

Mortality Prediction of Older Adults Using Random Forest and Deep Learning

Junhyeok Park[†] · Songwook Lee^{††}

ABSTRACT

We predict the mortality of the elderly patients visiting the emergency department who are over 65 years old using Feed Forward Neural Network (FFNN) and Convolutional Neural Network (CNN) respectively. Medical data consist of 99 features including basic information such as sex, age, temperature, and heart rate as well as past history, various blood tests and culture tests, and etc. Among these, we used random forest to select features by measuring the importance of features in the prediction of mortality. As a result, using the top 80 features with high importance is best in the mortality prediction. The performance of the FFNN and CNN is compared by using the selected features for training each neural network. To train CNN with images, we convert medical data to fixed size images. We acquire better results with CNN than with FFNN. With CNN for mortality prediction, F1 score and the AUC for test data are 56.9 and 92.1 respectively.

Keywords : Mortality Prediction, Convolutional Neural Network, Random Forest, Feature Selection, Deep Learning

랜덤 포레스트와 딥러닝을 이용한 노인환자의 사망률 예측

박 준 혁[†] · 이 성 옥^{††}

요 약

우리는 응급실을 방문한 65세 이상 노인환자의 의료 데이터를 각각 피드 포워드 신경망과 합성곱 신경망에 학습하여 사망률을 예측하였다. 의료 데이터는 노인환자의 성별, 연령, 체온, 심박 수 등의 기초적인 정보뿐 아니라 과거 병력, 다양한 혈액 검사 및 배양 검사 결과 등 다양하고 복잡한 정보를 포함하여 총 99가지의 자질로 구성된다. 이 중 사망률 예측에 크게 기여하는 자질을 선택하기 위해 랜덤 포레스트를 이용하여 자질의 중요도를 계산하였고, 그 결과 중요도가 높은 상위 80개의 자질을 선택하였다. 선택된 자질을 각각 피드 포워드 신경망과 합성곱 신경망의 학습에 사용하여 두 신경망의 성능을 비교하였다. 합성곱 신경망 학습을 위해 의료 데이터를 고정된 크기의 이미지로 변환하였으며 합성곱 신경망이 피드 포워드 신경망을 이용한 것보다 성능이 좋았다. 합성곱 신경망의 사망률 예측 성능으로 테스트 데이터에 대해 F1 점수는 56.9, AUC는 92.1을 각각 얻었다.

키워드 : 사망률 예측, 합성곱 신경망, 랜덤 포레스트, 자질 선택, 딥러닝

1. 서 론

오늘날 인구의 고령화로 인해 발생하는 노인환자들은 의료 분야에서 해결해야 할 주요 이슈 중 하나이다. 특히 응급실을 방문하는 노인환자의 수는 점점 증가하기 때문에 향후 노인환자들을 위한 의료 서비스의 중요도는 더욱 높아질 것으로 보인다[1].

노인환자들은 대체로 일반 성인 환자와 달리 복잡하고

다양한 의학적, 사회적 및 신체적 문제를 가지고 있어 응급실을 방문하였을 때 사망에 이르는 위험성이 더 높다[2]. 이러한 복합적인 문제들은 노인환자에 대한 정확한 진단과 치료를 방해하기 때문에 의료분야에 종사하는 의사나 간호사들도 이들의 치료 및 사망률을 예측하기는 쉽지 않다[3, 4].

현재 응급실에 내원하는 노인환자의 사망률을 예측하거나 이들의 위험성을 계층화(stratification) 할 수 있는 공인된 도구는 존재하지 않는다. 만약 응급실을 방문한 당시 노인환자의 의료 데이터를 기반으로 학습한 서비스가 개발된다면 조금 더 생명이 위급한 환자들에게는 보다 더 적극적인 치료를 수행하여 한정된 의료 자원을 효율적으로 배분할 수 있다. 또 이러한 서비스는 의사와 간호사들이 노인환자들을 더 정확하게 치료할 수 있도록 기여한다. 우리는 65세 이상 노인환자들이 응급실을 방문한 당시의 의료 데이터를 각각 Feed

* 이 논문은 2019학년도 한국교통대학교의 해외파견연구교수지원금을 받아 수행한 연구임.

† 비 회 원 : 한국교통대학교 컴퓨터정보공학과 석사

†† 정 회 원 : 한국교통대학교 컴퓨터정보공학전공 교수

Manuscript Received : July 16, 2020

First Revision : August 7, 2020

Accepted : August 26, 2020

* Corresponding Author : Songwook Lee(lesw@ut.ac.kr)

Forward Neural Network(FFNN)와 Convolutional Neural Network(CNN)에 학습하여 이들의 사망률을 예측하고 두 신경망의 성능을 비교한다. 추가로 사망률과 좀 더 연관성이 높은 자질을 선택하는 모델을 새롭게 제안한다.

본 논문의 구성은 다음과 같다. 2장에서 응급실을 방문하는 환자의 사망률을 예측하거나 환자의 특정 질병의 유무를 판단하는 연구들을 살펴본다. 3장에서 우리의 이전 연구[5]에서 제안한 사망률 예측 시스템과 새롭게 제안하는 시스템을 비교하여 설명한다. 4장에서 연구 결과를 보이며 5장에서 결론을 맺는다.

2. 관련 연구

환자의 사망률을 예측하는데 있어 다양한 정보들이 필요로 하며 이는 환자의 성별, 연령, 체온, 심박 수 등의 기초적인 정보뿐 아니라 과거 병력, 다양한 혈액 측정 검사 및 배양 검사 결과, MRI, CT 촬영과 같은 영상 데이터 정보 등 다양하고 복잡한 의료 정보들로 구성된다. 학습에 사용되는 데이터는 크게 각 병원 내 입원한 환자의 상태를 기록한 의료 정보, MIMIC(Multi parameter Intelligent Monitoring in Intensive Care) II Clinical Database 등 환자의 상태가 기록된 정보, PubMed, EMBASE 등의 의학사전정보로 구성된다. 이러한 정보들을 회귀분석(Logistic Regression), 지지벡터기계(Support Vector Machine), 인공신경망(Artificial Neural Network)등을 이용하여 학습한 후 환자의 사망률을 예측하는 연구들이 진행되어왔다.

[5]는 응급실을 방문한 65세 이상 노인환자 758명을 각각 지지벡터기계와 FFNN에 학습하여 사망률을 예측하고 그 성능을 비교하였다. 학습에 사용한 자질은 환자의 성별, 연령, 증상 명, 과거 병력, 피검사, 균 배양검사 등 총 89의 자질을 사용하였다. 사망률과 생존율 예측에서 FFNN이 지지벡터기계보다 나은 성능을 보였다.

[6]은 Medicare claims data에서 4가지 질환(심부전, 치매, 만성 폐쇄성 폐질환 및 종양)을 지닌 65세 이상의 환자를 대상으로 한다. 사망률 예측에는 총 6가지(나이브 베이저안, 지지벡터기계, k-NN, 인공신경망, 회귀분석, 랜덤 포레스트)의 모델을 이용한다. 환자의 연령, 성별 등 정적인 데이터만을 학습의 자질로 사용하거나 환자의 질병이 발생한 시점부터 1년간 순차적으로 환자의 정보를 기록한 시계열 데이터를 함께 자질로 사용하여 성능을 비교한다. up-sampling, down-sampling을 도입하여 생존, 사망 데이터 집합의 불균형을 맞춘 후, 무작위로 뽑은 학습집합을 이용하여 파라미터를 최적화하고 검증집합을 이용하여 성능을 평가한다. rbf(Radial Basis Function) 커널의 지지벡터기계를 사용하였고, 2개의 은닉층으로 구성된 인공신경망을 사용하였으며, 1,000개의 트리로 구성된 랜덤 포레스트를 사용하여 각 모델을 학습한다. 그 결과 대체로 시계열 데이터를 포함한 자질이 좋은 성능을 보이며 4가지 질병 중 종양을 지닌 환자의 사망률 예측이 가

장 높은 성능을 보인다. 따라서 시계열 데이터와 사망률의 주요 원인이 되는 자질은 사망률을 예측하는데 큰 역할을 한다.

[7]은 65세 이상 노인환자가 응급실을 방문한 당시의 의료 데이터와 측정된 적혈구 크기 분포 폭(RDW, Red cell Distribution Width)을 자질로 사용하여 회귀분석으로 사망률을 예측한다. 생존자의 적혈구 크기 분포 값은 대체로 낮았고 이 값이 낮을수록 혈액 내의 적혈구의 크기가 고르다. 반대로 사망자는 이 값이 높았으며, 실제로 입원 당시의 환자의 적혈구 크기 분포는 병원 내 사망률의 모든 원인들과 연관이 있음을 보인다. [8]은 패혈증(sepsis)을 지닌 환자를 MIMIC II Clinical Database에서 추출하여 회귀분석으로 사망률을 예측한다. 그 결과 시간에 따라 순차적으로 기록된 환자의 심박 수와 수축기혈압, 이완기혈압이 패혈증을 예측할 수 있는 중요한 자질임을 밝혔다.

환자의 의료 정보는 성별, 연령, 체온 등 단순한 정보 뿐 아니라 전문적인 의학 용어와 혈액, 배양, 소변 등 다양한 검사로부터 나오는 수많은 측정값들로 구성된다. 이러한 특성들로부터 자질을 추출할 때 사망률을 예측하거나 특정 질병들을 진단하는데 있어 불필요한 자질을 제거하는 것이 중요하다. [9]는 Wisconsin Breast Cancer Diagnosis and Prognostic Dataset(WBCDD, WBCPD)에서 유방암을 지닌 환자를 분류하는데 있어 학습에 불필요한 자질들을 랜덤 포레스트를 이용하여 제거한다. 공정함을 위해 무작위로 실험 집합을 10개의 학습 및 평가 집합으로 나누는 후 자질의 중요도를 계산하며, 이 때 랜덤 포레스트의 지니 계수(GINI index)를 이용한다. 모든 자질을 사용한 모델과 중요도가 높은 자질들을 조합한 모델과의 성능을 검증 집합을 이용하여 비교한다. 실험 결과 유방암을 진단하는 데 있어 불필요한 자질을 제거한 모델의 성능이 모든 자질을 사용한 모델보다 우수하다.

X-Ray와 CT 촬영 정보와 같은 의료 영상 데이터를 CNN에 학습하여 증상의 유무를 판단하는 연구들이 진행되었다. 의료 영상 데이터는 수집되는 양이 불충분하고, ImageNet[10]에서 제공하는 수많은 영상 데이터와는 특징이 달라 과적합(over-fitting) 문제가 발생하기 쉬워 일반화된 모델을 구축하기 어렵다. 따라서 ImageNet 데이터를 기반으로 학습한 AlexNet[11], GoogLeNet[12] 등을 사용하지 않고 새롭게 CNN을 설계하여 의료 영상 데이터를 학습한다. CNN을 설계할 때 layer의 개수, 활성화 함수, 파라미터 등 고려해야 할 사항들이 많기 때문에 다양한 시도가 필요하다.

[13]은 폐 결절(lung nodules) CT 영상 데이터베이스인 Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI)[14]로부터 총 1,280장의 폐 결절 이미지를 집합으로 구축한다. 인공신경망에 학습하기에는 집합의 크기가 작기 때문에 집합을 구성하는 이미지를 회전하거나 크기를 조정하고, 인공으로 폐 결절 이미지를 생성하는 등 다양한 방법으로 집합을 확장한다. 집합별로 학습집합과 평가집합을 같은 비율로 분류한 후, 같은 조건으로

CNN에 학습하여 집합의 변형이 학습에 어떠한 영향을 주는 지 확인한다. 그 결과 상대적으로 집합의 크기가 작은 1,280 장의 원본 이미지만을 학습하는 것보다 원본 이미지를 회전 혹은 크기를 조정하여 집합의 크기를 확장한 것이 불충분한 데이터의 양을 늘려 효과적인 성능을 보인다.

본 연구는 [5]에서 제안한 FFNN을 사망률 예측의 기본 모형으로 삼고, 랜덤 포레스트를 이용하여 사망률 예측에 기여도가 높은 자질을 선택한 후, 이를 FFNN과 CNN에 각각 학습한다. CNN 학습을 위해서 선택된 자질의 의료 데이터를 이미지로 변환하였으며 그 결과를 FFNN의 성능과 비교한다.

3. 사망률 예측 시스템

본 연구에서는 [5]에서 사용한 의료 데이터에 특정 질환 정보와 환자수가 추가로 수집된 데이터를 사용한다. 이 의료 데이터는 응급실을 방문한 65세 이상 노인환자들의 의료 데이터이며 환자의 진료 정보를 순차적으로 기록한 시계열 데이터는 존재하지 않는다.

Fig. 1은 의료 데이터를 이용하여 사망률 예측 시스템을 구축하는 과정을 나타낸다.

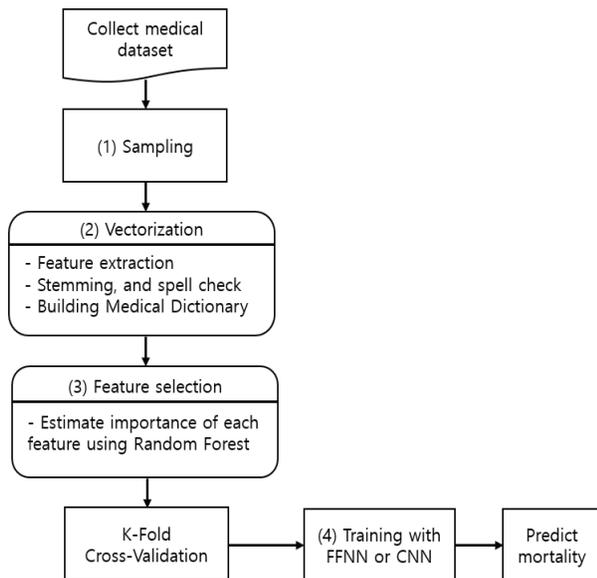


Fig. 1. Overall View of the Proposed System

3.1 견본 추출

노인환자의 수는 총 3,302명이며 이 중 사망자는 161명에 불과하다. 생존자의 비율이 사망자보다 매우 높기 때문에 Fig. 1의 과정 (1)과 같이 생존자의 데이터를 down sampling 하였다.

응급실의 특성상 환자의 특정 데이터가 존재하지 않는 경우가 발생할 수 있다. 예를 들어 응급실을 방문한 환자의 상태에 큰 이상이 없어 피검사, 동맥혈검사와 같은 특정한 검사들을 수행하지 않을 수 있다. 이 경우의 환자의 데이터는 총

분하지 않은 상황이므로 원활한 학습에 방해가 되기 때문에 실험 집합에서 우선적으로 제외한다. 또, 호소하는 증상 명(Chief Complaint)을 갖고 있지 않는 생존자도 실험 집합에서 제외한다.

특정 데이터가 존재하지 않은 생존자를 제거하였으나 그 비율은 사망자에 비해 여전히 높았다. 우리는 생존자의 데이터를 무작위로 제거하여 생존자와 사망자의 비율이 9대1이 되도록 유지한다. 따라서 본 연구에서는 총 1,610명 중 생존자 1,449명 사망자 161명의 의료 데이터를 이용한다.

3.2 자질 추출 및 벡터화

Fig. 1의 과정 (2)에서는 수집된 의료 데이터를 인공 신경망에 학습하기 위한 벡터화(vectorization)한다. 먼저 의료 데이터로부터 사망률에 기여하는 자질을 추출한다. 추출한 자질은 환자의 성별, 연령, 체온, 심박 수 등 기본적인 정보뿐 아니라 과거 병력, 다양한 혈액 검사 및 균 배양검사 등 총 99가지로 구성된다. 이는 기존 연구 [5]에서 사용한 자질에 폐렴, 패혈증, 균혈증 등 특정 질환의 유무가 추가되었다. 추출한 자질의 특성은 수치, 클래스, 단어 총 3가지로 구성된다.

Table 1은 추출한 총 99가지의 자질 중 일부에 대한 설명과 특성을 나타낸다.

Table 1. Type and Description of Features

category (# of features)	description of features	type
Initial Information (17)	Personal Information Temperature ... Chief Complaint(C.C) Blood Pressure(SBP, DBP)	numeric class word
Past History (10)	Diabetes Mellitus(DM) Hypertension(HTN) Coronary Artery Disease(CAD) ... Malignancy	numeric class word
Complete Blood Count (10)	White Blood Count(WBC) ... Hemoglobin, etc	numeric
Blood Chemistry (34)	UA, WBC ... UA, nitrite.,etc	numeric
ABGA (7)	FiO2 ... SaO2	numeric
Culture (9)	Blood Culture ... Sputum Culture	class word
Influenza PCR (2)	Influenza A Virus Influenza B Virus	class
CT (2)	Chest CT Abdomen CT	class
Diagnosis (8)	Pneumonia Sepsis ... Liver abscess	class

사람이 수동으로 의료 데이터를 구축하기 때문에 대소문자 혹은 공백, 특수문자와 같은 실수로 인한 오류를 제거하는 작업은 추출한 자질에 대한 사전을 구축하기 위한 필수적인 과정이다. 특히 환자가 호소하는 증상 명과 과거 병력과 같이 데이터가 전문적인 의료 용어와 약어로 구성되는 경우가 많아 약어의 원형을 복원시켜야 한다. 예를 들어 과거 병력이 ‘agc’, ‘agca’인 경우 두 약어들을 풀어서면 “advanced gastric cancer”를 뜻하며, 단어 사전 구축을 위해 약어의 원형을 복원시킨다.

구축한 사전을 이용하여 벡터 공간에 사상시키며 이는 Table 1과 같이 자질의 특성(c)에 따라 각각의 다른 방법을 적용해야 한다.

자질의 특성이 수치(numeric)인 경우 관찰된 수치들을 0 과 1 사이의 값으로 변환한다. 연구 [5]에서는 최댓값과 최소값을 이용한 MinMaxScaling을 적용하였으나 본 연구에서 측정된 수치 중 몇몇 이상점들을 발견하였다. 우리는 이러한 이상점들로 인하여 편중된 결과가 발생할 수 있다고 판단하여 본 연구에서는 양자화(Quantization)를 적용한다. 먼저 결측치를 제외한 모든 데이터(D)가 관찰되면 이를 오름차순으로 정렬한다. 정렬된 데이터를 나눌 그룹의 개수 k를 지정한 후, Equation (1)과 같이 데이터(x)가 속한 그룹 g을 그룹의 개수 k로 나누어 대푯값 r_k을 취한다. 최종적으로 Equation (2)와 같이 차원이 1인 벡터를 생성하며, 본 연구에서는 그룹 k를 20로 지정한다.

$$r_k = \frac{g}{k} \quad \text{if } D_{k-1} < x \leq D_k \quad k=1,2,\dots,g \quad (1)$$

$$v_{f,c=numeric} = \langle q(x) \rangle \quad q(x) \begin{cases} r_k & x \in dic_f \\ 0 & x \notin dic_f \end{cases} \quad (2)$$

자질의 특성이 클래스(class)인 경우 구축한 사전을 이용하여 Equation (3)과 같이 one-hot방식으로 인코딩한다. 이 때 벡터의 차원은 사전의 크기(d)와 같다.

$$v_{f,c=class.word} = \langle w_1, \dots, w_d \rangle, \quad w_d \begin{cases} 1, & x \in dic_f \\ 0, & x \notin dic_f \end{cases} \quad (3)$$

자질의 특성이 단어(word)인 자질은 증상 명, 과거 암 병력, 배양검사 결과로 구성되며 먼저 공백을 단위로 토큰화한다. 예를 들어 데이터(x) “advanced gastric cancer”는 용어(t)의 개수(s)가 3인 [‘advanced’, ‘gastric’, ‘cancer’]로 구성된다. 이 용어들을 구축된 사전을 이용하여 Equation (3)의 one-hot 방식으로 벡터를 생성한다.

위에서 제시한 방법들을 모두 적용하면 환자(i)에 대하여 Equation (4)와 같이 벡터를 생성할 수 있다.

$$V(i) = v_{f_1c} \oplus v_{f_2c} \oplus \dots \oplus v_{f_ic} \quad (4)$$

3.3 자질 선택

앞에서 살펴본 자질을 모두 사용하면 매우 큰 차원의 벡터가 생성된다. 이러한 자질 중에는 환자의 사망률을 예측하는데 크게 기여하는 자질이 있지만 그렇지 않은 자질은 예측을 방해한다. 우리는 랜덤 포레스트를 이용해서 사망률 예측에 기여하는 주요 자질을 선택한다.

랜덤 포레스트는 학습 데이터의 부분 집합들을 서로 다른 다수의 결정 트리(Decision Tree)에 학습하여 결합한 분류기로 단일 결정 트리에 비해 정확도와 안정성 향상을 도모한 기계학습 방법이다. 우리는 지니 계수를 이용하여 랜덤 포레스트를 구축한다[9]. 지니 계수는 결정 트리를 구성하는 노드에서 선택된 자질을 기준으로 분류할 때 데이터의 불순도(impurity)를 재는 척도로서 자질의 중요도를 파악할 수 있다. 각 자질의 중요도는 지니 계수를 랜덤 포레스트를 구성하는 모든 결정 트리들의 개수로 평균하여 구할 수 있다.

랜덤 포레스트에 입력되는 벡터는 하나의 자질이 하나의 차원을 가져야 한다. 하지만 3.2에서 제시한 벡터화 방법은 자질의 특성이 클래스나 단어인 경우 사전의 크기(d)만큼의 차원을 가진다. Equation (5)를 적용하여 각 자질에 대한 벡터의 차원을 1로 고정한다. 이 때 특성이 단어인 자질은 토큰화하지 않는다.

$$FS(V(i)) = FS(v_{f_1c}) \oplus FS(v_{f_2c}) \oplus \dots \oplus FS(v_{f_ic}),$$

$$FS(v_{f,c}) = \frac{\operatorname{argmax}(v_{f,c})}{d_{f,c}} \quad (5)$$

(3b) Equation (5)를 적용한 벡터를 랜덤 포레스트에 학습하면서 지니 계수를 이용하여 자질의 중요도를 계산한다. 실험은 중요도가 높은 자질을 10개씩 그룹화하면서 모델을 생성하고 모든 자질을 사용한 모델과 성능을 비교한다.

3.4 학습 모형 설계

우리는 FFNN과 CNN을 이용하여 노인환자의 사망률을 예측하고 두 모형의 성능을 비교한다. FFNN 모형은 [5]에서 제안한 구조를 사용하였으며 Fig. 2와 같다.

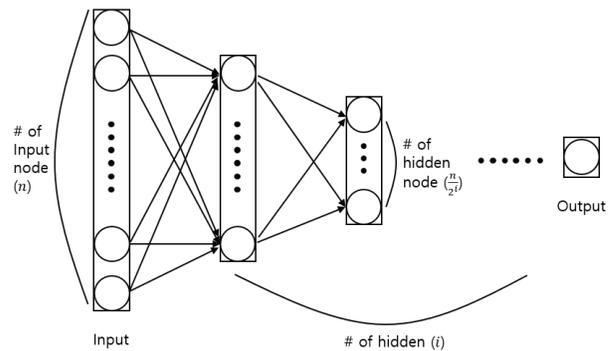


Fig. 2. A structure of Feed Forward Neural Network

Fig. 2와 같이 FFNN의 경우 입력층의 노드 수(n)는 벡터의 차원수와 같으며 다음 은닉층의 노드 수는 이전 층의 50%을 사용한다. 본 연구는 생존 혹은 사망을 판단하는 이항 분류 문제를 다루기 때문에 한 개의 출력층으로부터 0과 1 사이의 값이 발생하는 구조로 설계하며, 값이 0.5 이상인 경우 사망으로 판단한다.

CNN의 입력은 고정된 크기의 이미지를 받기 때문에 우리는 의료데이터 벡터를 선형 보간법(Bilinear interpolation)을 이용하여 $36 \times 36 \times 1$ 의 크기를 지닌 회색조(Grayscale) 이미지로 변환한다. Fig. 3은 의료데이터 벡터를 이미지로 변환한 예시이며, CNN 모형을 Fig. 4와 같이 설계하여 벡터 이미지를 학습한다.



Fig. 3. An Example of Images of Medical Data

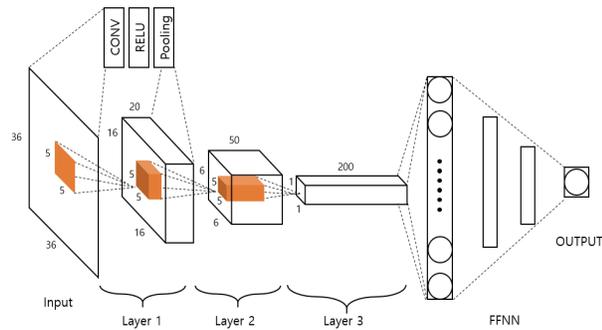


Fig. 4. A Structure of Convolutional Neural Network

Fig. 4와 같이 CNN의 구성은 convolutional, ReLU, pooling 순으로 이루어지며 총 3개의 feature map을 생성한다. convolutional 연산을 수행하는 필터의 크기는 각각 $5 \times 5 \times 1 \times 20$, $5 \times 5 \times 20 \times 50$, $5 \times 5 \times 50 \times 200$ 이며, 연산을 수행하는 간격(stride)을 1로 모두 통일하고 zero-padding을 적용한다.

convolutional 연산을 마친 후 ReLU를 통과하고 pooling 연산을 수행한다. pooling에 사용하는 필터의 크기는 모두 2×2 로 통일하며, 연산을 수행하는 간격(stride)은 2로 통일하고 가장 큰 값을 선택하는 max-pooling을 적용한다.

최종적으로 $1 \times 1 \times 200$ 의 feature map이 생성되며 이를 Fig. 2의 FFNN 모형의 입력으로 받아 결과를 출력한다.

4. 실험 및 결과

우리는 랜덤 포레스트를 통해 주요 자질을 추출한 모델과 모든 자질을 사용한 모델을 3.4절에서 제안한 FFNN과 CNN

에 학습하여 두 신경망의 성능을 비교한다. 실험에 사용한 컴퓨터 사양은 Ubuntu 16.04 LTS, Intel(R) Core(TM) i5-7500 3.40GHz, 16GB DDR4, GeForce GTX 1080이었으며, 랜덤 포레스트와 인공 신경망은 각각 scikit-learn과 Tensorflow-gpu 1.10.1을 이용하여 구현하였다.

Table 2는 실험에 사용한 랜덤 포레스트와 FFNN의 기본 파라미터 값을 나타낸다.

Table 2. Basic Parameter Values in the Experiment

method	parameter	
Random Forest	# of trees	400
	# of max features	\sqrt{k}
FFNN	# of input node	n
	# of hidden node	$\frac{n}{2^i}$
	# of output node	1
	weight initialization	Xavier
	Activation Function	ReLU
	Optimizer	Adam

Fig. 5는 3.3절에서 제시한 랜덤 포레스트를 구축한 결과 환자의 사망에 영향을 주는 상위 10개의 자질을 나타낸 그래프이다. 랜덤 포레스트 분류기 자체의 성능은 baseline 신경망 시스템의 성능보다 낮아 랜덤 포레스트는 각 자질의 중요도 계산에만 활용하였다. 자질의 중요도는 소변검사를 통한 아질산염 양상반응 유무(UA, nitrite), 젖산탈수소효소(LDH), 혈액요소질소(BUN) 순으로 높았다. 특히 상위 10개의 자질은 패혈증(Sepsis)을 제외하면 모두 혈액 내 성분들의 수치들로 구성되었다.

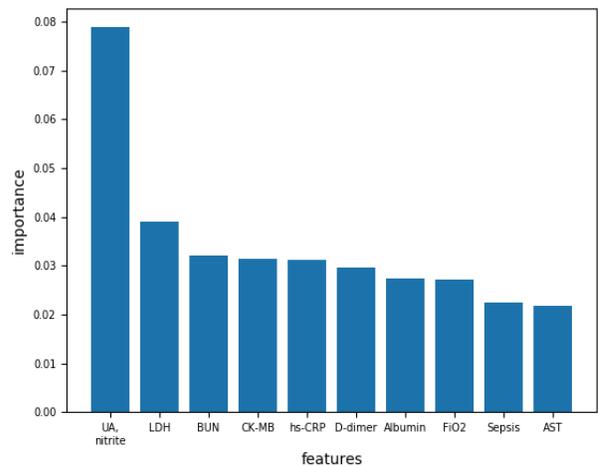


Fig. 5. Top 10 Important Features

Table 3은 5-fold 교차검증에 사용된 환자의 수를 나타낸다. 실험에 사용된 노인환자는 총 1,610명이며 이 중 1,449명의 생존자와 161명의 사망자의 데이터로 구성된다.

Table 3. The Number of Patients in the Set

set		# of survivals	# of deaths	# of total
set 1	train	1,160	128	1,288
	test	289	33	322
set 2	train	1,159	129	1,288
	test	290	32	322
set 3	train	1,164	124	1,288
	test	285	37	322
set 4	train	1,153	135	1,288
	test	296	26	322
set 5	train	1,160	128	1,288
	test	289	33	322

의학계에서의 시스템 성능은 진단율과 오진단을 모두를 낮추는 것이 중요하며 주로 AUC(Area Under Curve)를 평가 척도로 사용한다. 우리는 사망률을 기준으로 시스템의 성능을 평가하며 AUC와 F1 점수를 사용한다.

모델에 상관없이 공정함을 위해 학습은 모두 동일하게 500 epoch에서 종료하며 mini-batch의 크기는 256로 고정한다. 선택된 자질의 수에 따라 입력 벡터의 차원이 다르기 때문에 차원에 비례하도록 은닉 층의 수를 1~5로 설정하며, 학습 진도율(learning rate)을 1E-03~1E-05로 설정한다. 또, 과적합(over-fitting)을 막기 위해 FFNN의 은닉 층마다 dropout을 0.7로 설정한다.

우리는 모든 자질을 사용한 모델을 baseline으로 지정한다. 최적의 자질의 개수를 파악하기 위해 3.3에서 제시한 방법을 이용하여 중요도가 높은 자질을 상위 10개씩 사용하여 때마다 사망률을 기준으로 성능을 평가하였고 Table 4에 나타낸다. 이때 사용한 신경망 모형은 FFNN이다.

Table 4. The Result of Mortality Prediction Based on the Number of Features

The number of features (# of dimension)	F1	AUC	# of hidden layer	learning rate	training time (sec)
10 (13)	51.2	92.0	1	1E-03	18.11
20 (23)	52.6	92.1	1	1E-03	18.38
30 (33)	52.6	93.3	1	1E-03	19.58
40 (43)	52.3	92.1	1	1E-03	19.70
50 (336)	56.3	91.7	4	7E-05	49.27
60 (422)	55.4	91.9	4	5E-05	52.17
70 (583)	56.6	91.7	5	5E-05	71.72
80 (605)	57.0	91.8	5	5E-05	74.24
90 (622)	56.1	92.0	5	5E-05	74.84
99 (640) (baseline)	54.7	91.6	5	4E-05	74.18

실험 결과 baseline의 F1 점수는 54.7로 연구 [5]의 F1 점수인 52.0보다 2.7 높으며, baseline의 AUC는 91.6으로 연구 [5]의 AUC인 88.6보다 3.0 높았다. 본 실험에 사용한 노인환자는 총 1,610명으로 연구 [5]에서 사용한 758명보다 약 2.1배 많아 직접적인 비교는 어려우나 유의미한 성능 차이가 보이는 것을 고려하면 본 연구에서 제안하는 벡터화 방법이 좀 더 우수한 성능을 거뒀다고 할 수 있다.

AUC를 기준으로 가장 좋은 성능을 보인 자질 개수는 상위 30개를 사용했을 때이며, 이 때 AUC는 93.3이다. 하지만 F1 점수가 52.6으로 baseline보다 낮으며, 벡터의 차원이 33으로 다수의 은닉층을 사용하기에 적합하지 않았다.

전체적으로 상위 80개의 자질을 사용하였을 때 성능이 가장 좋았으며, 이 때 F1 점수는 57.0, AUC는 91.8이다. 우리는 상위 80개의 자질을 최적의 자질 개수로 판단하며, 이 모델을 FSRF(Feature Selection using Random Forest)라 지정한다.

Fig. 4에서와 같이 CNN의 경우 36×36×1의 고정된 크기의 이미지를 입력으로 받기 때문에 자질의 개수와 상관없이 차원이 1,296로 동일하다. CNN은 최종적으로 200차원의 벡터를 FFNN으로 넘겨주기 때문에 CNN에 사용되는 FFNN의 은닉층의 수는 3, 학습 진도율은 2E-05로 통일하였다.

Table 5는 생존율과 사망률을 기준으로 성능을 평가한 결과이고, Table 6은 Table 5의 결과를 평균하여 2-class 분류기로서의 성능을 평가한 결과이다.

Table 5. The Result Based on Survival and Death Rate

model		based on survival rate			based on death rate		
		P	R	F1	P	R	F1
FFNN	baseline	94.9	95.4	95.1	57.2	54.2	54.7
	FSRF	95.3	95.0	95.1	56.7	58.4	57.0
CNN	baseline	95.1	95.5	95.2	59.1	55.8	56.1
	FSRF	95.3	95.3	95.3	57.5	57.5	56.9

Table 6. The Result of 2-class Classification

model		P	R	F1	Acc	AUC	training time (sec)
FFNN	baseline	76.0	74.8	74.9	91.2	91.6	74.18
	FSRF	76.0	76.7	76.1	91.2	91.8	74.24
CNN	baseline	77.1	75.7	75.7	91.4	91.6	127.07
	FSRF	76.4	76.4	76.1	91.5	92.1	131.12

Table 5에서 사망률 예측 결과가 생존율 예측보다 저조한 이유는 실험 집합의 생존자와 사망자의 비율이 9:1로 편중된 결과로 분석된다. 사망률 예측 기준으로 CNN의 경우 FSRF와 baseline의 성능 변화가 크게 나지 않았지만 FFNN

의 경우 FSRF의 재현율은 58.4로 baseline의 재현율인 54.2보다 4.2 높았다. 비록 정밀도는 baseline에 비해 0.5 낮았지만 결과적으로 F1 점수가 3.3 높으며, 이는 사망률 예측에 기인하지 않는 불필요한 자질을 제거한 결과로 볼 수 있다.

Table 6에서 전반적으로 CNN이 FFNN보다 학습은 더 오래 걸리지만 성능은 약간 높다는 것을 알 수 있다. 특히 baseline의 경우 CNN과 FFNN의 AUC는 차이가 없지만, 다른 평가 척도들의 결과가 CNN이 FFNN보다 좋았다. 전체적인 성능이 가장 높은 모델은 CNN의 FSRF이며, 사망률 예측 기준으로 FFNN의 FSRF에 비해 F1 점수는 0.1 낮지만 AUC가 92.1로 가장 높다. 이는 연구 [5]의 F1 점수는 4.9, AUC는 3.5 더 높다.

Fig. 6은 CNN의 FSRF의 결과를 ROC 곡선으로 나타낸 그림이다.

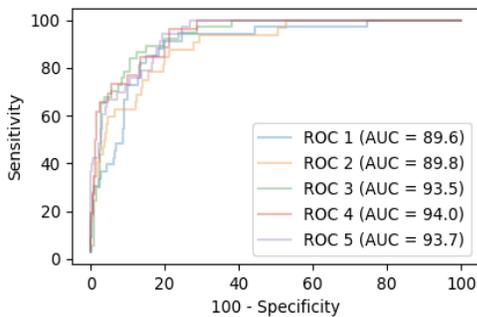


Fig. 6. The ROC Curves of FSRF Model using Convolutional Neural Network

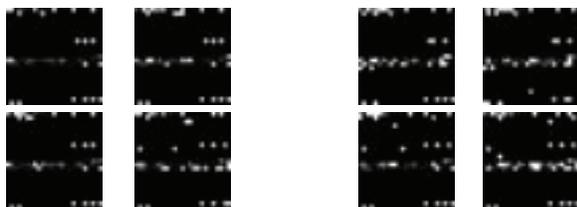


Fig. 7. Images of Medical Data of Survivor (left side) and Dead (Right side)

Fig. 7은 CNN의 FSRF를 시각적으로 표현한 예시이며 왼쪽의 4장의 이미지는 생존 데이터, 오른쪽의 4장의 이미지는 사망 데이터이다. 일반적인 영상 데이터와 달리 의료 데이터를 이미지화한 영상 데이터는 생존과 사망의 구별이 시각적으로 뚜렷하지 않아 이미지의 특성을 살리기 쉽지 않다. 이는 CNN이 FFNN에 비해 성능이 크게 향상되지 않은 이유로 해석된다.

5. 결 론

본 연구는 노인환자가 응급실을 방문한 당시의 의료 데이터로부터 노인환자의 사망률을 결정하는 데 크게 기여하는

자질을 선택하기 위해 중요도를 계산하였고, 랜덤 포레스트의 지니 계수를 이용하였다. 학습에 사용한 총 99개의 자질 중 중요도가 높은 자질을 10개씩 사용하면서 FFNN에 학습하였다. 그 결과 중요도가 높은 상위 80개의 자질을 사용하였을 때 최적의 개수임을 보였고, 이를 각각 FFNN과 CNN에 학습하였다. 선택된 자질로 학습한 두 신경망 중 CNN의 성능이 FFNN보다 더 높았다. 이 때, F1 점수는 56.9, AUC는 92.1로 비교 연구보다 각각 4.9, 3.5 높았다.

사실 신경망을 학습하는 과정에서 중요한 자질의 가중치는 높게 중요하지 않은 자질의 가중치는 낮게 학습되기 때문에 자질의 중요도를 따로 계산할 필요는 없어 크게 중요하지 않다. 하지만 우리는 일반적으로 신경망 학습에 사용하는 데이터와 다른 특성을 지닌 의료데이터를 좀 더 효율적으로 신경망에 학습하는 방법을 모색하기 위해 랜덤 포레스트를 이용하였다. 또, 노인환자는 일반 환자와 달리 폐렴, 패혈증, 패혈성 쇼크 등 일반적으로 사망과 직관적이라 생각되는 자질들보다 아질산염 양상반응 유무, 젖산탈수소효소, 혈액요소질소 등 혈액 내의 측정된 값들이 더 사망률과의 연관성이 높다는 것을 알 수 있었다. 향후 자질의 중요도를 좀 더 정확하게 측정한다면 의사나 간호사들이 노인환자의 비정상적인 생리학적 반응을 더 정확하게 예측할 수 있도록 도움을 줄 것으로 기대된다.

결과적으로 사망률 예측의 성능은 생존율 예측보다 저조하며, 이는 수집된 사망 환자의 데이터의 양이 생존 환자의 데이터보다 작으면서 발생하는 편중된 결과로 보인다. 또 수집된 데이터에는 시계열 데이터가 존재하지 않아 보다 다양한 신경망 모형을 설계하기 쉽지 않다. 향후 환자의 데이터 중 MRI나 CT 영상 데이터 등 의료 영상 데이터가 추가된다면 제안된 신경망 모형에 더 다양한 신경망 모형을 결합한다면 성능의 향상을 기대할 수 있을 것이다.

References

- [1] Richard Niska, Farida Bhuiya, and Jianmin Xu, "National Hospital Ambulatory Medical Care Survey: 2007 emergency department summary," *National Health Stat Report*, Vol.26, No.26, pp.1-31, Aug. 2010.
- [2] Lorraine Mion, Robert Palmer, Georgia J. Anetzberger, and Stephen W. Meldon, "Establishing a casefinding and Referral System for At-risk Older Individuals in the Emergency Department Setting: the SIGNET Model," *Journal of the American Geriatrics Society*, Vol.49, No.10, pp.1379-1186, Oct. 2001.
- [3] Nikolaos Samaras, Thierry Chevalley, Dimitrios Samaras, and Gabriel Gold, "Older Patients in the Emergency Department: A Review," *Annals of Emergency Medicine*, Vol.56, No.3, pp.261-269, Sep. 2010.

- [4] G. R. Strange and E. H. Chen, "Use of emergency departments by Elder Patients: a Five-year Follow-up Study," *Annals of Emergency Medicine*, Vol.5, No.12, pp.1157-1162, Dec. 1998.
- [5] Junhyeok Park and Songwook Lee, "Mortality Prediction of Older Adults Admitted to the Emergency Department," *KIPS Transactions on Software and Data Engineering (KTSDE)*, Vol.7, No.7, pp.275-280, Jul. 2018.
- [6] Maggie Makar, Marzyeh Ghassemi, David Cutler, and Ziad Obermeyer, "Short-term Mortality Prediction for Elderly Patients Using Medicare Claims Data," *International journal of Machine Learning and Computing*, Vol.5, No.3, pp. 192-197, Jun. 2015.
- [7] Soo Hyun Kim, Jeong Hoon Yeon, Kyu Nam Park, Sang Hoon Oh, Seung Pill Choi, Young Min Kim, Han Joon Kim, and Chun Song Youn, "The Association of Red Cell Distribution Width and in-hospital Mortality in Older Adults Admitted to the Emergency Department," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, Jun. 2016.
- [8] Supreeth P. Shashikumar, Matthew D. Stanley, Ismail Sadiq, Qiao Li, Andre Holder, Gari D. Clifford, and Shamim Nemati, "Early Sepsis Detection in Critical Care Patients using Multiscale Blood Pressure and Heart Rate Dynamics," *Journal of Electrocardiology*, Vol.50, No.6, pp.739-743, Aug. 2017.
- [9] Cuong Nguyen, Yong Wang, and Ha Nam Nguyen, "Random Forest Classifier Combined with Feature Selection for Breast Cancer Diagnosis and Prognostic," *Journal of Biomedical Science and Engineering*, Vol.6, No.5, pp.551-560, May 2013.
- [10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale Hierarchical Image Database," *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," in *Neural Information Processing Systems*, 2012.
- [12] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition 2015*.
- [13] He Yang, Hengyong Yu, and Ge Wang, "Deep Learning for the Classification of Lung Nodules," arXiv preprint, arXiv:1611.06651, 2016.
- [14] Samuel G Armato III, "The lung image database consortium (LIDC) and Image Database Resource Initiative (IDRI): a Completed Reference Database of Lung Nodules on CT Scans," *Medical Physics Online*, Vol.38, No.2, pp.915-931, Feb. 2011.



박 준 혁

<https://orcid.org/0000-0002-3653-9425>

e-mail : jhpark@a.ut.ac.kr

2017년 한국교통대학교 컴퓨터정보공학과 (학사)

2019년 한국교통대학교 컴퓨터정보공학과 (석사)

2020년 ~ 현 재 포터투마루 연구원

관심분야 : 자연언어처리, 기계학습, 의미분별, 인공지능



이 성 욱

<https://orcid.org/0000-0002-6224-4241>

e-mail : leesw@ut.ac.kr

1996년 서강대학교 전자계산학과(학사)

1998년 서강대학교 컴퓨터학과(석사)

2003년 서강대학교 컴퓨터학과(박사)

2003년 ~ 2004년 서강대학교 산업기술연구소 연구원

2003년 ~ 2005년 서강대학교 정보통신대학원 대우교수

2004년 ~ 2005년 LG전자 기술원 선임연구원

2005년 ~ 2007년 동서대학교 컴퓨터공학과 전임강사

2007년 ~ 현 재 한국교통대학교 컴퓨터정보공학전공 교수

관심분야 : 자연언어처리, 대화인터페이스, 기계학습, 인공지능