the heart muscle and cause not pump sufficient blood to the organs and bolster themselves [3]. Nowadays, numerous sorts of cardiac diseases can be treated with different drugs or surgical strategies. In arrange to minimize the degeneration caused by abnormal heart rates, rapid, effective, and precise discovery of identifying anomalies has significant importance [4]. Clinically investigation of the ECG flag is a relative strategy and not enough to rapidly identify abnormalities within the heart beat. Automatic anomalies detection is one of the foremost prevalent investigate points of recent years, encouraging quick reaction to intense ailments and alleviating the workload of the doctors moreover improving diagnostic productivity and precision [5-7].

2 Related Work

Numerous ponders within the literature distinguish the unsettling influences in ECG signals by utilizing different flag handling, including preprocessing of signals, feature extraction, and classification procedures [8-11]. Anomaly detection by utilizing conventional machine learning strategies comprises 4 primary stages; ECG signal preprocessing, heartbeat segmentation, including extraction, and classification. In each of the four steps, an activity is taken and the last objective is the distinguishing proof of the type of heartbeat [12].

de Albuquerque et al. [13], used and compared six distance metrics, six feature extraction algorithms, and three classifiers and obtain support vector machines (SVM)-based classifier presented the highest accuracy.

Amrita Rana; Kyung Ki Kim.[14], used RNN and single layer LSTM and compared the result with CNN where there proposed LSTM method showed the highest accuracy.

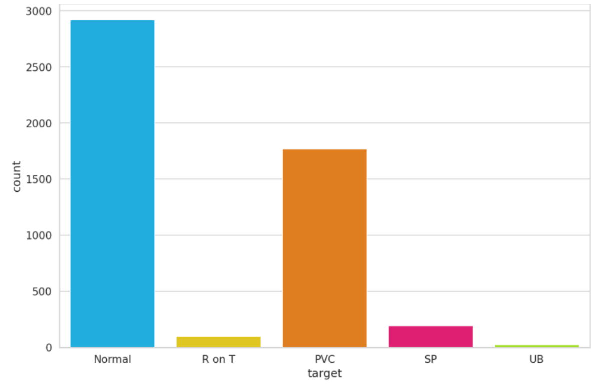
Yuanyuan Qu; Nina Zhang; Yue Meng; Zhiliang Qin; Qidong Lu; Xiaowei Liu.[15] used Deep Wavenet LSTM, CNN, SVM and Random Forest where Deep Wavenet LSTM showed 96.8% accuracuy.

In this paper, I proposed a deep learning approach for the detection of abnormal heartbeats from ECG signals.

3 DATA

The dataset is a 20-hour long ECG downloaded from Physionet. The name is BIDMC Congestive Heart Failure Database(chfdb) and it is record "chf07". It was originally published in "Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23)". The data was preprocessed in two steps: (1) extract each heartbeat, (2) make each heartbeat equal length using interpolation. This dataset was originally used in paper "A general framework for never-ending learning from time series streams", DAMI 29(6). After that, 5,000 heartbeats were randomly selected. The patient has severe congestive heart failure and the class values were obtained by automated annotation.

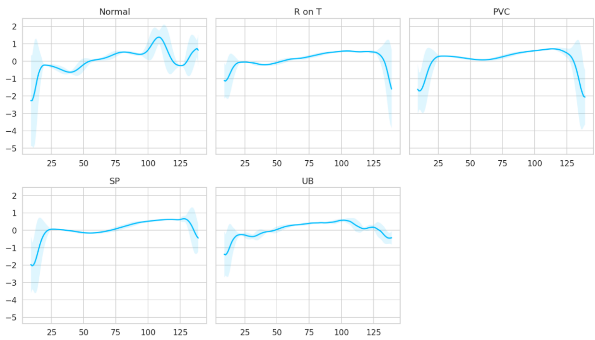
The dataset have 5 types of hearbeats (classes): Normal (N), R-on-T Premature Ventricular Contraction (R-on-T PVC), Premature Ventricular Contraction (PVC), Supra-ventricular Premature or Ectopic Beat (SP or EB), Unclassified Beat (UB).



4 Methods and Tools

4.1 Data Analyzing

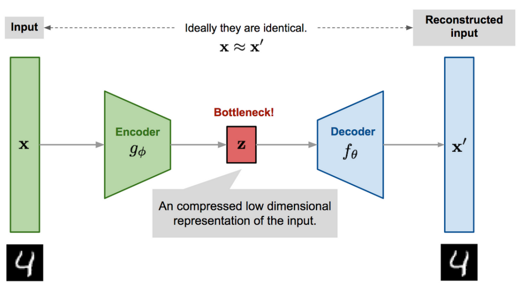
Because of the imbalance of the dataset, all the classes without normal were merged together in order to use that as testing data and the normal class as training. One Standard deviation was measured for each of the classes to check the distribution of each class from the mean.



The figure above shows that class normal has distinctive pattern than other classes.

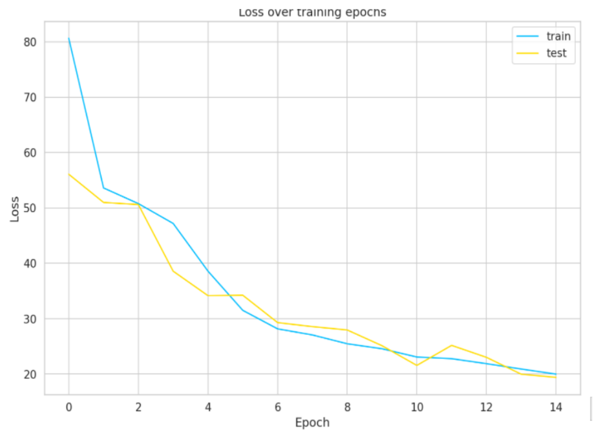
4.2 Proposed Architecture

LSTM Autoencoder is used to classify abnormal hearbeats. The Autoencoder's work is to urge a few input information, pass it through the show, and get a remaking of the input. The recreation ought to coordinate the input as much as conceivable. The trick is to utilize a small number of parameters, so the model learns a compressed representation of the data. In a sense, Autoencoders try to memorize as it were the most critical highlights (compressed adaptation) of the information. To classify a grouping as Abnormal or an Normal, a threshold was chosen over which a pulse is considered unusual.

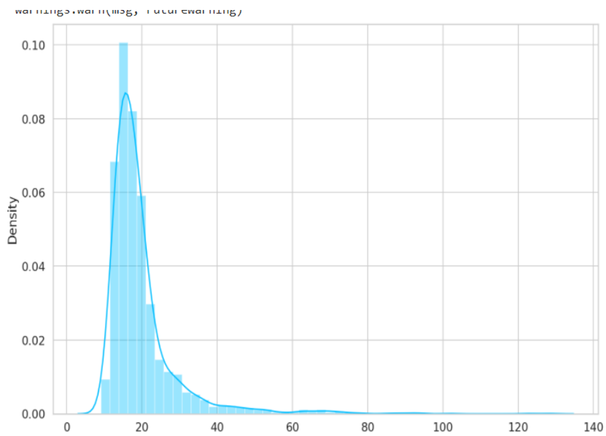


The following image is collected from https://lilianweng.github.io/lil-log/2018/08/12/from-autoencoder-to-beta-vae.html

The encoder has two layers to compress the data and the decoder also has two layers with an output layer to get the final reconstruction in our model. Total 15 number of epochs were run where at each epoch the model feed all the training examples and evaluated the performance on the validation set. The idea is to minimize the mean absolute error in order to get better reconstruction.



After training the model, the reconstruction error was calculated over training set.



It can be seen from the figure that reconstruction loss is around 40 for the training examples. Using this idea the problem is solved using binary classification problem. If the reconstruction loss is less than 40 or any threshold value set based on the changes in epochs the class can be classified as Normal and if its higher than it can be classified as Abnormal heart beat.

4.3 KNN and SVM without SMOTE

For KNN the dataset was splitted into attributes and labels. To avoid over-fitting, dataset was divided into training and testing splits, which gave better idea as to how the algorithm performed during the testing phase. This way the algorithm is tested on un-seen data. The split was 80% train data and 20% test data. Before making any predictions, feature scaling was performed so that all of them can be uniformly evaluated. The gradient descent algorithm was used to normalize features. After training the model KNN performed really good predicting normal classes.

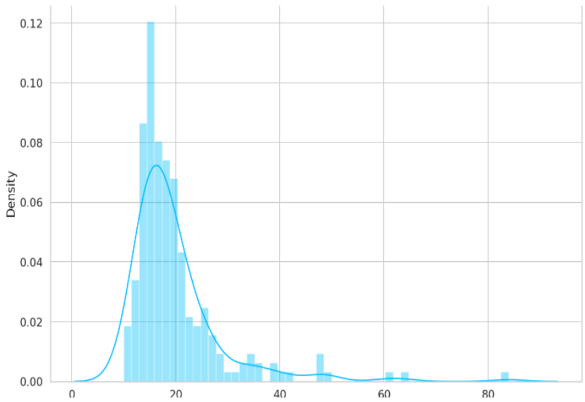
For SVM the dataset was splitted in the same portion as KNN and no feature scaling was performed. SVM also showed better results predicting Normal Classes.

4.4 SVM with Oversampling

From the results before SMOTE it can be seen that because of very less value of class 4 and 5, the algorithms could not predict anything for them. So I increased all the classes with the same number of samples that the highest class have. After doing that the algorithm showed some prediction results but it reduced the accuracy of the algorithms.

5 Results

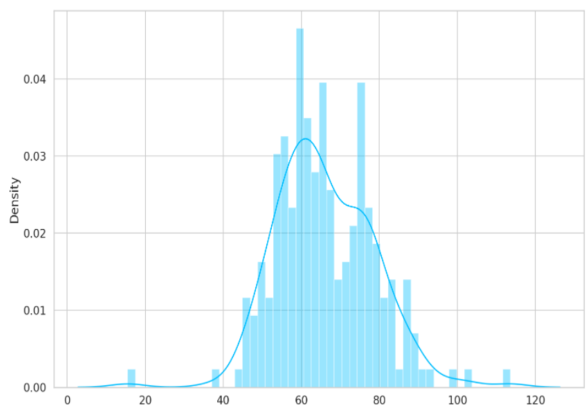
The LSTM auto-encoder model predicting losses for Normal heartbeats



Based on the Threshold the number of correct normal beats detection is

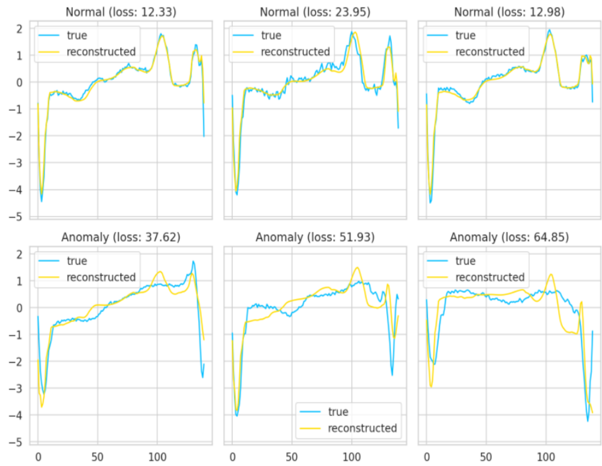
```
Correct normal predictions: 189/219
```

And predicting the Abnormal Heartbeat loss is



In the same way the number of total abnormal heartbeat will be all the points that are more than the Threshold.

Some of the reconstructed and real values can show the difference of how close they are.



For normal heartbeats the loss for real and reconstruction is less than the Threshold where for Abnormal heartbeats the loss is higher than the Threshold.

KNN showed 95% training accuracy where the model actually performed better predicting Normal Heartbeat.

```
[[576   1   1   0   0]
 [  1 345   4   2   0]
 [  5   6   9   2   0]
 [  1  25   0  19   0]
 [  3   0   0   0   0]]
              precision    recall  f1-score   support

           1       0.98      1.00      0.99       578
           2       0.92      0.98      0.95       352
           3       0.64      0.41      0.50        22
           4       0.83      0.42      0.56        45
           5       0.00      0.00      0.00         3

    accuracy                           0.95      1000
   macro avg       0.67      0.56      0.60      1000
weighted avg       0.94      0.95      0.94      1000
```

Performance of SVM before Oversampling

```
[[572   3   2   0   1]
 [  1 340   7   4   0]
 [  1   7  12   2   0]
 [  2  30   1  12   0]
 [  3   0   0   0   0]]
              precision    recall  f1-score   support

           1       0.99      0.99      0.99       578
           2       0.89      0.97      0.93       352
           3       0.55      0.55      0.55        22
           4       0.67      0.27      0.38        45
           5       0.00      0.00      0.00         3

    accuracy                           0.94      1000
   macro avg       0.62      0.55      0.57      1000
weighted avg       0.93      0.94      0.93      1000
```

Dataset after Oversampling

```
[('1', 2919), ('2', 2919), ('3', 2919), ('4', 2919), ('5', 2919)]
```

Performance of SVM after Oversampling

```
[[447  29  87  11   4]
 [  4 318  18  11   1]
 [  1   6  15   0   0]
 [  2  37   2   4   0]
 [  2   0   0   0   1]]
              precision    recall  f1-score   support

           1       0.98      0.77      0.86       578
           2       0.82      0.90      0.86       352
           3       0.12      0.68      0.21        22
           4       0.15      0.09      0.11        45
           5       0.17      0.33      0.22         3

    accuracy                           0.79      1000
   macro avg       0.45      0.56      0.45      1000
weighted avg       0.86      0.79      0.81      1000
```

6 Discussion and conclusion

In this paper, I used different approaches to classify Normal and Abnormal heartbeats. LSTM autoencoder successfully detected all the numbers of Normal or Anomaly heart beats. I also showed how through calculating reconstruction loss the problem can be solved by Binary classification using Threshold value. Also I showed how KNN performed with feature scaling and without SMOTE for the dataset. And the performance of SVM before and after performing Oversampling.

References

[1] E. Suaste-Gómez, D. Hernández-Rivera, A. Sánchez-Sánchez, and E. Villarreal-Calva, ”Electrically insulated sensing of respiratory rate and heartbeat using optical fibers,” Sensors, vol. 14, no. 11, pp. 21523-21534, 2014.

[2] G. Tse, ”Mechanisms of cardiac arrhythmias,” Journal of arrhythmia, vol. 32, no. 2, pp. 75-81, 2016.

[3] E. Benmalek and J. Elmhamdi, ”Arrhythmia ECG signal analysis using non parametric time-frequency technique,” in 2015 International Conference on Electrical and Information Technologies (ICEIT), 2015, pp. 281-285: IEEE.

[4] 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,” Information sciences, vol. 405, pp. 81-90, 2017.

[5] J. Cao, M. L. Brown, E. J. Higgins, and P. J. Degroot, ”Automatic thresholds for atrial tachyarrhythmia detection in an implantable medical device,” ed: Google Patents, 2019.

[6] J. F. Li, K. Chen, J. Ling, Y. Zhan, and G. Manogaran, ”Automatic diagnosis of cardiac arrhythmia in electrocardiograms via multigranulation computing,” Applied Soft Computing, vol. 80, pp. 400-413, 2019.

[7] J. Z. Wu, T. Lan, C. Yang, and Z. Nie, ”A Novel Method to Detect Multiple Arrhythmias Based on Time-frequency Analysis and Convolutional Neural Networks,” IEEE Access, 2019.

[8] J.A. J. Prakash and S. Ari, ”A system for automatic cardiac arrhythmia recognition using electrocardiogram signal,” in Bioelectronics and Medical Devices: Elsevier, 2019, pp. 891-911.

[9] M. Alfaras, M. C. Soriano, and S. Ortín, ”A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection,” Frontiers in Physics, vol. 7, p. 103, 2019.

[10] R. He et al., ”Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional LSTM,” IEEE Access, vol. 7, pp. 102119-102135, 2019.

[11] E. Izci, M. A. Ozdemir, R. Sadighzadeh, and A. Akan, ”Arrhythmia detection on ECG signals by using empirical mode decomposition,” in 2018 Medical Technologies National Congress (TIPTEKNO), 2018, pp. 1-4: IEEE.

[12] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, ”ECG-based heartbeat classification for arrhythmia detection: A survey,” Computer methods and programs in biomedicine, vol. 127, pp. 144-164, 2016.

[13] V. H. C. de Albuquerque et al., ”Robust automated cardiac arrhythmia detection in ECG beat signals,” Neural Computing and Applications, vol. 29, no. 3, pp. 679-693, 2018/02/01 2018.

[14] A. Rana and K. K. Kim, ”ECG Heartbeat Classification Using a Single Layer LSTM Model,” 2019 International SoC Design Conference (ISOCC), 2019, pp. 267-268, doi: 10.1109/ISOCC47750.2019.9027740.

[15] Y. Qu, N. Zhang, Y. Meng, Z. Qin, Q. Lu and X. Liu, ”ECG Heartbeat Classification Detection Based on WaveNet-LSTM,” 2020 IEEE 4th International Conference on Frontiers of Sensors Technologies (ICFST), 2020, pp. 54-58, doi: 10.1109/ICFST51577.2020.9294765.