

딥 러닝 알고리즘을 이용한 중환자실 내  
위급한 부정맥 증상 탐지에 관한 연구

The research on detecting life threatening  
arrhythmias in the intensive care unit using  
deep learning algorithm

산업 및 시스템공학과

연구책임자 신 하 용

참여연구원 박 희 환

K A I S T

## 제 출 문

KAIST 총장 귀하

본 보고서를 “딥 러닝 알고리즘을 이용한 중환자실 내 위급한 부정맥 증상 탐지” 과제의 보고서로 제출합니다.

2015.12.28.

주관연구부서명 : 한국과학기술원

주관연구책임자 : 신 하 용

연 구 원 : 박 희 환

## 보고서 초록

과제관리번호 (계정번호)	N01150659		총 연구기간		5 개월 (2015. 7. 13. ~ 2015. 12. 11.)		
연구사업명	중 사 업 명		글로벌교육연구선도사업				
	세부사업명		석박사 모험연구 프로그램 운영				
연구과제명	중 과 제 명						
	세부(단위)과제명		딥 러닝 알고리즘을 이용한 중환자실 내 위급한 부정맥 증상 탐지에 관한 연구				
연구책임자	신 하 용	총 참여연구원수	총 : 1 명 내부 :   명 외부 :   명	총 연구비	정부: 7000 천원 기업:   천원 계: 7000 천원		
연구부서명			참여기업명				
국제공동연구	상대국명 :		상대국연구기관명 :				
위 탁 연 구	연구기관명 :		연구책임자 :				
요약(연구결과를 중심으로 개조식 500 자 이내)					보고서 면수	38	
<p>중환자 의료에서 환자의 상태를 정확히 알려주는 경보 시스템은 매우 중요하다. 그러나 무려 86%에 달하는 경우가 보고되었을 정도로 높은 오경보율은 중환자의 회복과 의료진의 적절한 처치를 방해하는 방해물이 되어왔다. 특히 중환자의 생명까지 좌우할 수 있는 치명적인 부정맥 증상에 대한 오정보는 중대한 결과를 유발할 수 있을 정도로 중요하기에, 이를 줄이기 위해선 무엇보다 심전도에서 부정맥 증상을 정확히 탐지해내야 한다. 이 연구에서 우리는 이 문제의 해결책으로 지난 10년 간 이미지 인식이나 음성 인식과 같은 패턴 인식 문제에 대해서 놀랄만한 성취를 이루어낸 딥 러닝 방법론을 고려한다. ECG에서 부정맥 증상을 탐지하기 위해, 우리는 새롭게 개발한 두 종류의 심층 신경망 구조와 convolutional neural networks를 이 연구에서 제안하며, 실험 결과 우리가 제안한 딥 러닝 알고리즘 중에서도 특히 convolutional neural networks의 경우 민감도, 특이도, 그리고 정확도가 각각 89.47%, 88.03%, 88.67%를 달성하여, 우리의 알고리즘이 충분히 경쟁력이 있음을 보여주었다.</p>							
색 인 어 (키워드 / 각 5개 이상)	한 글	부정맥, 심전도, 딥 러닝, 심층 신경망, Convolutional neural network					
	영 어	Life-threatening arrhythmias, ECG, Deep learning, Deep neural network, Convolutional neural network.					

## 요 약 문

### I. 제 목

딥 러닝 알고리즘을 이용한 중환자실 내 위급한 부정맥 증상 탐지

### II. 연구개발의 목적 및 필요성

현재의 중환자실 내 경보 시스템의 높은 오경보율은 중환자의 회복을 방해할 뿐만 아니라 의료진이 정말 심각한 경보를 놓치는 결과를 낳고 있다. 특히 위급한 부정맥 증상에 대해 이런 오경보는 치명적인 결과를 낳을 수도 있으므로, 중환자에게 발생하는 위급한 부정맥을 정확하게 탐지할 수 있는 새로운 방법이 필요하다.

### III. 연구개발의 내용 및 범위

심장 활동을 도표로 나타내는 심전도로부터 위급한 부정맥 증상을 찾기 위해 이 연구에서는 딥 러닝 기술을 활용하였다. 기본적인 딥 러닝 알고리즘인 심층 신경망(Deep neural networks)를 먼저 적용해보고, 심층 신경망의 성능을 올리기 위해 ECG 분석에 적합한 추가 구조를 제안하였다. 그리고 마지막으로 이미지 인식 등에서 높은 성능을 보이는 convolutional neural networks를 이 문제에 적용하였다.

### IV. 연구개발결과

추가구조를 도입한 심층 신경망과 convolutional neural network 알고리즘의 경우 기존의 경보 시스템보다 높은 정확도를 보여주었으며, 특히 convolutional neural network의 경우 민감도, 특이도, 그리고 정확도가 각각 89.47%, 88.03%, 88.67%를 달성하였다.

### V. 연구개발결과의 활용계획

본 연구는 딥 러닝 기술로 생체 신호를 분석하는 후속 연구들에 대한 시작점으로 활용될 수 있다. 먼저 Full bayesian learning 등을 이용해 정확도를 더욱 향상시킨 후 2016 PhysioNet/CinC Challenge에 대한 기반 연구로 활용할 계획이며, 이후 웨어러블 기기에서 수집되는 생체 신호 분석 연구로 확장할 수 있을 것이다.

## S U M M A R Y

### I. Title

The research on detecting life threatening arrhythmias in the intensive care unit using deep learning algorithm

### II. The purpose and necessity of the research

Alarms in the intensive care unit play a major role in critical care. However, patients and medical staffs have suffered from high false alarm rates, which was high as 86% in one report. These false alarms can be more critical for patients with arrhythmia. To reduce them, it is essential to detect arrhythmias in electrocardiogram (ECG) signals accurately.

### III. Contents of the research

we consider deep learning methods to solve this problem. Deep learning, which consists of deep and complex neural networks architectures, has applied successfully in pattern recognition problems such as image recognition and speech recognition, and has made major advances in recent 10 years. We suggest two novel deep neural networks architectures and a convolutional neural network to find life-threatening arrhythmias from critical patients' ECG signals.

### IV. Results of the research

Experimental results show that the convolutional neural network yielded a sensitivity, specificity, and accuracy of 89.47%, 88.03%, and 88.67% in life-threatening arrhythmia detection.

### V. Application plan of the research

This research can be a starting point for analyzing biological signal using deep learning approach. First, it will be a stepping stone for 2016 PhysioNet/CinC challenge by applying full bayesian learning to the result. In addition, we consider to advance this research to analyze biological signal from wearable instrument and platform.

## C O N T E N T S

1. The purpose and necessity of the research
  - 1.1 The necessity of the research
  - 1.2 The purpose of the research
2. The current method for detecting arrhythmias and its limitation
3. Contents and results of the research
  - 3.1 Deep neural networks
  - 3.2 Additional architecture for deep neural networks
  - 3.3 Convolutional neural networks
  - 3.4 Data and experimental setting
  - 3.5 Experiment result
4. Achievements and contributions of the research
  - 4.1 Result evaluation by comparing with 2015 PhysioNet/CinC challenge
  - 4.2 Contributions of the research
5. Application plans
6. Reference

## 목 차

### 제 1 장 연구개발과제의 개요

#### 제 1 절 연구개발의 필요성

#### 제 2 절 연구개발의 목적

### 제 2 장 국내외 기술개발 현황

#### 제 1 절 심전도(Electrocardiogram)

#### 제 2 절 위급한 부정맥 증상에 대한 현행 기준

#### 제 3 절 현재의 부정맥 탐지 방법과 그 한계

### 제 3 장 연구개발수행 내용 및 결과

#### 제 1 절 심층 신경망(Deep neural networks)

#### 제 2 절 심층 신경망을 위한 추가 구조

#### 제 3 절 Convolutional neural networks

#### 제 4 절 데이터 및 실험 설계

#### 제 5 절 결과

### 제 4 장 목표달성도 및 관련분야에의 기여도

#### 제 1 절 2015 PhysioNet/CinC challenge 결과와의 비교를 통한 목표 달성 평가

#### 제 2 절 중환자 생체신호 분석에의 기여

### 제 5 장 연구개발결과의 활용계획

### 제 6 장 참고문헌

## 제 1 장 연구개발과제의 개요

### 제 1 절 연구개발의 필요성

오늘날 중환자실 환자 모니터링 시스템이 만드는 경보의 상당수는 실상 의료적으로 긴급하지 않은 경보들이다. 과거부터 지금까지 많은 연구가 이러한 높은 오경보율 문제를 지적하였는데, 초기에 이 문제를 지적한 Lawless(1994)는 오경보율이 무려 86%에 달한다는 보고가 존재하고, 6%~40%의 경보가 맞는 경보이긴 했지만 실제로 그리 시급하지 않은 상황에 대한 경보였다는 점을 지적했다. Chambrin 등(1999)은 조사 결과 전체 경보 중 무려 72%가 어떤 의료적 행위도 이끌어 내지 못했음을 보여주었고, 후속 연구에서 Chambrin (2001)은 소아 및 성인 중환자 치료를 대상으로 한 여러 연구를 검토한 결과 오직 10%도 되지 않는 경보만이 실제로 의료적 처치를 이끌어 냈다는 걸 발견하였다. Tsien(1997)은 고작 단 2% - 9% 만이 진정으로 시급한 대처를 요하는 경보였다고 보고하였다.

비록 동 연구에서 중환자를 모니터한 298시간 동안 단 하나의 false negative 경보도 발생하지 않았음을 Tsien이 함께 언급하긴 하였으나, 단 한번의 실수가 환자의 생명을 앗아가는 결과로 나타날 수 있음을 감안하고서라도 현재의 오경보율은 너무 높은 수준으로, Imhoff 등(2009)은 이 문제가 Lawless(1994)에 의해 제기된 이래 지금까지 거의 개선되지 않았음을 지적하고 있다.

무엇보다도 높은 오경보율은 환자를 보살피는 데 심대한 어려움을 끼친다. Chambrin(2001)에 따르면, 중환자실에서 발생하는 경보는 80dB 이상의 소음을 내며, 이는 환자와 의료진 모두의 수면을 방해하고 지속적인 스트레스를 강요하며, 그 결과 의료진의 경보에 대한 반응 속도가 느려질 뿐만 아니라 진정으로 심각한 경보를 의료진이 놓치는 결과까지 낳게 된다. 심지어 이러한 환경 속에서는 숙련된 간호사조차도 오직 38%의 위급한 경보를 인지했다는 연구까지 있다. (Cropp 등, 1994)

그렇다면 지금까지의 경보 시스템은 어떻게 동작하는가? 현재 중환자실에서 사용되는 경보 시스템은 환자로부터 얻어지는 생체 신호가 일정 기준을 넘을 때 발생한다. 즉, 경보 시스템은 중환자로부터 측정되는 여러 종류의 생체 신호



가 ‘정상’ 범주 내에 있는지를 바탕으로 경보를 울릴지 아닐지를 결정하는 것이다. 그러나 실제로는 환자의 생체 기능에 변화가 없는데도 불구하고 측정되는 생체 신호(파라미터)가 급격하게 변하는 일이 관찰되며, 이때 발생하는 경보가 바로 오경보(false alarm)이다. (Chambrin, 2001)

Imhoff와 Kuhls(2006)에 따르면, 오경보의 이유로는 크게 다음 세 가지를 들 수 있다.

- 기술적인 요인 때문에(예: 젖은 바닥) 실제로는 파라미터가 경보 기준을 넘지 않았는데도 기준을 넘은 것으로 인식하여 경보가 발생할 수 있다.
- 임상적으로, 파라미터가 실제로 기준을 넘어 경보가 발생하였으나, 그 상황이 의료적으로 그다지 심각하지 않은 상황일 수 있다.
- 의료진의 개입으로 인해 경보가 발생할 수 있다.

의료진의 개입과 기술적인 요인으로 인한 오경보가 의료진의 훈련과 기기의 올바른 설치 등으로 상대적으로 쉽게 해결할 수 있는데 반해, 임상적인 요인으로 인한 오경보는 해결하기가 매우 까다롭다. 환자로부터 실시간으로 얻어지는 파라미터를 분석하여 환자의 상태 변화를 정확히 탐지해내는 것이 선행되어야 하기 때문이다.

그러나 경보 시스템의 정확도를 높이기 위해서는 임상적인 요인으로 인한 오경보를 필히 줄여야 하며, 특히 중환자의 생명을 위태롭게 할 수 있는 부정맥 증상이 발생했을 때 이를 정확하게 탐지해내는 것은 환자의 생명과도 직결된 문제이다. 부정맥이란 심장이 정상적으로 뛰지 않는 것을 말하는데, 전 세계적으로 매년 약 6백만명이 부정맥의 일종인 심실빈맥증 때문에 일어난 급성심장사로 사망하며 급성심근경색이 일어났을 때의 생존률은 1%가 채 되지 않으므로 (Mehra, 2007), 부정맥이 발생한 시점에서 이를 정확히 감지하는 일은 매우 중요하다. 그러므로 **생명에 치명적인 부정맥 증상을 정확히 탐지해내는 것이야말로 중환자를 대상으로 하는 의료 행위의 질을 높이는 데 필수적**이라 할 것이다.

## 제 2 절 연구개발의 목적

본 연구에서는 이 문제를 해결하기 위해 딥 러닝(Deep learning) 기술을 적용할 것을 고려한다. 딥 러닝 기술이란 오늘날 각 분야에서 패턴 인식 문제를 풀기 위해 많이 시도하는 접근방법으로, 이미지 인식, 음성 인식 등 패턴 인식에 대해 기존에 존재하는 기계 학습 알고리즘보다 훨씬 높은 정확도를 보이고 있기 때문이다. 특히 딥 러닝 알고리즘은 기존의 다른 전문가가 직접 데이터를 처리하여 특징을 뽑아낼 필요 없이 데이터 자체를 바로 학습할 수 있어서 인과 관계를 인간이 바로 알기 어려운 상태에서의 분류(classification)를 잘 해내는 장점이 있는데(LeCun 등, 2015), 이러한 장점이야말로 실시간으로 들어오는 환자의 생체 신호를 분석해 정상(normal)과 비정상(abnormal)을 정확히 판단해야 하는 경보 알고리즘에 적합하다고 할 수 있다.

이 연구에서 우리는 **딥 러닝 기술을 활용하여 중환자로부터 모니터링 기기가 수집한 ECG 신호 데이터를 분석하고, 이를 통해 중환자에게 위급한 부정맥 증상이 발작하였는지 아닌지를 정확하게 감지하는 것을 그 목적으로 한다.**

## 제 2 장 국내외 기술개발 현황

### 제 1 절 심전도

부정맥을 포함하는 심장의 기능 이상을 진단하기 위해서 가장 많이 사용하고 가장 중요한 신호가 바로 심박동의 주기 중에 일어나는 심장의 전기적 활동 상태를 그래프 상에 나타낸 **심전도(Electrocardiogram, ECG)**이다.

심장은 스스로 전기를 만들고 그 전기 신호에 의해 움직인다. 간단하게 그 기작을 설명하면, 이완된 상태에서의 심근 세포를 살펴보면 세포막을 중심으로 세포 밖에는 양이온, 안에는 음이온이 차 있는데 이를 ‘분극’ 되어 있다고 한다. 그런데 전기 자극을 받게 되면 세포 안으로 양이온이 들어오면서 근육의 수축이 일어나는데 이를 ‘탈분극’ 이라고 한다. ECG 파형은 바로 이 전기 신호를 그린 것이므로, 심장의 움직임에 따라 구별되는 몇 가지 요소들로 구성되게 된다.

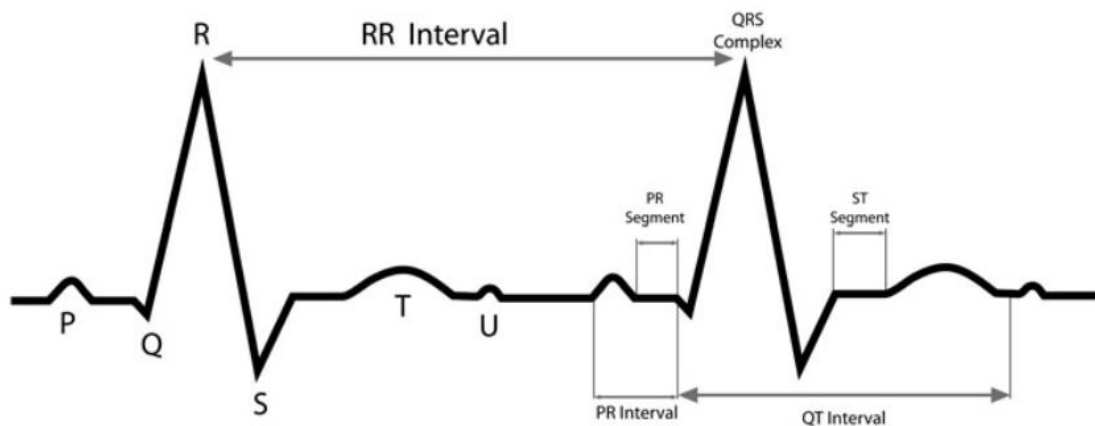


그림 1 The ECG waves, segments, and intervals. Gacek, A., & Pedrycz, W. (2012). ECG Signal Processing, Classification and Interpretation.

먼저, P-wave는 ECG에서 처음으로 굴절되는 부분으로, 심방의 탈분극으로 나타나는 파형이다. 심방의 재분극은 심실의 탈분극 동안 일어나며, QRS complex라 불리는 뾰족한 모양의 굴절은 이 심실의 탈분극에 해당한다. 뒤이어 나타나는 T-wave는 심실의 탈분극을 보여주며, 인구 전체의 약 1/4에게서는 U-wave가 나타난다. 그러므로 심장의 어느 부분에 이상이 발생하면, 이상

에 대응하는 ECG 파형의 요소들에서도 역시 문제를 발견할 수 있게 된다.

## 제 2 절 위급한 부정맥 증상에 대한 현행 기준

이 연구에서는 ECG를 이용하여 부정맥 중에서도 환자의 생명에 치명적인 증상인 **무수축(Asystole)**, **심한 서맥(Extreme bradycardia)**, **심한 빈맥(Extreme tachycardia)**, **심실 빈맥(Ventricular tachycardia)**, **심실 조동/세동(Ventricular flutter/fibrillation)** 다섯 가지 부정맥 증상(Aboukhalil, A et al, 2008)에 한정하여 논의를 진행한다.

- 무수축(Asystole)

간호학대사전에 따르면, 심장 무수축이란 어떤 일정한 시간, 심장이 수축하지 않는 상태를 말한다. 각종 심(心)블록이나 심실정지 등이 일어났을 대나 심실세동으로 효과적인 심수축이 일어나지 않을 때 생긴다. 심장 모니터링에 대한 미국 공인 기준인 ANSI/AAMI EC13:2002 에서는 ECG 상에서 4초 동안 QRS 파가 나타나지 않을 때 경보를 발하도록 규정하고 있다.

- 심한 서맥(Extreme bradycardia)

간호학대사전에 따르면, 맥박은 건강한 사람의 경우 성인은 매분 60~80, 유아 90~140, 노인 70~80이고 1분간 60이하를 서맥이라고 한다. ANSI/AAMI EC13:2002 에서는 심박수가 분당 40 이하이면 심한 서맥으로 판단하여 경보를 발하도록 규정하고 있다.

- 심한 빈맥(Extreme tachycardia)

간호학대사전에 따르면, 심박수는 보통 매분 60~80회 인데, 매분 100회를 넘을 경우 이것을 빈맥이라고 한다. 심박수가 분당 140 이상이면 심한 빈맥으로 판단하여 경보를 발하도록 규정하고 있다(ANSI/AAMI EC13:2002).

- 심실 빈맥(Ventricular tachycardia)

간호학대사전에 따르면 심실빈맥이 발작하면 심계항진을 느끼고 흉부압박감, 호흡촉진, 불안감을 호소하고 식은땀, 오심, 구토 등을 수반하는 수도 있다. 심박수는 매분 150~250이 된다. 심전도의 특징은 QRS의 폭이 넓고 심실성 기외수축이 연속해서 나타나는 형태를 취한다. ANSI/AAMI EC13:2002 기준으로 연속 5회 이상의 심실 수축이 심박수 100 이상으로 관측될 때 경보를 발한다.

● 심실 조동/세동(Ventricular flutter/fibrillation)

심실조동, 세동은 심실근의 일부가 불규칙하게 수축할 뿐이고 심장 전체의 수축은 볼 수 없는 상태이다. 심전도 상으로는 QRS, T파도 인정되지 않고 기선이 크고 작게 흔들리고 있을 뿐이고 이 흔들림이 비교적 규칙적으로 사인커브 모양의 것을 조동, 전혀 불규칙인 것을 세동이라고 하는데 임상적 의미는 조동도 세동도 같다(간호학대사전, 1996). 심전도 상에서 흔들리는 파형이 4초 이상 지속될 시 경보를 발한다(ANSI/AAMI EC13:2002).

이를 요약하면 다음 표와 같다.

경보 종류	경보 발생 기준
무수축	4초 동안 QRS 파가 관측되지 않음
심한 서맥	심박수가 분당 40 이하
심한 빈맥	심박수가 분당 140 이상
심실빈맥	심실수축이 연속 5회 이상, 심박수 100 이상으로 관측
심실조동/세동	흔들리는 파형이 4초 이상 지속

표 1 치명적인 부정맥 증상과 그 경보 발생 기준

### 제 3 절 현재의 부정맥 탐지 방법과 그 한계

ECG 정보에 기반한 부정맥 분석에 대해서는 세 종류의 접근법이 존재한다. (Clifford, 2006) 첫 번째는 가장 일반적인 접근법으로, 주어진 ECG에 QRS 탐지를 수행하고 거기서 얻은 각 맥박이 어떤 종류의 맥박인지 파악하여 그 정보를 기반으로 부정맥을 판단하는 것이다. 그러나 만약 QRS 탐지를 통해 맥박을 찾기 어려운 ECG에 대해선 정확도가 크게 떨어진다는 단점이 있다.

두 번째 접근법은 첫 번째에 대한 대안으로, 만약 ECG 상에서 QRS 탐지를 수행하기 어렵다면 대안으로 ECG의 일정 구간에 대해서 분산을 구하거나 Power-Frequency 분석을 수행하여 부정맥을 판단하는 것이나, 지금까지 이러한 접근 방법은 한정된 수의 데이터베이스에만 적용되었으므로 다양한 ECG 데이터를 통한 검증이 필요하다는 한계가 있다.

마지막으로, 부정맥 증상 시 ECG 리듬이 어떻게 움직일지에 대한 모델을 세우고 관측한 ECG 신호를 이 모델과 비교하는 방법이 있으나, 중환자로부터 얻은 ECG 신호의 경우 환자마다, 그리고 같은 환자라도 시간에 따라 급격히 다른 모습을 보여주는 것을 감안하면 이 문제에 대해 적합한 접근방법이라 하기 어렵다.

## 제 3 장 연구개발수행 내용 및 결과

딥 러닝은 낮은 단계의 표현(Representation)을 보다 높은 단계의 추상적인 표현으로 전이하는, 단순하지만 비선형적인 모듈을 여러 단계로 배치함으로써 각 단계의 표현을 학습하는 표현 학습(Representation-learning) 방법으로, 이러한 전이가 충분히 이루어지도록 구성하게 되면 딥 러닝 알고리즘은 매우 복잡한 함수라도 학습할 수 있게 된다(LeCun 등, 2015). 탐지(detection)나 분류(classification)에 필요한 특징(feature)을 데이터로부터 자동적으로 찾아낸다는 장점 덕분에, 딥 러닝은 지난 수 년간 이미지 인식 등 인간이 손쉽게 해내면서도 어떻게 해내는지 명확히 모르던 과제들에 대해 놀라운 성취를 이루어냈으며, 오늘날에는 자연어 처리, 차량 자율 주행, DNA 분석 등 여러 분야에서 폭넓게 시도되고 있다.

이 장에서는 세 절에 걸쳐 우리가 이 연구에서 사용한 딥 러닝 알고리즘에 대해 각각 설명하고, 그 후 실험 세팅과 그 결과에 대해 논한다.

### 제 1 절 심층 신경망(Deep neural networks)

여기서는 기본적인 딥 러닝 알고리즘인 심층 신경망에 대해 설명한다. 독자의 이해를 돕기 위해 먼저 가장 간단한 형태의 심층 신경망인 은닉 층이 하나인 신경망에 대해서 알아보자. 그러기 위해서, 입력 변수에 대해서 다음과 같은  $M$  개의 선형 결합(linear combination)을 아래와 같이 구성한다.

$$a_j^1 = \sum_{i=1}^D w_{ji}^1 x_i + b_j^1, \text{ where } j = 1, \dots, M.$$

이때 파라미터  $w_{ji}^1$ 를 가중치(weight)로, 파라미터  $b_j^1$ 를 편향(biases)로 부르고, 선형 결합의 결과로 구해진  $a_j^1$ 는 액티베이션(activation)으로 부르기로 하자. 각 파라미터의 지수 ‘1’은 파라미터가 신경망의 ‘첫 번째’ 층에 해당됨을 나타내고, 이러한 층을 심층 신경망에서는 은닉 층(hidden layer)으로 부르는데, 입력 변수와 출력값 사이에 ‘숨어’ 있기 때문이다. 이렇게 구한 각 액티베이션을 미분 가능한 비선형 함수인 활성화 함수  $h(\cdot)$ 를 사용해 변환하면

$$z_j^1 = h(a_j^1).$$

출력값을 얻게 된다. 이때 활성화 함수로는 주로 sigmoid 함수나 tanh 함수

수 등이 널리 사용되며, 최근에는 ReLU(Rectified linear unit)도 많이 사용된다. 그럼 이 출력값  $z_1^1, \dots, z_M^1$  을 다시 아래와 같이 새롭게 선형 결합한다.

$$a_k^2 = \sum_{j=1}^M w_{kj}^2 z_j^1 + b_k^2, \text{ where } k = 1, \dots, K.$$

이때  $K$ 는 최종 출력값의 개수를 의미한다. 각 파라미터의 지수 ‘2’ 는 파라미터가 신경망의 ‘두 번째’ 층에 해당됨을 뜻하며, 이 신경망에서 은닉 층은 하나뿐이므로 이 층은 출력 층(Output layer)이 된다. 마지막으로, 액티베이션  $a_k^2$ 를 적절한 활성화 함수로 변환시켜주면 신경망의 최종 출력값  $y_k$ 를 얻을 수 있다. 만약, 이 신경망의 목적이 다중 분류라면 활성화 함수로 소프트맥스

함수  $y_k = \sigma(a_k^2)$ , where  $\sigma(a_k) = \frac{\exp(a_k)}{\sum_l \exp(a_l)}$  가 일반적으로 사용된다.

이를 그림으로 표현하면 아래와 같다.

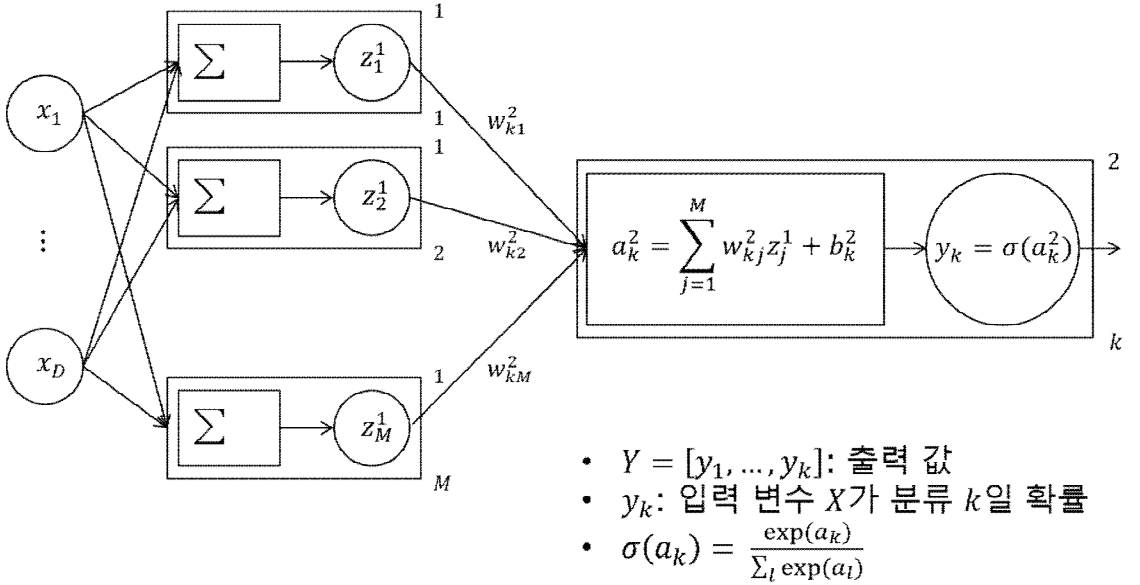


그림 2 Feed-forward networks

이때 하나 하나의 연산이 이루어지는 각각의 노드를 보면 입력 변수를 받아 계산하여 출력하는 것이 마치 신경망에서의 신경 세포와 같이 동작하기 때문에, 각 노드를 뉴런(neuron)이라 부른다. 또한 좌측의 입력 변수가 우측으로 진행



하면서 계산되어 최종 출력 값으로 변환되는 이 과정을 정보가 입력 변수에서 최종 출력 값까지 앞으로 전파되는 것으로 해석할 수 있으므로 이를 피드포워드 망 함수(Feed-forward network functions)라고 부른다. 보다 대중적인 이름으로는 다층 퍼셉트론(Multilayer perceptron)라고도 한다.

은닉 층을 하나로 한정하지 않고 여러 층을 사용할 수도 있는데, 은닉 층이 하나인 경우에 비해 더 적은 수의 뉴런으로 보다 복잡한 함수를 모사할 수 있다는 장점이 있다. 그러나 그 경우 뒤에서 살펴볼 역전파 알고리즘으로 신경망을 훈련시킬 때, 즉 파라미터를 조정할 때 좋지 않은 지역 최적해를 얻게 되는 단점이 있다. 이를 보완하기 위해 Hinton 등(2006)은 제한 볼츠만 기계(Restricted Boltzmann machine)을 이용하여 파라미터의 초기값을 정하는 심층 신뢰 신경망(Deep belief networks)를 제안하였고, 이 심층 신뢰 신경망은 딥 러닝 초기의 성공을 견인하였으나 이 연구에서 다루는 내용과는 다소 거리가 있으므로 이에 대해서는 여기서 갈음하고자 한다.

앞에서 본 피드포워드 망 함수는 정보를 입력 변수에서 출력값으로 보내기만 할 뿐이므로, 우리가 원하는 함수를 제대로 모사하기 위해서는 신경망 내의 파라미터를 조정해야 한다. 그러기 위해서는 신경망을 이용해 입력값으로부터 계산한 예측값과 실제 데이터에 있는 목표값 사이의 차이를 구하고, 이를 최소화하는 방향으로 파라미터  $W$ 를 조정해야 한다.

인공 신경망으로 분류 과제를 수행할 경우 예측값  $y_{nk}$ 는 입력 벡터  $X_n$ 이 분류  $k$ 에 속할 확률을 나타내고, 목표값  $t_{nk}$ 는 입력 벡터  $X_n$ 이 분류  $k$ 에 속하는지 아닌지를 나타내는 이진수이며  $\sum_k t_{nk} = 1$ 이어야 한다. 그럼 negative log likelihood로 구해지는 오류 함수  $E(W)$ 는 다음과 같게 된다.

$$E(W) = - \sum_{n=1}^N \sum_{k=1}^K t_{nk} \ln y_{nk}$$

인공 신경망에서는 일반적으로 오류 함수를 최소화시키는 파라미터  $W$ 를 단숨에 찾을 수 없으므로 파라미터를 최적화시키는 방법으로 확률적 경사 하강법(Stochastic gradient descent, 이하 SGD)을 주로 사용한다. 즉, 우리는 피드포워드 망으로 계산한 확률을 바탕으로 오류 함수값을 계산하고, SGD 방법으로 오류 값을 줄이는 방향으로 파라미터를 조정하는 일을 반복해가며 지역 최적해를 찾아가게 된다. 이 알고리즘을 피드포워드 망의 흐름과 반대방향으로 오

류 정보를 전파한다고 하여 오류 역전파 알고리즘(Error backpropagation algorithm)이라고 부른다.

## 제 2 절 심층 신경망을 위한 추가 구조

비록 인공 신경망이 한 개의 은닉 층과 유한한 개수의 은닉 뉴런으로 모든 함수 관계를 모사할 수 있음이 Hornik (1991)의 Universal approximation theorem 로 증명되기는 했으나, 실제로 전처리를 하지 않은 ECG 데이터 자체에서 우리가 원하는 특징을 바로 찾아내는 걸 기대하기는 매우 어려웠다. 그 때문에 은닉 층이 하나인 인공 신경망을 ECG 분석에 사용한 많은 연구자들은 ECG 신호의 노이즈를 제거한 후 QRS 탐지를 수행하여 신호 안 맥박의 위치를 찾고, 맥박의 특징을 수치화하여 이 값을 인공 신경망에 입력하는 방식을 주로 택하였는데, (Martis 등, 2014) 기존의 이런 방법은 분석을 위해 연구자가 맥박의 특징을 수치화할 방법을 일일이 찾아내야 했을 뿐만 아니라, 맥박을 찾기 어려운 ECG 데이터에서는 효과를 발휘하기 힘든 문제가 있었다.

그러나 분류가 되어 있는 충분한 데이터를 얻기란 쉽지 않은 현실에서, ECG에 대한 적절한 전처리는 인공 신경망의 분류 성능을 높일 수 있는 유력한 방법이다. 그러므로 우리에게는 QRS 탐지에 의지하지 않고서도 주어진 ECG 데이터의 특징을 명확히 추출해낼 수 있는 전처리 방법이 필요하다. 그래서 우리는 ECG 신호의 특징을 추출하여 다음 층에 전달하는 역할을 맡은 인공 신경망 내 추가 구조를 두 가지 고안하여 심층 신경망에 도입하였다.

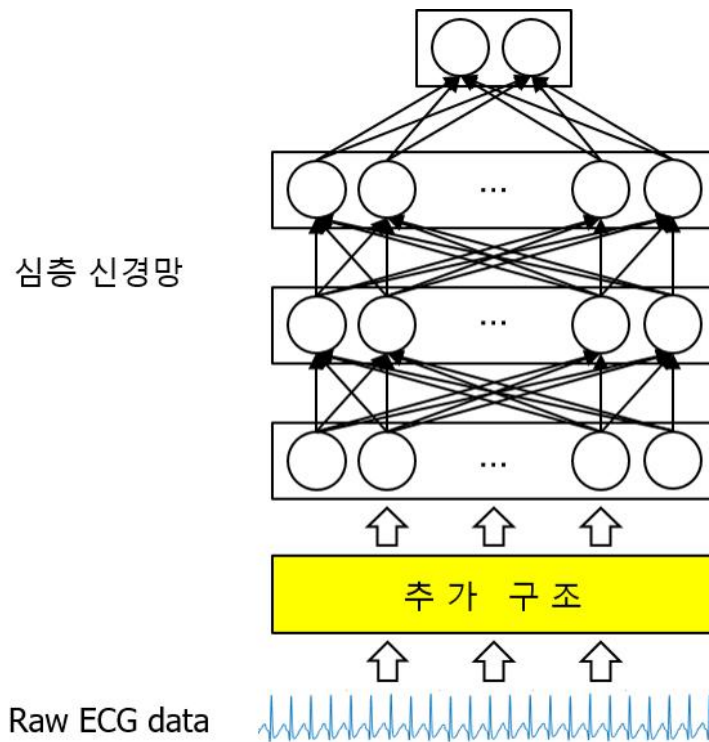


그림 3 Figure of additional structure for ECG analysis

### 1. Max-pooling 층 + 심층 신경망

입력받은 데이터에 대해 sub-sampling을 수행하는 풀링 층은 신경망의 크기를 줄여주며, 신경망이 입력 데이터의 작은 변화에도 큰 영향을 받지 않게 해주는 장점이 있다. 이는 올바른 인식을 방해할 수 있는 작은 노이즈가 무수히 존재하는 ECG 신호를 처리하는 데 유용할 수 있다.

예를 들어 입력 데이터가 100개의 값으로 이루어진 시계열 데이터이고 풀링 크기가 4라면, 입력 데이터를  $100/4=25$ 개의 구간으로 나눈 후 각 구간의 최대 값을 뽑아 25개의 값으로 이루어진 새로운 시계열 데이터를 구성하는 식이다.

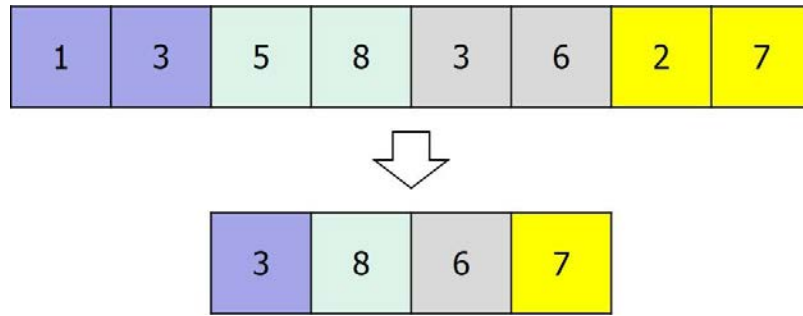


그림 4 Max-pooling layer

## 2. Max-Min 층 + 심층 신경망

많은 전문가들이 ECG를 분석할 때 중요하게 생각하는 것 중 하나가 R peak의 진폭과 P-QRS-T wave의 너비이다. 그러므로 이 부분을 명확히 잡아내어 심층 신경망을 훈련할 때 사용한다면 보다 나은 성과를 낼 것으로 기대할 수 있다. 이를 위해 새로이 고안한 추가 구조가 바로 Max-Min 층이다. Max-Min 층은 다음과 같이 동작한다. 정해진 크기의 window를 ECG 신호 위에서 시간 방향으로 움직이면서 window 내의 최대값과 최소값을 선택한 후, 그 두 값의 차이로 이루어진 새로운 신호를 만드는 것이다.

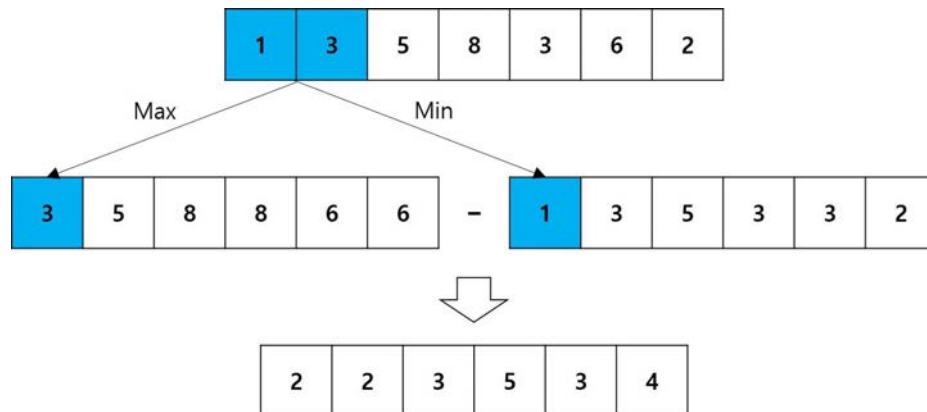


그림 5 Max-Min layer

Max-Min 층은 몇 가지 장점이 있다. 먼저, P-QRS-T wave의 진폭과 너비를 명확하게 보여주며 ECG에 들어가기 쉬운 잡음을 상당부분 제거해준다. 또한 ECG 내부의 불필요한 추세를 제거하고 P-QRS-T wave를 강조해준다. 실제 ECG 신호를 Max-Min 층을 이용해 처리한 결과는 다음과 같다.

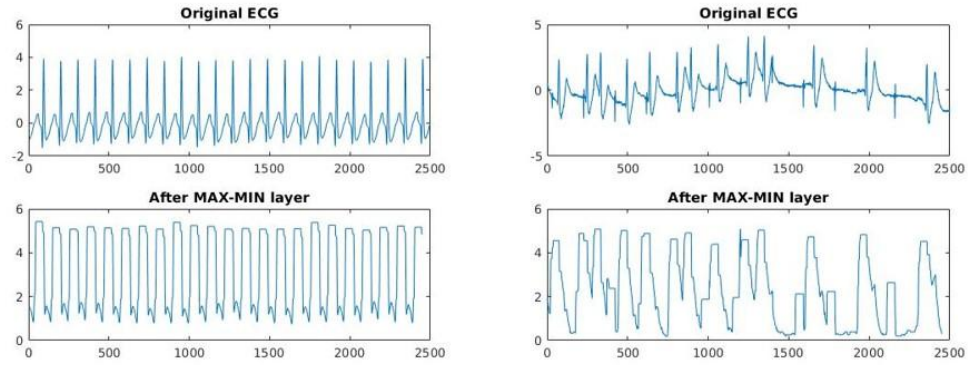


그림 6 Effect of the Max-Min layer

### 제 3 절 Convolutional neural networks

Convolutional neural networks (이하 CNN) 는 복수의 배열로 이루어진 데이터를 처리하기 위해 고안된 인공 신경망으로, 2000년대 초반 이후 이미지에서의 물체나 부분의 탐지, 구분, 인식에 있어서 엄청난 성공을 거둔 신경망 알고리즘으로(LeCun 등, 2015), 현재에도 활발히 연구가 진행 중인 신경망이다. 이러한 CNN 에 대해 이해하기 위해서는 먼저 convolution 이란 개념에 대해 알 필요가 있는데, Convolution(합성곱) 이란 두 함수에 대한 연산으로, 한 함수의 값과 다른 함수를 반전 이동한 값을 곱한 후 구간에 대해 적분(연속함수)하거나 합산(이산함수)하는 연산이다. 1차원에서의 convolution 연산을 수식으로 표현하면 다음과 같다.

$$s(t) = (x * w)(t) = \int_{-\infty}^{\infty} x(a)w(t-a)da, \text{ or } \sum_{a=-\infty}^{\infty} x[a]w[t-a].$$

이때  $x$ 를 CNN 용어로는 입력값(input),  $w$ 를 커널(kernel)로 부르며, convolution의 결과로 얻은 출력값  $s$ 를 feature map 이라고도 부른다. 한편, 많은 경우에 보다 고차원의 함수에 대해서도 convolution 연산을 수행하는데, 예를 들어 2차원 이미지  $I$  와 2차원 커널  $K$  를 가지고 convolution 연산을 수행한다면 다음과 같이 쓸 수 있다.

$$s[i,j] = (I * K)[i,j] = \sum_m \sum_n I[m,n]K[i-m,j-n].$$

Convolution은 교환법칙이 성립하므로, 위 식은 아래와 같이 바꿀 수 있다.

$$s[i, j] = (I * K)[i, j] = \sum_m \sum_n I[i - m, j - n] K[m, n].$$

그러나 많은 인공 신경망 라이브러리에선 CNN을 실제로 구현할 때 convolution 대신 상호상관(cross-correlation)으로 구현하는데, 상호상관이란 convolution과 유사하지만 함수의 반전 이동이 없는 연산이며 수식으로 다음과 같이 쓸 수 있다.

$$s[i, j] = (I * K)[i, j] = \sum_m \sum_n I[i + m, j + n] K[m, n].$$

많은 기계 학습 라이브러리에서는 실제로 상호상관으로 구현하면서도 이를 convolution으로 부르기 때문에, (Bengio 등, 2015) 이 연구에서도 이러한 관습대로 커널의 반전 이동과 상관없이 convolution이라는 이름으로 통칭한다.

일반적인 CNN은 크게 세 종류의 층으로 구성된다.

- Convolution 층

Convolution 층에서는 입력된 신호 혹은 이미지와 개의 다른 커널에 대해 각각 convolution 연산을 수행하고, 그 결과로 얻은 개의 feature map을 활성화 함수  $h(\cdot)$ 를 통해 변환하여 다음 층으로 내보낸다. 수식으로는 아래와 같이 표현할 수 있다.

$$s^k[i, j] = h((w^k * x)[i, j]), \text{ where } k = 1, \dots, K.$$

- Pooling 층

Convolution 층에서 출력된 개의 feature map에 대해서 sub-sampling을 수행하는 층이다. 정해진 범위 안에서 최대값 하나를 골라 출력하는 max-pooling이나 평균값을 계산해 출력하는 average-pooling이 많이 사용된다. 풀링을 하게 되면 데이터 포인트의 수가 줄어들 뿐만 아니라, 입력 데이터의 작은 변화에도 크게 영향을 받지 않는 이점이 생긴다. 일반적인 CNN에서는 convolution 층과 pooling 층을 같이 묶어 여러 번 반복하여 배치함으로써 최초의 입력 데이터로부터 특징을 추출해낸다.

- 다층 퍼셉트론

다층 퍼셉트론은 일반적으로 여러 개의 convolution-pooling 층 다음에 위치하며, 앞서의 convolution-pooling 층에서 얻어진 특징들을 가지고 입력된 데이터를 분류하는 역할을 맡는다.

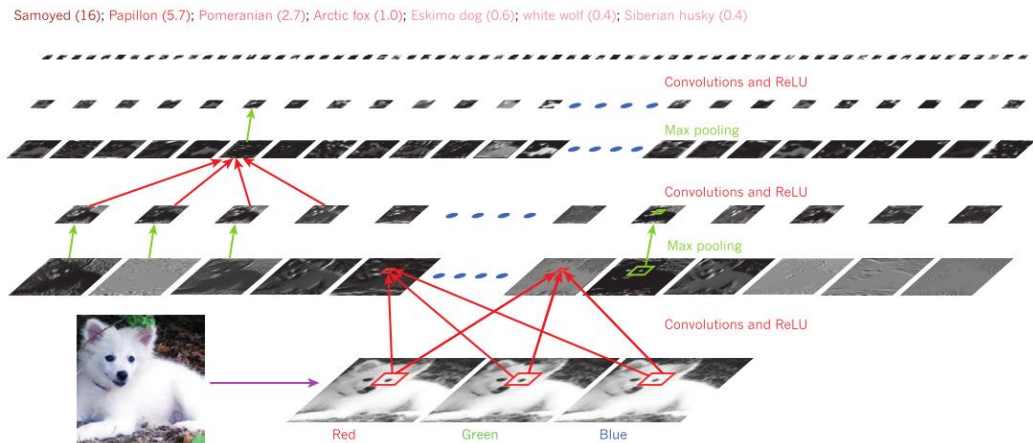


그림 7 Inside a convolutional network. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

위 그림은 convolution 신경망을 통해 사모예드 강아지를 분류하는 장면을 묘사한 것이다. 그림에서 볼 수 있듯, convolution-pooling 층이 반복하여 여러 차례 배치되어 있는 것을 볼 수 있다.

기존의 심층 신경망과 CNN을 구별짓는 차이점은 세 가지다. 먼저, 전통적인 심층 신경망은 서로 인접한 층에 있는 뉴런들이 모두 연결되어 있는데 반해 CNN은 커널의 크기를 입력 데이터보다 훨씬 작게 설정함으로써 일부의 뉴런끼리만 서로 연결되게끔 한다. 두 번째로, CNN에서는 한 커널 안의 파라미터를 입력 데이터 전체에 대해 공유하므로 심층 신경망에 비해 연산이 보다 효율적이다. 마지막으로, 커널이 입력 데이터 전체에 걸쳐 convolution 연산을 수행하므로 입력 데이터 내에 변화가 생기면 출력될 때도 같은 식으로 변한다. 예를 들어, 시계열 데이터 안에 존재하는 특정 이벤트가 나중에 발생한 것처럼 뒤로 움직인다면, 새로 변한 입력 데이터를 CNN에 통과시켜 얻은 출력 결과에서도 그 특정 이벤트에 해당하는 feature가 나중 부분에 나타나게 된다.

## 제 4 절 데이터 및 실험 설계

### 1. 데이터

이 연구를 위해 우리는 두 종류의 다른 데이터베이스에서 ECG 기록을 수집하였다. 첫 번째 데이터베이스(이하 DB1)는 2015 PhysioNet/CinC Challenge에서 공개된 훈련 데이터로, 중환자실의 모니터링 기기에서 수집한 750개의 레

코드로 구성되어 있다. 이 데이터의 특징은 레코드 시작 시점에서 5분 후에 제 2 장에서 설명한 다섯 종류의 치명적인 부정맥 중 하나가 발생했음을 알리는 경보가 올었다는 점이다. 일단의 전문가 집단이 경보가 옳은 경보였는지 잘못된 경보였는지를 심사하여 각 레코드를 분류하였으며, ANSI/AAMI EC13 Cardiac Monitor Standards의 기준에 따르면 부정맥이 발작한 지 10초 내에 반드시 경보가 발생해야 하므로 경보를 발생시킨 직접적 요인은 경보 발생 이전 10초 간의 데이터에 존재한다. 모든 레코드는 12bit, 250Hz로 리샘플되었으며, 각각 두 종류의 ECG 신호와 하나 이상의 추가적인 생체 신호(예: photoplethysmogram (PPG) 등)를 포함한다.

두 번째 데이터베이스(이하 DB2)는 PhysioNet이 보유한 Multi-parameter Intelligent Monitoring in Intensive Care II (MIMIC-II) version 2 데이터로, 마찬가지로 중환자실의 모니터링 기기에서 수집한, 길이가 제각기 다른 4458개의 레코드로 되어 있으며 경보가 올린 시점과 그 경보를 일으킨 것으로 의심되는 증상이 레코드에 함께 기록되어 있다. 그러나 이 레코드 중 일부만이 제 2 장에서 설명한 치명적인 부정맥에 해당하는 ECG 신호를 담고 있으며 부정맥으로 인한 경보가 참인지 거짓인지에 대한 주석이 부재한 상태였다. 그래서 A. Aboukhalil 등(2008)은 이 데이터 중 498개의 레코드를 선정하여 한 팀의 전문가들과 함께 해당 레코드 내의 부정맥 경보를 일일이 분석하여, 최종적으로 환자 447명에 대한 41,301 시간 동안의 레코드에서 발생한 부정맥 경보를 참과 거짓 경보로 분류하고 이 결과를 PhysioNet에 공유하였다. 이 데이터는 8bit, 125Hz로 리샘플되었으며, 각각 복수의 ECG 신호 및 추가적인 생체 신호를 포함한다.

우리는 이들 데이터 베이스 중에서 오직 ECG II, 즉 전극 II에서 수집된 ECG만을 연구에 사용하였다. 또한, 위급한 부정맥 증상에 대한 ECG 정보는 경보 발생 이전 10초 간의 데이터에 존재하므로, 전체 데이터 베이스에서 이 부분만을 따로 추출하여 사용하였다. 데이터 베이스 내 일부 레코드에 ECG II 데이터가 없었기 때문에, 최종적으로 DB1에서 얻은 10초 길이의 728개 레코드와 DB2에서 얻은 10초 길이의 5748개 레코드를 추출하였고, 125Hz로 샘플된 DB2 레코드를 DB1 레코드와 함께 사용하기 위해 interpolation을 수행하여 250Hz로 변환하였다.



각 레코드는 경보를 일으킨 증상에 대한 정보와 경보가 참인지 아닌지에 대한 정보를 가지고 있어 이들 정보를 사용해 신경망을 훈련시키고 그 성능을 시험하였다. 신경망의 훈련에는 무작위로 선택한 DB1의 600개 레코드와 DB2의 5620개 레코드를 합친 총 6220개 레코드를 사용하였고, 신경망의 테스트에는 DB1의 128개 레코드와 DB2의 128개 레코드를 합친 총 256개 레코드를 사용하였다.

	Training		Test		계
	True alarm	False alarm	True alarm	False alarm	
DB1	232	368	50	78	728
DB2	3318	2302	69	59	5748
계	3550	2670	119	137	6476

표 2 Summary of the training and test data

## 2. 실험 설계

먼저, 이 연구의 목적에 적합한 인공 신경망 구조를 찾기 위해 총 일곱 가지의 인공 신경망 구조를 시험하였다. 시험한 모든 신경망 구조가 동일한 입력 층 크기(2500,  $250\text{Hz} \times 10\text{초}$ )와 출력 층 크기(2, Positive/Negative)를 가지므로, 입력 층과 출력 층 사이에 배치되는 은닉 층의 수와 층 안의 뉴런 개수에 따라 각각의 신경망 구조가 결정된다. 시험한 은닉 층의 종류는 다음과 같다: [250-250], [500-500], [750-750], [1000-1000], [250-250-250], [500-500-500], [750-750-750].

그 다음으로 고려해야 할 것은 하이퍼 파라미터이다. Bengio(2012)는 ‘데이터를 학습 알고리즘에 적용하기 전에 미리 설정해야 하는 변수로, 학습 알고리즘 자체에서 바로 선택되지는 않는 것’을 하이퍼 파라미터로 정의하였고, 하이퍼 파라미터를 선택하는 것이 어떤 모델을 선택할지와 동등하게 중요한 문제라고 지적한 바 있다. 여러 인공지능 커뮤니티에서 딥 러닝에 최적화된 하이퍼 파라미터 탐색을 주제로 연구를 진행하고 있으나, 현재까지 대부분의 연구자

들이 하이퍼 파라미터 선택을 위해 많은 실험에 기반한 경험적인 방법에 의존하고 있으며, 우리 또한 반복 실험을 통해 하이퍼 파라미터도 통상적으로 시도되는 파라미터 범위 내에서 보다 좋은 파라미터를 얻고자 하였다.

최종적으로 이 연구에서 적용된 여러 딥 러닝 알고리즘에 공통적으로 적용된 하이퍼 파라미터는 다음과 같다. 먼저, SGD를 통해 신경망의 가중치를 얼마나 개선할지를 결정하는 학습률(learning rate)은  $1 \times 10^{-3}$ 로 하고, 신경망의 학습이 반복될수록 학습률을 감소시키기 위한 감소율은  $1 \times 10^{-5}$ 로 설정하였다. 이 때 update iteration  $n$ 에서의 학습률은 다음과 같다.

$$n\text{에서의 학습률} = \frac{\text{초기 학습률}}{1 + (n \times \text{감소율})}$$

또, 신경망을 학습시킬 때마다 사용하는 데이터의 개수를 의미하는 mini-batch size는 30으로 설정하였으며, 무작위로 훈련시킬 뉴런을 선택함으로써 인공 신경망의 과적응(overfitting) 문제를 완화하는 드롭 아웃(drop-out)의 확률은 0.5로 설정하였다.

Convolutional neural network는 4장에서 설명한 것처럼 convolution-pooling 층이 심층 신경망 앞에 연이어 배치된 형태이므로, convolution-pooling 층의 개수, 풀링 크기, 그리고 커널 크기에 대해서도 정해줘야 한다. 커널 크기에 비해 convolution-pooling 층의 개수와 풀링 크기가 신경망의 성능에 훨씬 큰 영향을 준다는 것을 경험적으로 알고 있었으므로, 시간 절약을 위해 DB1 만 이용하여 가장 좋은 층의 수와 풀링 크기를 먼저 탐색했다.

Accuracy		Pooling size									
		1	2	3	4	5	6	7	8	9	10
# of Conv. Layers	1	62.50%	63.28%	64.84%	61.72%	63.28%	64.06%	63.28%	63.28%	65.63%	64.84%
	2	62.50%	64.06%	67.19%	69.53%	74.22%	80.47%	81.25%	84.38%	82.03%	82.81%
	3	63.28%	69.53%	77.34%	79.69%	83.59%	83.59%	84.38%	84.38%	-	-
	4	64.84%	75.78%	85.16%	82.03%	-	-	-	-	-	-
	5	63.28%	78.91%	-	-	-	-	-	-	-	-
	6	67.19%	55.47%	-	-	-	-	-	-	-	-
	7	62.50%	-	-	-	-	-	-	-	-	-
	8	55.47%	-	-	-	-	-	-	-	-	-
	9	55.47%	-	-	-	-	-	-	-	-	-
	10	55.47%	-	-	-	-	-	-	-	-	-

그림 8 Outpt of different convolutional layer and pooling size

그 결과 위 그림에서 보이는 것과 같이 convolution-pooling 층이 4층이고 풀링 크기가 3일 때 가장 좋은 결과를 보였다. 각 convolution 층의 feature map 수는 75로 동일하게 설정하였고, 커널 크기는 0.2초에 해당하는 50으로 설정하였다.

마지막으로, 실험에 사용한 인공 신경망은 오픈소스 기계학습 프레임워크 Torch7 로 구현하였고, i7-5930K CPU와 GTX 980 Ti GPU로 구성된 하드웨어에서 구동하였다.

### 3. 평가 기준

정확도만으로는 어떤 질병에 대한 검사의 성능을 판단하기 어려운 경우가 많다. 예를 들어, 발병률이 아주 낮은 질병의 경우 질병을 전혀 탐지하지 못하는 검사법이라 할지라도 정확도는 높게 나올 수 있기 때문이다. 이러한 문제를 해결하기 위해 가장 많이 쓰이는 수치로 바로 민감도(Sensitivity, 혹은 true positive rate)와 특이도(Specificity, 혹은 true negative rate)가 있다. 민감도와 특이도는 다음과 같이 구한다.

$$\text{민감도} = \frac{TP}{TP + FN}, \text{ 특이도} = \frac{TN}{TN + FP}.$$

또한, PhysioNet/CinC Challenge 2015에서는 Challenge 평가 기준으로 False Negative에 페널티를 준 다음과 같은 점수 기준을 제시하였다.

$$\text{Challenge score} = \frac{TP + TN}{TN + TN + FP + (5 \times FN)}.$$

이 연구에서는 각 알고리즘의 성능을 비교하기 위해 앞에서 제시한 정확도, 민감도, 특이도, 그리고 challenge 2015 score를 사용하도록 한다.

Outcome		Golden standard	
		True	False
Prediction	Positive	TP	FP
	Negative	FN	TN

표 3 Confusion matrix

## 제 5 절 결과

### 1. 기본 심층 신경망

심층 신경망 알고리즘으로 일곱 가지 신경망 구조에 대하여 실험한 결과는 아래 도표와 같다. 신경망을 훈련시키는 기준으로 사용한 negative log likelihood를 바탕으로 각 구조를 비교하면, 전반적으로 은닉 층이 두 개인 신경망에 비해 은닉 층이 세 개인 신경망이 보다 나은 성과를 보였으나, 층 내 뉴런의 수가 일정 개수 이상이 되면 오히려 성능이 떨어지는 모습을 보였다. 가장 높은 정확도를 보인 [750-750-750]에서 심층 신경망은 정확도 71.09%, 민감도 67.44%, 특이도 74.80%, Challenge 2015 score 47.39의 성능을 보였다.

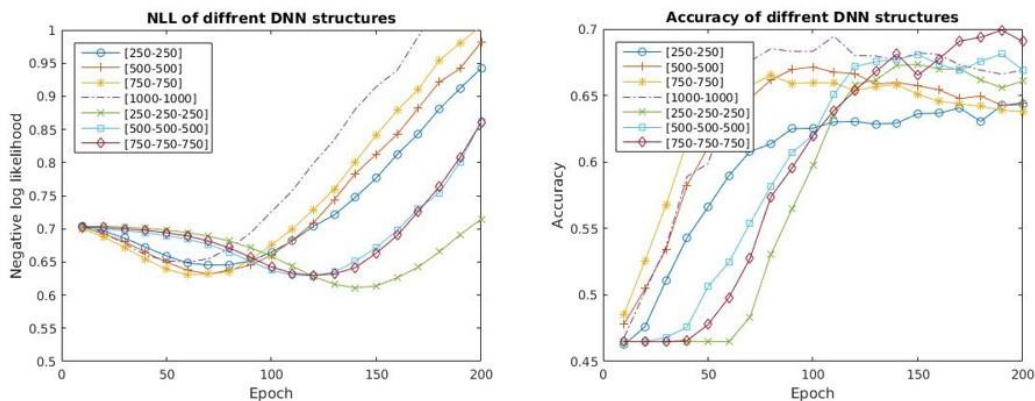


그림 9 Results of different structure of DNN. 10 epoch마다 평균을 낸 값을 나타내었다.

### 2. Max pooling 층 + 심층 신경망

Max pooling 층 + 심층 신경망의 경우, 풀링 크기에 따른 성능 향상 폭을 확인하기 위해 풀링 크기가 2인 인공 신경망과 4인 인공 신경망으로 나누어 실험을 진행하였다. 그 결과 도표에서 보이는 바와 같이 두 경우 모두, 풀링 층을 적용하지 않은 심층 신경망보다 negative log likelihood는 낮고 정확도는 높은 결과를 보여주었다. 그러나 풀링 크기가 2인 경우와 4인 경우를 각각 비교해보면, [1000-1000]을 제외한 다른 은닉 층 구조에선 개선이 거의 없으며, [1000-1000]의 경우에도 변화의 폭이 유의미할 만큼 크지 않다는 걸 볼 수 있다.

풀링 크기가 2인 인공 신경망의 경우 정확도 74.22%, 민감도 73.87%, 특이도 74.48%, Challenge 2015 score 47.03의 성능을 보였고, 풀링 크기가 4인 인공 신경망의 경우 정확도 74.61%, 민감도 69.85%, 특이도 80%, Challenge 2015 score 54.26의 성능을 보여주어 두 신경망 모두 기본 심층 신경망에 비해 더 나은 성능을 보여주었다.

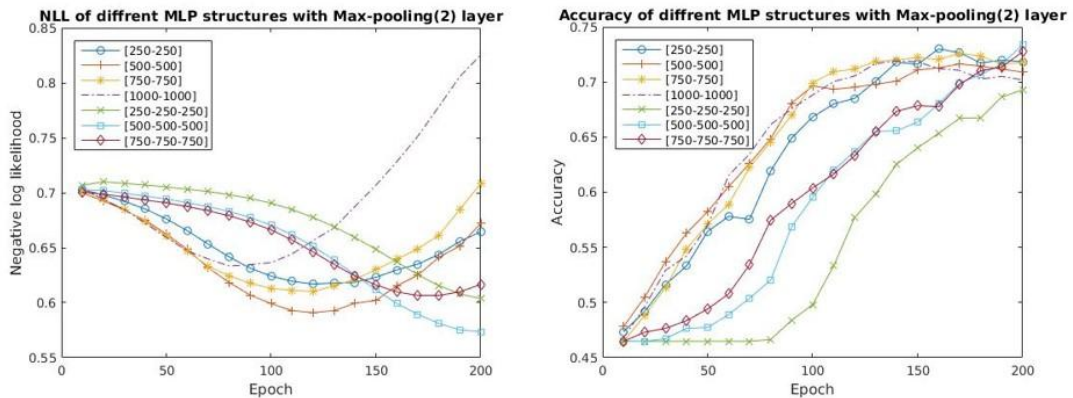


그림 10 Results of different structure of DNN with max-pooling(pooling size=2) layer. 10epoch마다 평균을 낸 값을 나타내었다.

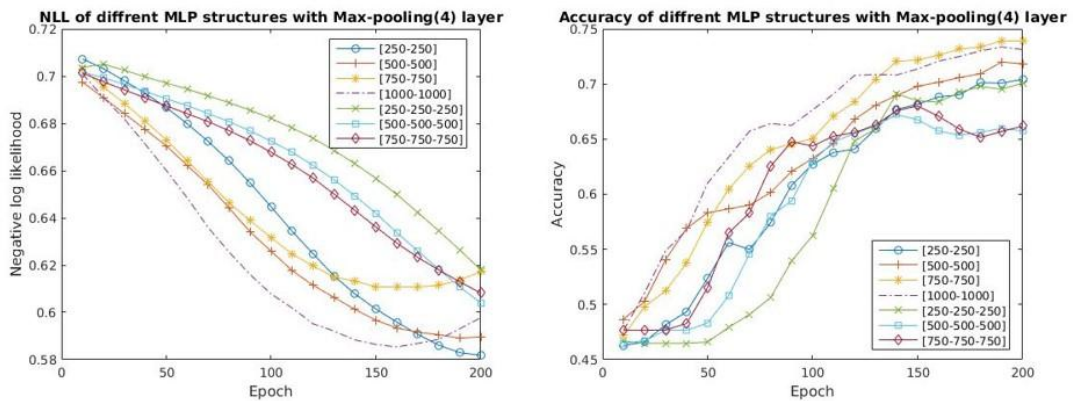


그림 11 Results of different structure of DNN with max-pooling(pooling size=4) layer. 10epoch마다 평균을 낸 값을 나타내었다.

### 3. Max-Min 층 + 심층 신경망

Max-Min 층 + 심층 신경망의 경우, 은닉 층의 종류에 따른 negative log likelihood와 정확도 모두 큰 차이를 보이지 않는 것을 발견할 수 있었다. Max-Min 층을 추가한 심층 신경망은 그렇지 않은 심층 신경망에 비해 뚜렷한

정도의 성능 향상이 이루어진 것을 볼 수 있다. 정확도 81.64%, 민감도 80%, 특이도 83.09%, Challenge 2015 score 60.06으로 기본 심층 신경망은 물론, Max pooling 층 + 심층 신경망의 경우보다 모든 수치에서 더 높은 결과를 보여 주고 있다.

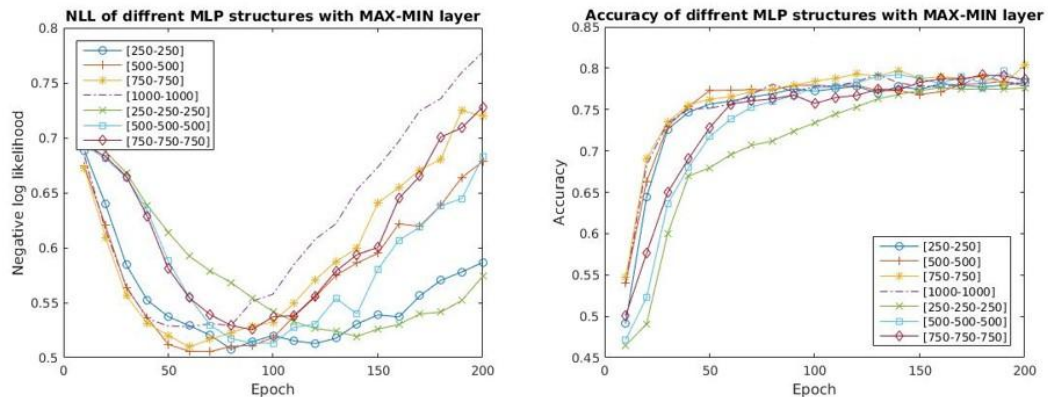


그림 12 Results of different structure of DNN with max-min layer. 10 epoch마다 평균을 낸 값을 나타내었다.

#### 4. Convolutional neural networks

Convolutional neural networks 의 경우, negative log likelihood와 정확도 모두 큰 차이를 보이지 않는 것을 발견할 수 있었다. 이 결과를 통해 4.2에서 언급한 것처럼 CNN의 성능은 그 무엇보다 convolution-pooling 층의 층수와 풀링 크기에 가장 큰 영향을 받으며, convolution-pooling 층에서의 연산을 통해 입력 데이터에 대한 적절한 특징이 추출되기만 하면 그 특징을 이용해 분류하는 역할을 담당하는 다층 퍼셉트론의 구조에는 성능이 큰 영향을 받지 않는다는 것을 알 수 있다. CNN은 정확도 88.67%, 민감도 89.47%, 특이도 88.03%, Challenge 2015 score 70.06의 성능을 보이면서 심층 신경망은 물론, 추가 구조를 포함한 심층 신경망들에 비해서도 놀라울 정도로 좋은 성능을 보여주었다.

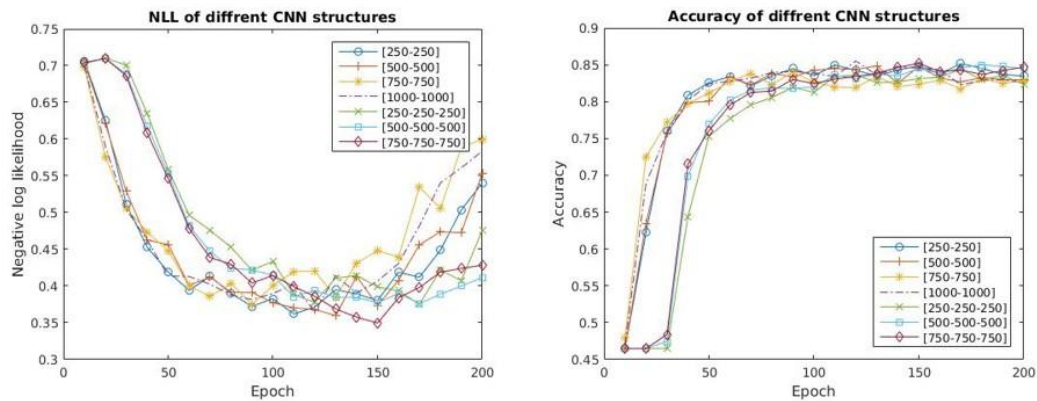


그림 13 Results of different structure of CNN. 10 epoch마다 평균을 낸 값을 나타 내었다.

Max-Min 층 + 심층 신경망과의 보다 정확한 비교를 위해, 데이터 집합 내 에서 서로 다르게 뽑은 10개의 training set/test set 을 준비한 후 이를 Max-Min 층 + 심층 신경망과 CNN 에 학습시켜 그 성능을 비교한 결과, Figure 17에서 볼 수 있듯이 CNN 알고리즘이 확실히 보다 더 우월한 것을 확 인할 수 있었다.

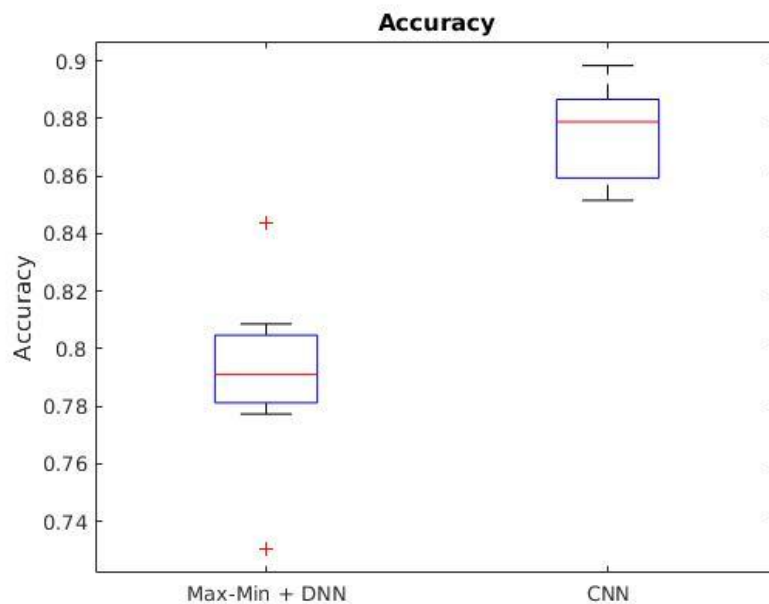


그림 14 Comparison between Max-Min + DNN and CNN



Convolutional neural networks 가 올바르게 훈련되었는지 판단하는 지표 중 하나는 훈련 후 커널의 형태가 특정한 패턴을 띄고 있는지를 확인하는 것이다. 특히 이미지 인식을 할 때 Convolutional neural networks 은 반복 훈련을 통해 처음에 무작위로 주어졌던 커널을 특정 패턴과 유사한 모습으로 바뀌어나가며, 그 결과 이미지 안에서 커널의 패턴과 유사한 모양을 정확히 찾아 인식할 수 있게 된다. 그러나 이 연구에서 Convolutional neural networks 의 훈련 후 커널의 모양을 보면, Figure 19에서 볼 수 있듯 분명히 훈련을 통해 커널의 모양은 변했지만, 인공적인 패턴을 보이고 있다고는 하기 어렵다. 즉, 이 연구에서 제시된 Convolutional neural networks 구조가 ECG 분석에 온전히 적합한 형태는 현재 아니며, 개선의 여지가 있음을 보여준다.

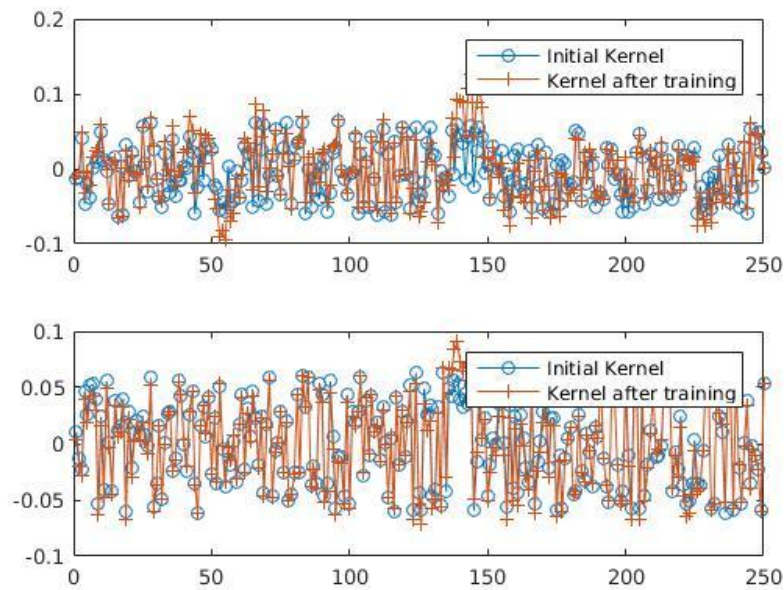


그림 15 Convolutional kernel before and after training



지금까지 살펴 본 인공 신경망의 성능을 요약하면 다음과 같다.

	경보 시스템	기본 심층 신경망	Max-pooli ng (2) layer	Max-pooli ng (4) layer	Max-Min layer	<b>CNN</b>
정확도	46.48%	71.09%	74.22%	74.61%	81.64%	<b>88.67%</b>
민감도	46.48%	67.44%	73.87%	69.85%	80%	<b>89.47%</b>
특이도	-	74.80%	74.48%	80%	83.09%	<b>88.03%</b>
Challenge	46.48	47.39	47.03	54.26	60.06	<b>70.06</b>

표 4 Summary of each deep learning algorithm

지금까지 우리는 딥 러닝 알고리즘을 일반적인 negative log likelihood function을 error function으로 삼아 훈련시켜 왔다. 그러나 사실 이 연구가 생명에 치명적인 부정맥 증상을 탐지하는 것을 목표로 삼는 이상, 정확도를 다소 희생하더라도 ‘실제로 위급한 증상인데 알고리즘이 위급하지 않다고 판단하는’ 이른바 false negative를 줄이는 것이 보다 현실에 적합할 것이다. 이를 위해 우리는 표 3의 Confusion matrix 내 각 요소, **true positive, true negative, false positive, false negative**에 대해 **1:1:1:5의 가중치**를 주고 이를 negative log likelihood function에 반영함으로써 false negative를 줄이는 방향으로 알고리즘을 훈련시켰다. 이렇게 훈련시킨 결과, **convolutional neural networks** 는 무려 **76.4**, **Max-Min층 + 심층 신경망의 경우엔 63.75의 challenge score**를 달성하였다.

## 제 4 장 목표 달성도 및 관련 분야에의 기여도

### 제 1 절 PhysioNet/CinC challenge 2015 결과와의 비교를 통한 목표 달성 평가

PhysioNet/CinC challenge는 연구 목적으로 사용할 수 있는 방대한 생체 신호 데이터베이스를 제공하는 PhysioNet.org와 Computing in Cardiology 학회가 공동 주관하는 챌린지 행사로, biomedical engineering 분야에서 공신력 있는 챌린지 중 하나이다. 올해인 2015년의 챌린지는 ‘Reducing False Arrhythmia Alarms in the ICU’ 라는 주제로 2015년 2월 16일부터 2015년 9월 10일까지 진행되었으며, 이 챌린지에서 참가자들은 ECG II, V 등 다양한 종류의 ECG와, ABP, PPG 등의 다른 생체 신호를 바탕으로 중환자실 내에서 부정맥에 대한 오경보를 줄이는 미션을 수행했다. 비록 본 연구와 이 챌린지 간 시간 차이로 인해 챌린지에 직접 참여하지는 못했으나, 챌린지에서 주어진 목표는 본 연구가 지향하는 바와 유사하기에 챌린지에 참여한 도전자들의 성적과 본 연구에서 개발한 알고리즘의 성적을 비교함으로써 우리는 알고리즘의 성능이 어느 정도인지 평가할 수 있다.

이러한 맥락에서 제 3 장 제 5 절 끝부분에서 구한 CNN과 Max-Min + 심층 신경망의 challenge score와 PhysioNet/CinC Challenge 2015의 최고 득점자들과 점수를 비교해보면 아래 표와 같으며, CNN의 경우 최고 득점자에 준하는 성적을 올린 것을 확인할 수 있다.

Challenge 2015 score	
Filip Plesinger	76.6
<b>CNN</b>	<b>76.4</b>
Sibylle Fallet	73.0
Christoph Hoog Antink	71.0
Linda Eerikainen	70.0
Vignesh Kalidas	64.6

표 5 Comparison with PhysioNet/CinC Challenge 2015 challengers

비록 모든 동일한 데이터로 성능을 측정한 것이 아니므로 동일선상에서 비교하기에는 다소 무리가 있다 할지라도, 본 연구 에서 ECG II 데이터만을 사용한 것과 달리 Challenge의 참여자들은 다른 ECG 데이터와 추가적인 생체 신호 데이터(ABP, PPG 등)을 사용했다는 것을 감안하면, 딥 러닝, 특히 Convolutional neural networks을 이용한 부정맥 탐지 알고리즘이 충분히 경쟁력이 있다고 할 수 있으며, 소기의 목표를 달성했다고 평가할 수 있을 것이다.

## 제 2 절 중환자 생체신호 분석에의 기여

사실 간단한 인공 신경망을 이용한 ECG 분석은 사실 과거부터 꾸준히 이루어져 왔으나, 인공 신경망을 여러 층으로 구성하여 훈련시키는 딥 러닝 방법론을 ECG 분석에 사용된 것은 최근의 일이며 그 수도 많지 않다. Rodrigues와 Couto(2012)는 세 개의 은닉 층을 가진 심층 신경망으로 ECG의 노이즈를 제거하는 방법을 제안하였고, Zheng 등(2014)은 ECG를 포함한 다중 채널의 생체 신호를 딥 러닝 알고리즘 중 하나인 Convolutional neural networks 에 훈련시켜 사람의 움직임과 심장 이상을 탐지하는 방법을 시도하였다. 또, Dmitrievich(2015)는 ECG에서 추출한 심박변이도 데이터를 Convolutional neural networks 와 Deep belief networks 에 학습시켜 환자의 질환을 진단하는데 활용한 전례가 있다.

본 연구는 중환자를 대상으로 한 경보 알고리즘에 딥 러닝을 적용한 최초의 사례이며, 적용 결과 기존의 경보 시스템을 능가하는 성능을 달성했다는 점에서 중환자 생체신호 분석 분야에 기여하였다. 이는 매우 노이즈가 많은 생체 신호에 대해서도 노이즈 제거나 QRS 탐지 같은 전처리 없이 바로 분석할 수 있음을 보여주는 것으로, 범용적인 ECG 분석에 딥 러닝 방법론이 적합할 수 있음을 제시한 것이다. 추후 후속 연구에서 ECG 분석은 물론, 다른 생체 신호 분석에 딥 러닝 방법론이 활용될 수 있고 그 전망도 밝다는 것이 이 연구를 통해 제시되었다는 점에서 본 연구의 의의가 있다고 할 수 있다.

## 제 5 장 연구개발결과의 활용계획

본 연구는 딥 러닝 기술로 생체 신호를 분석하는 후속 연구들에 대한 시작점으로 활용될 수 있다. 이 연구가 중환자를 대상으로 하는 만큼 높은 정확도를 달성하는 것이 그 무엇보다 중요하기 때문에, 과적응(overfitting)을 방지할 수 있는 full bayesian learning 기법 등을 이 연구 결과에 적용하여 정확도를 더욱 향상시키는 것이 선결 문제다. 이렇게 정확도가 향상된 딥 러닝 알고리즘을 바탕으로 PhysioNet/CinC Challenge 2016에 출전해 딥 러닝 방법의 유효성을 객관적으로 검증받을 계획이며, 이후 웨어러블 기기에서 수집되는 생체 신호를 분석하여 사용자의 상태를 진단하는 **웨어러블 기반 원격 의료**로 연구를 확장할 수 있을 것이다.

## 제 6 장 참고 문헌

- [1] Chambrin, M. C., Ravaux, P., Calvelo-Aros, D., Jaborska, A., Chopin, C., & Boniface, B. (1999). "Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis." *Intensive care medicine*, 25(12), 1360-1366.
- [2] Chambrin, M. C. (2001). "Alarms in the intensive care unit: how can the number of false alarms be reduced?" *Critical Care (London, England)*, 5(4), 184-188.
- [3] Lawless, S. T. (1994). "Crying wolf: false alarms in a pediatric intensive care unit." *Critical care medicine*, 22(6), 981-985.
- [4] Tsien, C. L., & Fackler, J. C. (1997). "Poor prognosis for existing monitors in the intensive care unit." *Critical care medicine*, 25(4), 614-619.
- [5] Cropp, A. J., Woods, L. A., Raney, D., & Bredle, D. L. (1994). "Name that tone. The proliferation of alarms in the intensive care unit." *CHEST Journal*, 105(4), 1217-1220.
- [6] Imhoff, M., & Kuhls, S. (2006). "Alarm algorithms in critical monitoring." *Anesthesia and Analgesia*, 102(5), 1525-1537.
- [7] Imhoff, M., Fried, R. (2009). "The crying wolf: Still crying?" *Anesthesia and Analgesia*, 108(5), 1382-1383.
- [8] Rodrigues, R., & Couto, P. (2012). "A Neural Network Approach to ECG Denoising." *arXiv preprint arXiv:1212.5217*.
- [9] Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). "Time series classification using multi-channels deep convolutional neural networks." *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8485 LNCS, 298-310.
- [10] Dmitrievich, I. A. (2015). "Deep Learning in information analysis of electrocardiogram signals for disease diagnostics", Ph.D. Thesis,

Moscow Institute of Physics and Technology, Moscow, Russian Federation.

- [11] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." *Nature*, 521(7553), 436-444.
- [12] Mehra, R (2007). "Global public health problem of sudden cardiac death." *Journal of electrocardiology* 40 (6 Suppl): S118-22.
- [13] Gacek, A., & Pedrycz, W. (Eds.). (2011). *ECG signal processing, classification and interpretation: a comprehensive framework of computational intelligence*, Springer Science & Business Media.
- [14] Aboukhalil, A., Nielsen, L., Saeed, M., Mark, R. G., & Clifford, G. D. (2008). "Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform." *Journal of Biomedical Informatics*, 41(3), 442-451.
- [15] American National Standards Institute. (2002). *American National Standard for Cardiac Monitors, Heart Rate Meters and Alarms*, Association for the Advancement of Medical Instrumentation.
- [16] Clifford, G. D. (2006). "ECG statistics, noise, artifacts, and missing data." *Advanced Methods and Tools for ECG Data Analysis*, 55-99.
- [17] Hinton, G. E. Osindero, S., & Teh, Y. W. (2006). "A fast learning algorithm for deep belief nets." *Neural computation*, 18(7), 1527-1554.
- [18] Yoshua Bengio and Ian J. Goodfellow and Aaron Courville. (2015). *Deep Learning*, Book in preparation for MIT Press.
- [19] Bishop, C. M. (2006). *Pattern recognition and machine learning*, springer.
- [20] Hornik, K. (1991). "Approximation capabilities of multilayer feedforward networks." *Neural networks*, 4(2), 251-257.
- [21] Martis, R. J., Acharya, U. R., & Adeli, H. (2014). "Current methods in electrocardiogram characterization." *Computers in Biology and Medicine*, 48(1), 133-149.

- [22] Bengio, Y. (2012). “Practical recommendations for gradient-based training of deep architectures.” *Neural Net*, 437-478.
- [23] Collobert, R., Kavukcuoglu, K., & Farabet, C. (2011). “Torch7: A matlab-like environment for machine learning.” *BigLearn, NIPS Workshop* (No. EPFL-CONF-192376)
- [24] 간호학대사전. (1996). 한국사전연구소

## 주 의

1. 이 보고서는 KAIST의 글로벌교육연구선도사업으로 지원하는 석박사과정 모험연구사업의 연구보고서입니다.
2. 이 보고서 내용을 발표할 때에는 반드시 석박사과정 모험연구사업의 연구결과임을 밝혀야 합니다.
3. 국가과학기술 기밀유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 안됩니다.