CMP611

ADVANCED BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

— ANOMALY DETECTION FOR ARRHYTHMIA —

[DISORDER – IRREGULARITY]

ELECTRO CARDIAGRAM SEQUENCES

PREPROCESSING

SYNTHETIC SEQUENCES

SEGMENTATION

RESULTS

Kürşat Çakal
N21138461
kursatckl@gmail.com

MILESTONES

01 Dataset

MIT-BIH Arrhythmia
QRS Complex
AHA Codes
American Hospital Association

02 Dynamics of New Approach

Wavelet Denoise (White Gaussian Noise Filter)
R Peak Detection - Pan-Tompkins Algorithm
Synthetic Data Generation - Z-Score
Novel Augmentation Method
ECG Sequence Segments
(Segment X & Segment Y)

03 Architecture

Architecture
Training Process
Equations

04 Results

Confusion Matrix

05 Comparison

Literature Review

PROBLEM

Problem : Heart Attacks – Strokes – Irregular Heartbeats

Why Important ?

The increasing **industry needs** and the fast-paced **life marathon** in our lives have **brought** about various **changes** in our **lifestyle** like **diet**, the form of **entertainment**, **stress relief**, **lack of sports** activities that result in **unhealthy routines**. These routines bring humankind's life lots of disease in its root. Although technological developments and humanity increase its presence in discoveries, **investments** for the solution of health problems cause very **high costs**.

One of the most important effects of **unhealthy routines** is based on blood vessels and **heart problems** that tackle under the **Cardiovascular System Diseases**. Key symptoms of CVDs are included **hypertension**, **obesity**, **diabetes**, **smoking**, **hyperlipidemia**, and **lack of physical activities** as mentioned in [9]. All of these factors cause lots of deaths. Over 17 million people died from CVDs in 2019 as an estimation that equivalent to 32 percent of all cases in the world and 85% of these mortalities are caused by **heart attacks** and **stroke** according to [10]. The way that we can keep track of these abnormalities **controlling** some of the morphological sides of the human body is **arrhythmia monitoring (ECG Signal Sequence Analyzing)**.

Is is possible to early detection ?

Yes. It is effectively possible with the proposed study.

DATASET

MIT-BIH Arrhythmia Database [1,2,3]

- 4000 - Holder Record
- 48 Set / 25 File
- Each Set 30 Min at 360 Hz
  - 600.000 Sequence Point for A/Signal
- Age 24-89
- Years 1975-1979

@Beth Israel Hospital ArrLab

Definitions	Heartbeat Class Conformity																
AHA Codes	N							V	F	O	N		E	P	F	O	Q
MIT-BIH Type	.	L	R	A	a	J	S	V	F	!	e	j	E	P	f	O	Q
Signal Amount	75052	8075	7259	2546	150	83	2	7127	803	472	16	229	106	7028	982	193	33
Total	110156																
%	68,13	7,33	6,58	2,31	0,13	0,075	0,001	6,45	0,72	0,42	0,014	0,207	0,096	6,38	0,89	0,175	0,02

DATASET

N	S	V	F	Q
· Normal	· Atrial premature	· Premature ventricular contraction	· Fusion of ventricular and normal	· Paced
· Left bundle branch block	· Aberrant atrial premature	· Ventricular escape	· Fusion of paced and normal	
· Right bundle branch block	· Nodal (junctional) premature	· Unclassifiable		
· Atrial escape	· Supra-ventricular premature			
· Nodal (junctional) escape				

Different Functionalities
Different Wave Distributions

PRE-PROCESSING - TWO TAIL Z SCORE

Z	1 - P	P	Z	1 - P	P
1.237	0.784	0.216	0.792	0.572	0.428

Eliminating Data Point Offset Effects

(Compare Wave Points to obtain normal distribution interval for augmentation phase)

- Denoising
  - White Gaussian Noise
- Synthetic Data Generation
- R Peak Detection
  - Pan Tompkins Algorithm
- Z-Score
- Augmentation
- Segmentation

PRE-PROCESSING

QRS Complex

1

2

4

3

QRS Complex

4

5

7

6

Raw Signal

1

2

4

3

PRE-PROCESSING

PRE-PROCESSING

NOVEL ECG AUGMENTATION

Real Life Input Flow Simulation

- Corruption
- Time Shifting
- Genetic Effects
- Input Source Precision
- External Effects
- Divergent Time Phase

?

NOVEL ECG AUGMENTATION

NOVEL ECG AUGMENTATION

ARCHITECTURE-FLOW DIAGRAM

- Lightweight Convolutional Layer Model
- Activation Function (Sigmoid)
- Loss Function (MSE)

Regularization Rate (λ)	0.2
Learning Rate (x)	0.1
Momentum (m)	0.7

- Augmentation
  - 348.43 sec
- Epoch
  - 5947.29 sec
  - 2731.91 sec
  - 5817.32 sec

Layer No	Layer Type	Layer Size	Layer Depth	Kernel Size
1	Input	1 x 260	N/A	N/A
1	Convolutional	1 x 260	3	5
2	Activation	1 x 196	3	5
3	Max-pooling	1 x 196	3	5
4	Flatten	1 x 153	1	5
5	Connected	1 x 765	1	1
6	Activation	1 x 5	1	N/A

Laptop

- 1 x Intel Core 2.80 GHz
  - (4200H)
- 16 GB RAM
- Set X
  - ✓ 2-500 Ep
- Set Y
  - ✓ 2-500 Ep
- Matlab R2021a

RESULTS

~(20/80)

Automized Augmentation

- 5947.29 sec

75722, 27330

- 2731.91 sec

41560, 12920

- 5817.32 sec

79421, 26595

The detailed dataset signal amounts during the CNN operations for preprocessed train, preprocessed test, preprocessed & mixed test, and preprocessed & augmented test are respectively 75722, 27330, 79421, 26595, 41560, and 12920. Thus, we achieve better results in case considering the lack of literature and real-life needs for ECG Heartbeat Classification.

LITERATURE COMPARISON

Literature	Year	Info	Acc. [Set A Set B]
This Study	2022	CNN based Heterogeneous Harmonization • Robust Augmentation Ability • Robust Accuracy • Robust Detection • Main advantage of this study is the massive ability to adapt systems in case of platform and environment independent input heterogeneity. • For example, average literature study accuracy will down about %56 (as seen Table IV) in case of missing robust augmentation ability. Therefore, other results and accuracies are not important in lack of robust augmentation for heterogeneous environments. This is the reason of importance of proposed study for real life implementations. The proposed study is stronger than other proposals even accuracy comparison seems not best.	93.5% 97.1%
Acharya et al. [11]	2017	CNN (10-Fold) • Complex Model • Poor Data Distribution • Long Training Hourse	93.47% 94.03%
Lie et al. [12]	2021	Attention-Based CNN • Poor Data Distribution	98.57%
Schwab et al.[13]	2017	Recurrent Neural Networks • Poor Accuracy	90.0% 79.0%
Zhang et al.[14]	2021	Encoder based Adversarial CNN • Small Subject Number	91.5% 95.3%
Khan et al.[15]	2021	CNN • Poor Data Processing	95.2%
Hua et al.[16]	2020	CNN • Lack of Augmentation	97.02%
Kachuee et al. [17]	2018	Deep Transferable Representation • Lack of Augmentation	93.4% 95.9%
Zhu et al. [18]	2021	Extreme Gradient Boosting • Poor Data Distribution	98.68% 99.14%
Chen et al.[19]	2017	QRS Detection • Traditional Signal Processing Method • Poor Data Distribution • Lack of Segmentation • Lack of Augmentation	99.82% 99.81%
Mathunjwa et al. [20]	2022	Deep Residual CNN (5-Fold) • Complex Architecture • 2D Approach • Bad for Hardware Limitations • Poor Data Distrubution	98.05% 95.96%
Kedawat et al.[21]	2011	Fast Fourier Transform • Lack of Augmentation • Limited Data Distribution	98.48%
Ye et al. [22]	2012	Wavelet Transform Analysis • Lack of Augmentation • Existing Complex Class Spectrum	99.3% 99.7%
Kumari et al.[23]	2015	FPGA based Signal Processing • Poor Data Distribution	94.76%

Fig. 4: t-SNE visualization of the learned representation: (a) samples from MIT-BIH for ECG beat classification (b) samples from PTB dataset for MI classification. Labels for each task are indicated with colors (best viewed in color).

FIGURE 2: The proposed overall process of arrhythmia classification.

REFERENCES

[1]Dataset:
https://physionet.org/content/mitdb/1.0.0/

[2]Dataset Dictionary:
https://archive.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm

[3]Visualization of Data Records:
https://physionet.org/lightwave/?db=mitdb/1.0.0

[4]Mathunjwa, B. M., Lin, Y. T., Lin, C. H., Abbod, M. F., Sadrawi, M., & Shieh, J. S. (2022). ECG Recurrence Plot-Based Arrhythmia Classification Using Two-Dimensional Deep Residual CNN Features. Sensors, 22(4), 1660.

[5]Gothwal, H., Kedawat, S., & Kumar, R. (2011). Cardiac arrhythmias detection in an ECG beat signal using fast fourier transform and artificial neural network. Journal of Biomedical Science and Engineering, 4(04), 289.

[6]Kumari, L. R., Sai, Y. P., Balaji, N., & Viswada, K. (2015). FPGA based arrhythmia detection. *Procedia Computer Science*, *57*, 970-979.

[7] Isin, A., & Ozdalili, S. (2017). Cardiac arrhythmia detection using deep learning. Procedia computer science, 120, 268-275.

[8] Alfaras, M., Soriano, M. C., & Ortín, S. (2019). A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection. Frontiers in Physics, 7, 103.

[9]G. D. Flora and M. K. Nayak, "A Brief Review of Cardiovascular Diseases, Associated Risk Factors and Current Treatment Regimes", 2019. Current pharmaceutical design, 25(38), 4063–4084. https://doi.org/10.2174/1381612825666190925163827

[10]World Health Organization, "Cardiovascular diseases(CVDs)", 2017. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [Accessed: Mar 10, 2022].

[11] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats", 2017. Computers in biology and medicine, 89, 389–396. https://doi.org/10.1016/j.compbiomed.2017.08.022

[12] Z. Liu, and X. Zhang, "ECG-Based Heart Arrhythmia Diagnosis Through Attentional Convolutional Neural Networks", 2021. IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS), 156-162.

REFERENCES

[13] P. Schwab, G. Scebba, J. Zhang, M. Delai, and W. Karlen, “Beat by beat: Classifying cardiac arrhythmias with recurrent neural networks”, 2017. Computing in Cardiology (CinC), 1-4.

[14] J. Zhang, A. Liu, D. Liang, X. Chen, and M. Gao, “Interpatient ECG Heartbeat Classification with an Adversarial Convolutional Neural Network”, 2021. Journal of healthcare engineering, 2021, 9946596. https://doi.org/10.1155/2021/9946596

[15] M.M.K. Rahman, M.A.S. Bakr, S. Sakib, A. Aziz, A.K. Tanzeem, and Z. Hossain, “Electrocardiogram Heartbeat Classification Using Convolutional Neural Networks for the Detection of Cardiac Arrhythmia”, 2020. Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 915-920.

[16] X. Hua, J. Han, C. Zhao, H. Tang, Z. He, J. Tang, Q. Chen, S. Tang, and W. Zhou, “A novel method for ECG signal classification via one-dimensional convolutional neural network”, 2020. arXiv: Signal Processing.

[17] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, “ECG Heartbeat Classification: A Deep Transferable Representation”, 2018. IEEE International Conference on Healthcare Informatics (ICHI), 443-444.

[18] H. Zhu, Y. Zhao, Y. Pan, H. Xie, F. Wu, and R. Huan, “Robust Heartbeat Classification for Wearable Single-Lead ECG via Extreme Gradient Boosting”, 2021. Sensors (Basel, Switzerland), 21.

[19] C. L. Chen and C. T. Chuang, “A QRS Detection and R Point Recognition Method for Wearable Single-Lead ECG Devices”, 2017. Sensors (Basel, Switzerland), 17(9), 1969. https://doi.org/10.3390/s1709196

[20] Mathunjwa, B. M., Lin, Y. T., Lin, C. H., Abbod, M. F., Sadrawi, M., & Shieh, J. S. (2022). ECG Recurrence Plot-Based Arrhythmia Classification Using Two-Dimensional Deep Residual CNN Features. Sensors, 22(4), 1660.

[21] Gothwal, H., Kedawat, S., & Kumar, R. (2011). Cardiac arrhythmias detection in an ECG beat signal using fast fourier transform and artificial neural network. Journal of Biomedical Science and Engineering, 4(04), 289.

[22] Ye, C., Kumar, B. V., & Coimbra, M. T. (2012). Heartbeat classification using morphological and dynamic features of ECG signals. IEEE Transactions on Biomedical Engineering, 59(10), 2930-2941.

[23] Kumari, L. R., Sai, Y. P., Balaji, N., & Viswada, K. (2015). FPGA based arrhythmia detection. Procedia Computer Science, 57, 970-979.

[24] Two-Tailed Z Score : https://keydifferences.com/difference-between-one-tailed-and-two-tailed-test.html

QUESTIONS

- **GoTerms**
  - How can I reach GOTerms in order to understand Cardiovascular Arrhytmia process in more detail?
    - Search Results are around :
      - Heart Development Gene Ontology Term
      - Heart Contraction Gene Ontology Term

- **Life-Threatening vs Non-Life Threatening**
  - MIT-BIH Non-Life-Threatening is focusus for abnormalities. Life-Threatening is focuses mortal cases but unlabeled. How can I apply the label for this dataset. By auditivly, cooperatively or research-oriented?

- **Finding Richer Dataset**
  - Collection of more detailed data is costly. How can I achieve more quality data? How can I analysis real-life cases and external effects like corruption, time shifting factors, genetic effects, time phase divergent?

THANK YOU