



ECG를 통한 Feature Ensemble 기반 Wolf Parkinson White Syndrome classification



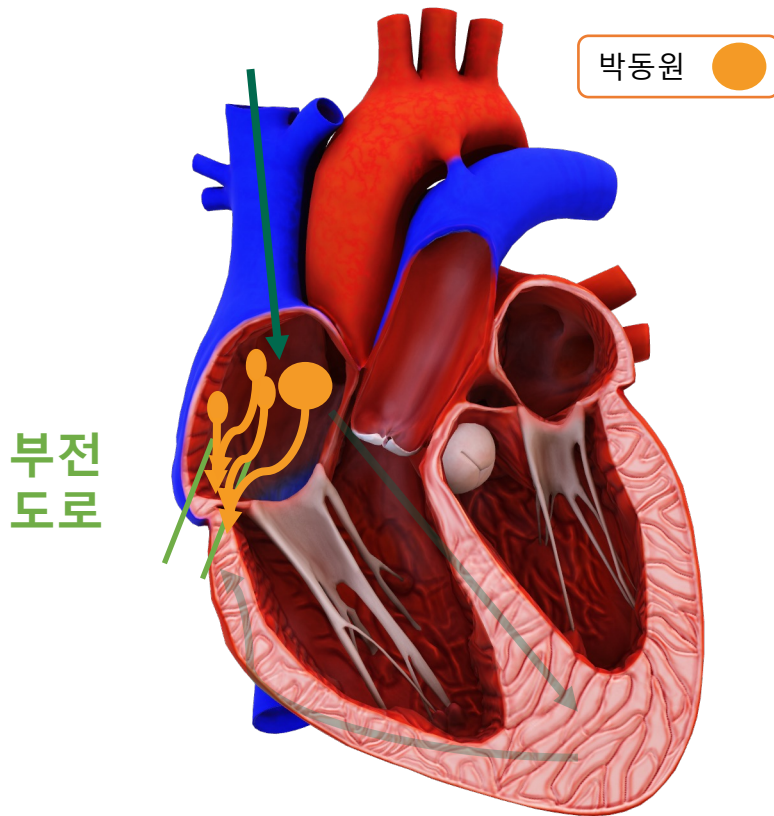
Algorithmic Machine Intelligence Lab.

**한국교통대학교 소프트웨어학과
AMI LAB
학부 연구생 오규태**

- 1 Introduction**
| Research purpose
- 2 Proposal Method**
| Feature Ensemble
- 3 Experiments**
| Experiments and Result
- 4 Conclusion**
| Conclusion
- 5 On Going Research**
| On Going Research

1. Introduction

Research purpose



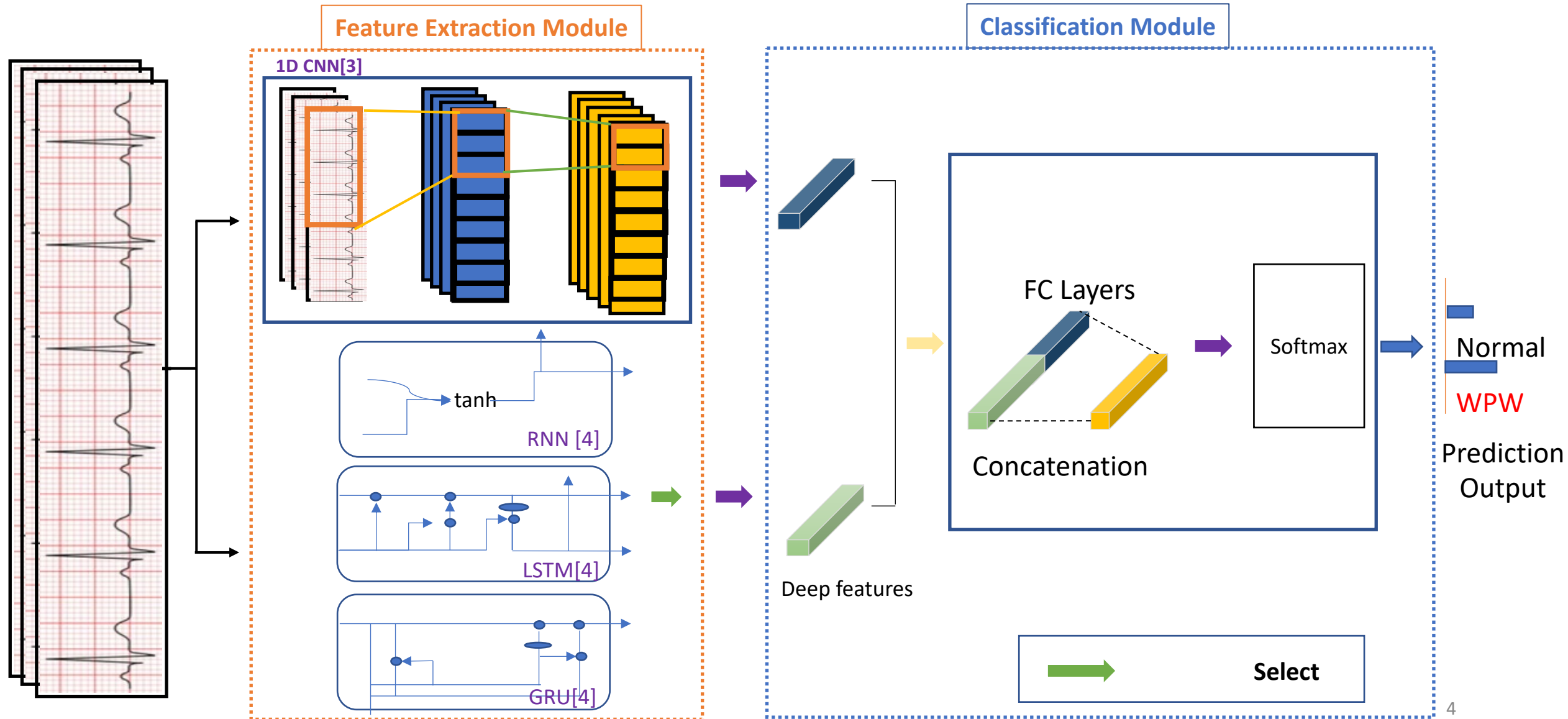
Wolf Parkinson White Syndrome(WPW)는 일반인과는 다르게 선천적으로 심방과 심실 사이에 **부전도로(Accessory Pathway)**가 존재하여 정상 전도와 비교하였을 때, 빠른 속도로 심실을 자극하여 부정맥을 일으키는 것을 의미하며, **심실세동**과 함께 발생하면 **생명에 위협**이 될 수 있을 정도의 빠른 심박수를 유발하여 **심인성 급사**를 유발한다.

부정맥이 주된 증상이기는 하나, 평소에는 무증상인 경우가 많고, 성인이 되어 갑작스럽게 발생하는 경우가 있어 인지하지 못하고 살아가는 환자들이 많다.[1,2]

이러한 특징은 갑작스러운 건강 악화가 타인의 목숨에 악영향을 줄 수 있는 트럭 운전기사나 의사와 같은 직업군 등의 경우 큰 문제가 될 수 있기에 조기에 발견하고 치료하는것이 중요하다.

따라서 본 논문에서는 WPW를 정확하게 분류할 수 있도록 Feature Ensemble 기반 기법을 제안한다.

2. Proposal Method

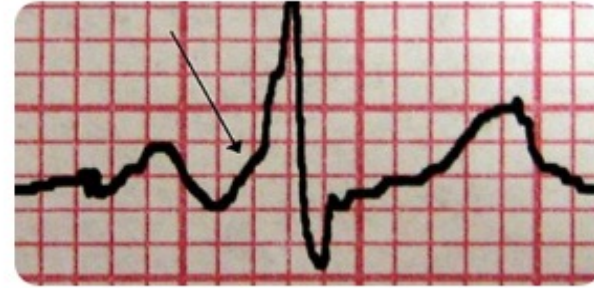


3. Experiments (Dataset / Kaggle)

Wolff-Parkinson-White ECG database

A database containing ECGs from patients with Wolff-Parkinson-White syndrome

→ 158개 존재[5]



<https://www.kaggle.com/datasets/bjoernjostein/wolfparkinsonwhitedata>

Normal-Sinus-Dataset

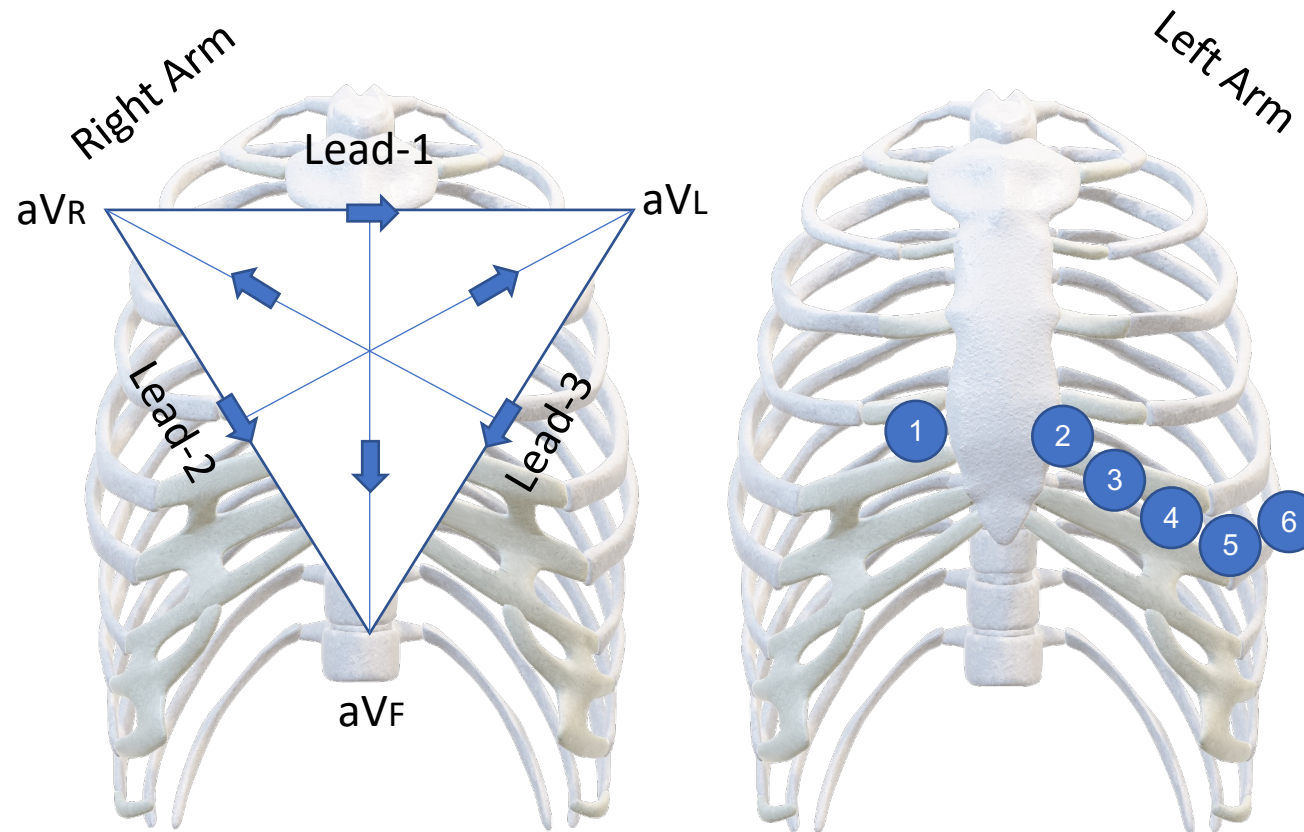
A dataset containing 15090 ECGs from patients with normal sinus rhythm

→ 15080개 존재[6]



<https://www.kaggle.com/datasets/bjoernjostein/normalsinusdataset>

3. Experiments (Dataset 12Lead[8])



3. Experiments (Dataset Noise)

시계열 데이터베이스의 상관관계 보존 노이즈 생성

시계열 데이터베이스의 상관관계
보존 노이즈 생성
(Correlation-aware Noise Generation on
Time-Series Databases)

홍 선 경 [†]
(Sun-Kyong Hong)

홍 준 호 ^{**}
(Junho Hong)

문 양 세 ^{***}
(Yang-Sae Moon)

요약 본 논문에서는 시계열 데이터에 대한 랜덤 노이즈 생성 시, 프라이버시뿐 아니라 상관관계를 보존하는 새로운 접근법을 제안한다. 민감한 시계열 데이터의 프라이버시 보호를 위해, 기존의 랜덤 노이즈 추가 기법은 균등(uniform) 혹은 가우시안(Gaussian) 분포를 기반으로 생성한 노이즈를 원본 데이터에 추가하여 제3자(데이터 마이너)에 공개한다. 그러나 이러한 랜덤 노이즈 추가 기법은 시계열 간 상관관계를 왜곡하거나 아예 없앨 수 있다는 문제점이 있다. 시계열 간 상관관계는 보다 정확한 마이닝 결과를 얻기 위해 매우 중요한 척도이다. 본 논문에서는 상관관계 보호를 위해 시계열의 원본 엔트리와 이에 대응하는 노이즈 엔트리의 부호를 일치시키는 간단하면서도 효율적인 전략을 제시하고, 이를 사용하는 상관관계 보존 노이즈 생성(correlation-aware noise generation)의 새로운 접근법을 제안한다. 엔트리들의 부호는 상관관계에 큰 영향을 미치기 때문에, 동일한 부호를 유지하는 것이 상관관계 보존에 매우 중요한 역할을 한다. 본 논문에서는 실제 데이터 대상의 실험을 통해, 시계열 상관관계를 보존하는 관점에서 제안한 기법의 우수성을 보인다.

키워드: 시계열 데이터, 상관관계, 프라이버시, 랜덤 노이즈 추가, 노이즈 생성

Abstract This paper proposes a novel approach that preserves *correlation* as well as *privacy* in generating random noise to time-series data. To preserve privacy of sensitive time-series data, the existing techniques generate random noise data based on uniform or Gaussian distributions and add them to the original data. These additive random noise techniques, however, have a critical problem of destructing (i.e., not preserving) the correlation among time-series. Correlation among time-series is very important to acquire the more accurate mining results. To preserve the correlation, in this paper we propose a *correlation-aware noise generation* approach, which is a simple but effective approach that generates a noise value to have the same sign with its corresponding original entry. This is because the signs of entries make a large influence on the correlation measure, and thus, preserving the same signs is important to improve the correlation preservation. We empirically show the superiority of the proposed approach in the viewpoint of preserving the time-series correlation.

Keywords: time-series data, correlation, privacy, additive random noise, noise generation

319

시계열 데이터 베이스의 상관 관계 보존한 노이즈 생성하기 위해 본 논문의 방법을 가져와 사용함.

CaNG는 평균은 0, 표준편차가 시그마인 가우시안 분포를 통해 임의의 값을 생성하고 각 기법의 검증을 거쳐 최종 Noise를 선정함.[7]

CaNG2 기법과 CaNG3 기법을 통해 데이터를 생성해 결과를 확인함.

CaNG2 기법

$$x_i \cdot n_i > 0$$

- x_i 는 기본 값이며, n_i 는 노이즈이다.
- x_i 와 n_i 를 비교하여 동일한 부호를 가진다면 선택 아니라면 n_i 새로 생성

CaNG3 기법

$$x_i \cdot n_i > 0$$

- x_i 는 기본 값이며, n_i 는 노이즈이다.
- x_i 와 n_i 를 비교하여 동일한 부호를 가진다면 선택 아니라면 $-1 * n_i$ 으로 채택

3. Experiments (Dataset Setting)

Noise 추가 데이터셋 사용

- 12 LEAD 모두 사용
- Time Step : 250[8] / Sample Rate : 500Hz[9]

<None Noise>

Normal : Abnormal (1 : 1)

Train : Normal 80 + Abnormal 79 = 159개

Validation : Normal 20 + Abnormal 19 = 39개

Test : Normal 60 + Abnormal 60 = 120개

<add Noise>

Normal : Abnormal (1 : 1)

Train : Normal 80 + Abnormal 79 = 159개

Validation : Normal 20 + Abnormal 19 = 39개

Test : Normal 0 + Abnormal 0 = 0개

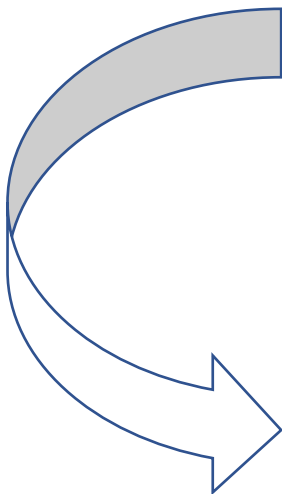
<최종 데이터셋>

Normal : Abnormal (1 : 1)

Train : Normal 160 + Abnormal 158 = 318개

Validation : Normal 40 + Abnormal 38 = 78개

Test : Normal 60 + Abnormal 60 = 120개



3. Experiments (Basic Model 세팅)

Model	기본 설정 값[8]
1D-CNN(Kernel size)	15 -> 10 -> 5
1D-CNN(Channel)	128 - > 256 -> 512
(BI)LSTM(Channel)	128 - > 256 -> 512
(BI)GRU(Channel)	128 - > 256 -> 512

3. Experiments (하드웨어 세팅)

Category	
OS	Ubuntu 18.04
CPU	i9-12900Ks
RAM	128GB
GPU	RTX 3090TI
CUDA	11.4
Framework	Tensorflow-GPU 2.8

3. Experiments (CaNG2 Result)

Model	Precision	Recall	F1-Score	Accuracy
1D-CNN	0.92500	0.92797	0.92487	0.92500
1D-CNN + GAP	0.87500	0.87594	0.87492	0.87500
LSTM	0.65000	0.65864	0.64517	0.65000
Bi-LSTM	0.69167	0.71303	0.68374	0.69167
GRU	0.67500	0.67622	0.67443	0.67500
Bi-GRU	0.60833	0.62699	0.59340	0.60833

빨강 : Best
파랑 : Worst

3. Experiments (CaNG2 Result)

Model	Precision	Recall	F1-Score	Accuracy
Bi-GRU + 1D-CNN	0.92500	0.92607	0.92495	0.92500
GRU +1D-CNN	0.94167	0.94476	0.94157	0.94167
GRU +LSTM	0.72500	0.73608	0.72173	0.72500
1D-CNN +LSTM	0.94167	0.94179	0.94166	0.94167
1D-CNN +LSTM +GAP	0.83333	0.83333	0.83333	0.83333
Bi-LSTM +1D-CNN	0.93333	0.93771	0.93317	0.93333
Bi-LSTM + Bi-GRU	0.67500	0.69943	0.66473	0.67500

3. Experiments (CaNG3 Result)

Model	Precision	Recall	F1-Score	Accuracy
1D-CNN	0.925000	0.92797	0.92487	0.92500
1D-CNN + GAP	0.87500	0.87594	0.87492	0.87500
LSTM	0.55833	0.56838	0.54149	0.55833
Bi-LSTM	0.67500	0.69943	0.66473	0.67500
GRU	0.65000	0.65017	0.64990	0.65000
Bi-GRU	0.60833	0.62699	0.59340	0.60833

빨강 : Best
파랑 : Worst

3. Experiments (CaNG3 Result)

Model	Precision	Recall	F1-Score	Accuracy
Bi-GRU + 1D-CNN	0.92500	0.92607	0.92495	0.92500
GRU +1D-CNN	0.94167	0.94476	0.94157	0.94167
GRU +LSTM	0.72500	0.73608	0.72173	0.72500
1D-CNN +LSTM	0.94167	0.94179	0.94166	0.94116
1D-CNN +LSTM +GAP	0.82500	0.82500	0.82499	0.82500
Bi-LSTM +1D-CNN	0.94167	0.94476	0.94157	0.94167
Bi-LSTM + Bi-GRU	0.68500	0.69943	0.66473	0.67500

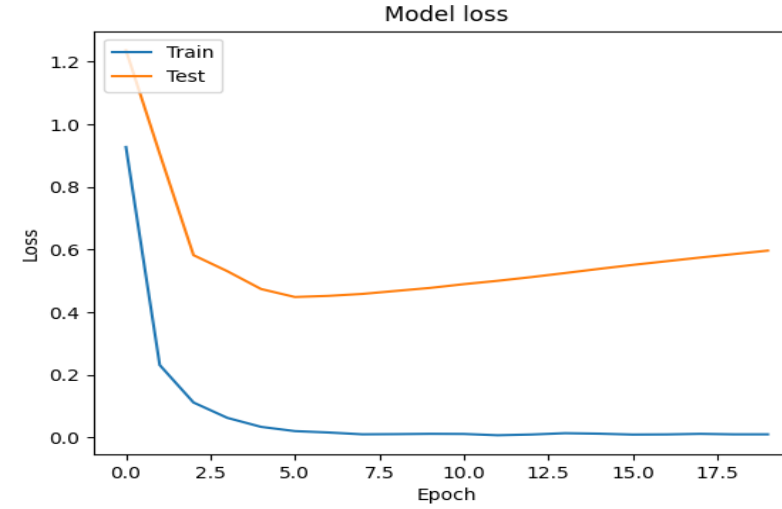
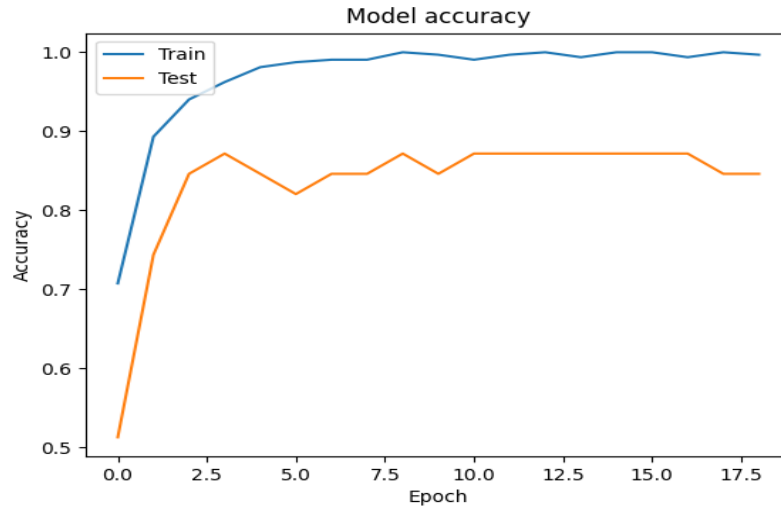
3. Experiments (CaNG2 VS Cang3 Result)

Model	Precision	Recall	F1-Score	Accuracy
Bi-GRU + 1D-CNN	0.92500	0.92607	0.92495	0.92500
GRU +1D-CNN	0.94167	0.94476	0.94157	0.94167
GRU +LSTM	0.72500	0.73608	0.72173	0.72500
1D-CNN +LSTM	0.94167	0.94179	0.94166	0.94167
1D-CNN +LSTM +GAP	0.83333	0.83333	0.83333	0.83333
Bi-LSTM +1D-CNN	0.93333	0.93771	0.93317	0.93333
Bi-LSTM + Bi-GRU	0.67500	0.69943	0.66473	0.67500

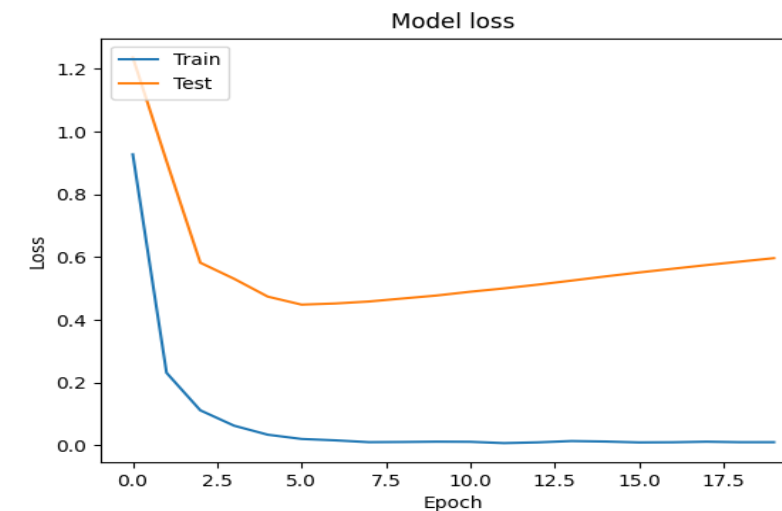
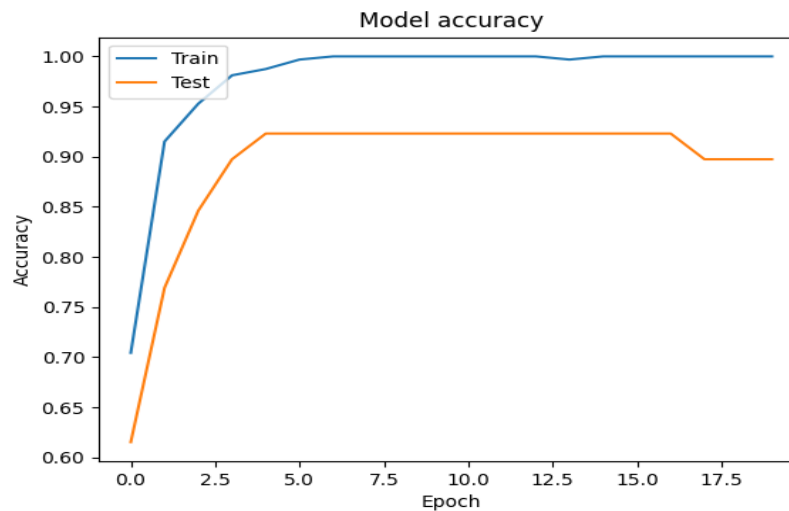
Model	Precision	Recall	F1-Score	Accuracy
Bi-GRU + 1D-CNN	0.92500	0.92607	0.92495	0.92500
GRU +1D-CNN	0.94167	0.94476	0.94157	0.94167
GRU +LSTM	0.72500	0.73608	0.72173	0.72500
1D-CNN +LSTM	0.94167	0.94179	0.94166	0.94117
1D-CNN +LSTM +GAP	0.82500	0.82500	0.82499	0.82500
Bi-LSTM +1D-CNN	0.94167	0.94476	0.94157	0.94167
Bi-LSTM + Bi-GRU	0.68500	0.69943	0.66473	0.67500

3. Experiments (Final Result)

1D - CNN

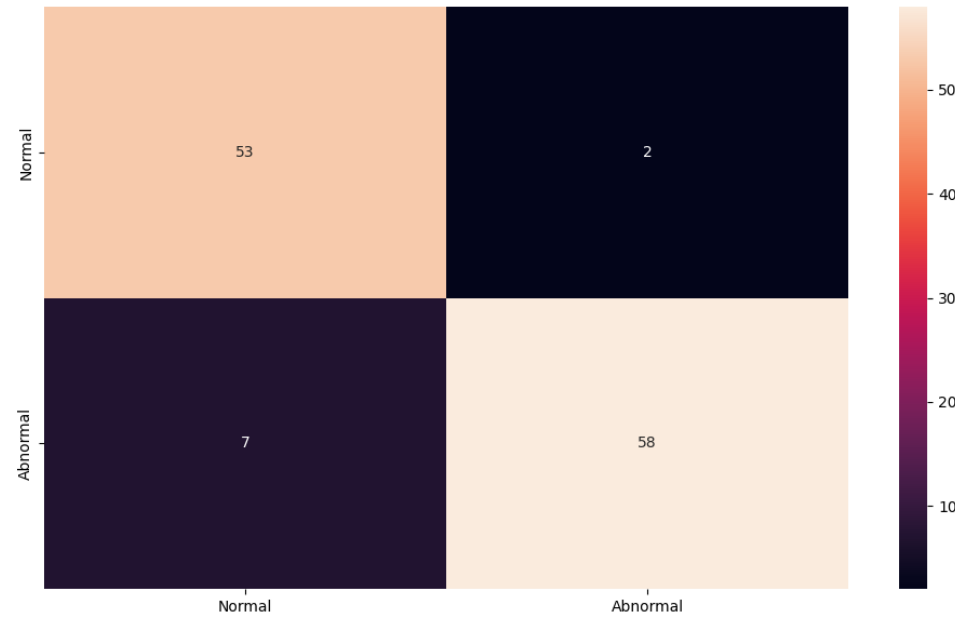


1D - CNN + LSTM

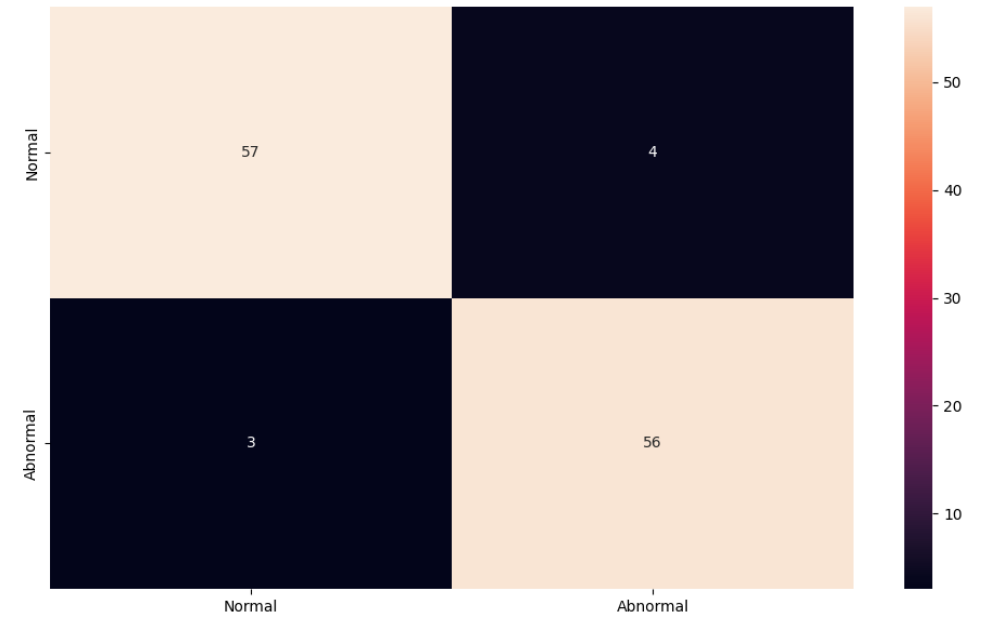


3. Experiments (Final Result)

1D - CNN



1D - CNN + LSTM



4. Conclusion

No	정리
1	적은 데이터셋을 증강하여 더 좋은 결과를 얻어냄.
2	데이터 증강을 위하여 적용한 기법 중 Cang3기법보다 CaNG2 기법을 적용하여 데이터 증강을 하였을때 더 좋은 결과를 얻음.
3	Feature ensemble시 단일 모델 대비 약 1.6~2.5%의 F-1 Score 향상

5. On Going Research

No	정리
1	Time Step / Sample Rate 최적 값 찾기
2	하이퍼파라미터 전반적인 수정
3	다른 순환 신경망 및 WPW 분류에 적합한 기법을 찾아 Feature Ensemble

References

- [1] Asan Medical Center, <https://www.amc.seoul.kr>
- [2] C. Lee, and S.R Oh, "Paroxysmal Supraventricular Tachycardia in a Patient with Wolff-Parkinson-White Syndrome Induced by Central Venous Cannulation and Surgical Stimuli during Operation," Korean Journal of Anesthesiology, Vol. 48, No. 3, pp. 308-310, 2005
- [3] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D.J Inman , "1D convolutional neural networks and applications:A survey," Mechanical systems and signal processing, Vol. 151, No. 107398, 2019
- [4] J,Y CH, C. Gulcehre, K.H Cho, Y.SH Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," arXiv, Vol. 1412, No. 3555, 2014
- [5]kaggle,<https://www.kaggle.com/datasets/bjoernjostein/wolfparkinsonwhitedata>
- [6]kaggle,<https://www.kaggle.com/datasets/bjoernjostein/normalsinusdataset>
- [7] S.K Hong, J.H Hong, Y.S Mong, "Correlation-aware Noise Generation on Time-Series Databased," 정보과학논문지: 데이터베이스, Vol. 40, No. 5, pp. 319-327, 2013
- [8]kaggle,<https://www.kaggle.com/code/bjoernjostein/wpw-detection-from-ecg-using-1d-cnn>
- [9] O.H Kwon, J.W Jeong, H.B Kim, I.H Kwon, S.Y Park, J.E Kim and Y.R Choi, "Electrocardiogram Sampling Frequency Range Acceptable for Heart rate Variability Analysis," Healthcare Informatics Research, Vol. 24, No. 3, pp. 198-206, 2018
- [10] B.C Jang, M.H Kim, G. Harerimana, S.U Kang, and J.W Kim, "Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism," Applied Sciences, Vol. 10, No. 17, 2020