

Mortality Prediction of Older Adults Admitted to the Emergency Department

Junhyeok Park[†] · Songwook Lee^{**}

ABSTRACT

As the global population becomes aging, the demand for health services for the elderly is expected to increase. In particular, The elderly visiting the emergency department sometimes have complex medical, social, and physical problems, such as having a variety of illnesses or complaints of unusual symptoms. The proposed system is designed to predict the mortality of the elderly patients who are over 65 years old and have admitted the emergency department. For mortality prediction, we compare the support vector machines and Feed Forward Neural Network (FFNN) trained with medical data such as age, sex, blood pressure, body temperature, etc. The results of the FFNN with a hidden layer are best in the mortality prediction, and F1 score and the AUC is 52.0%, 88.6% respectively. If we improve the performance of the proposed system by extracting better medical features, we will be able to provide better medical services through an effective and quick allocation of medical resources for the elderly patients visiting the emergency department.

Keywords : Mortality Prediction, Support Vector Machine, Artificial Neural Network, Deep Learning

응급실 방문 노인 환자의 사망률 예측

박 준 혁[†] · 이 성 욱^{**}

요 약

세계 인구의 고령화가 진행되는 오늘날 노인들을 위한 의료 서비스의 수요는 점차 증가할 것으로 보인다. 특히, 응급실을 방문하는 노인 환자는 일반 환자보다 다양한 질병을 갖고 있거나, 특이한 증상을 호소하는 등 복잡한 의학적, 사회적 및 신체적 문제를 가지고 있는 경우가 많다. 우리는 65세 이상의 응급실을 방문한 노인 환자의 사망률 예측을 위해 연령, 성별, 혈압, 체온, 혈액검사, 주증상명 등의 의료 데이터를 사용하였다. Feed Forward 신경망과 지지벡터기계를 각각 학습하여 사망률을 예측하고 그 성능을 비교하였다. 1개의 은닉층을 사용한 Feed Forward 신경망의 실험 결과가 가장 좋았으며, 이 때 F1 점수는 52.0%, AUC는 88.6%이다. 좀 더 좋은 의료 자질을 추출하여 제안 시스템의 성능을 향상시킨다면 응급실에 방문한 노인 환자들을 위한 효과적이고 신속한 의료 자원 배분을 통해 더 좋은 의료 서비스를 제공할 수 있을 것이다.

키워드 : 사망률 예측, 지지벡터기계, 인공신경망, 심층망

1. 서 론

전 세계적으로 고령화가 진행되는 가운데 노인 환자들은 복잡한 의학적, 사회적 및 신체적 문제를 가지고 있다[1]. 특히 우리나라의 인구 고령화 속도는 OECD 국가 중 가장 빠른 것으로 관측되며, 노인들을 위한 의료 서비스에 대한 수요가 현재보다 더욱 증가할 것으로 예상된다.

노인 환자들은 일반 성인 환자들보다 응급실을 자주 방문

하고 그 수는 해마다 증가하는 추세이다[2]. 특히 노인 환자들은 일반 성인 환자들과 비교하였을 때 위급한 수준이 더 높고, 응급실에 더 오래 머무르는 경향이 있다[3]. 이는 노인 환자들이 일반 성인 환자들보다 더 많은 질병을 가지고 있거나, 많은 약물을 복용 중이며, 비정형 징후 및 증상을 호소하여 정확한 진단과 치료 및 사망 예측을 복잡하게 하기 때문이다[4]. 의사나 간호사들도 응급실을 방문하는 노인 환자에 대한 처치와 치료가 일반 성인 환자들보다 더 어렵게 느끼고 있으며, 이러한 노인 환자에 대한 진단과 치료의 복잡성 때문에 이들의 사망률을 예측하기는 쉽지 않다[5]. 따라서 노인 환자들은 응급실에서 일반 병동으로 옮겨 장기간 치료를 받거나, 중환자실에서 집중 치료를 받거나, 사망에 이르는 위험이 일반 성인 환자보다 크다고 할 수 있다.

※ 이 논문은 한국연구재단(NRF-2017R1D1A1A02019411)의 지원을 받았음.

† 비 회 원 : 한국교통대학교 컴퓨터정보공학과 석사과정

** 정 회 원 : 한국교통대학교 컴퓨터정보공학전공 교수

Manuscript Received : April 16, 2018

Accepted : April 29, 2018

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

환자의 회복은 응급 질병의 치명성과 전체 치료 과정에서 받은 의료 서비스의 질에 의해 좌우되기 때문에 환자의 병원 내 사망률 또는 위험 계층화(stratification)를 예측하는 문제는 복잡하고 중요한 문제이다. 환자의 초기 정보를 이용하여 사망률을 예측한다면 조기에 적극적인 치료를 수행하여 회복을 높일 수 있을 것이다. 또, 경험이 많지 않은 의사와 간호사들의 적절한 처치에 도움을 줄 수 있을 것이다. 우리는 응급실을 방문한 65세 이상 노인 환자의 의료 데이터를 지지벡터 기계와 인공신경망을 이용하여 각각 학습한 후, 사망률을 예측하는 시스템을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서 응급실 방문 환자의 사망률 혹은 입원 환자의 사망률을 예측하는 연구들을 살펴보고, 3장에서 제안된 사망률 예측 시스템을 설명한다. 4장에서 연구 결과를 보이며 5장에서 결론을 맺는다.

2. 관련 연구

응급실 병원 내 환자의 사망률을 예측하기 위해 환자의 성별, 연령, 내원일시와 같은 기초적인 정보 뿐 아니라, 증상 명, 심박 수, 체온 등의 초기 정보, 과거 병력, 다양한 혈액 측정 검사 결과, 배양 검사 결과, CT 촬영 정보 등 최대한 많은 측정값들을 회귀분석(Logistic Regression), 지지벡터기계(Support Vector Machine), 인공신경망(Artificial Neural Network)등을 이용하여 학습한다. 학습집합에 사용되는 데이터는 각 병원내의 입원한 환자의 기록된 정보나, MIMIC(Multi parameter Intelligent Monitoring in Intensive Care) II Clinical Database 등 이미 측정된 환자의 정보, PubMed, EMBASE 등의 의학사 전정보를 이용하여 구축한다.

환자의 증상을 표기한 텍스트 정보는 대체로 불완전한 문장 구조를 지니며, 다양하고 풍부한 의료 전문 용어가 사용되는 특성 때문에 자연언어처리 기술을 적용시키기 쉽지 않다. 따라서 연령, 성별과 같은 정적 데이터와 심박 수, 체온 등과 같이 임의로 지정한 기간 동안 측정할 때 마다 변화하는 시계열 데이터를 적절히 활용해야 한다.

[6]은 연령, 성별, 입원시간, 증상 명, 혈압, 심박 수, 체온, 활력징후 등 총 8,607회의 입원을 통해 얻은 다양한 데이터를 기반으로 Multi-Variable Logistic Regression을 이용하여 사망률을 예측한다. 그 결과 호흡기 질환을 지닌 환자가 심혈관 질환, 신경계 질환, 외상을 입은 환자보다 사망 위험이 높았고, 위장 또는 비정상적인 문제가 있는 환자는 사망 위험이 높았다.

환자의 기록 정보 중 사망률의 주요 원인이 되는 특정한 자질은 사망률을 예측하는데 큰 역할을 한다. [7]은 [6]과 같이 1,990명의 응급실을 방문한 65세 이상 노인 환자의 데이터와 입원 당시의 측정된 RDW(Red-cell Distribution Width) 값을 기준으로 사망률을 예측한다. 실제로 비 생존자의 RDW 값은 생존자보다 더 높았으며, 연구 결과 입원 당시의 RDW 값이 병원 내 사망률의 모든 원인들과 관련이 있음을 나타낸다. [8]은 MIMIC II Clinical Database에서 추출한 2,155명의 sepsis(패혈증)를 지닌 환자의 측정값들을 추출한 후 회귀분

석으로 사망률을 예측하면서, 시간에 따라 순차적으로 기록된 환자의 Heart Rate(HR)와 Blood Pressure(BP)가 패혈증을 예측할 수 있는 중요한 자질임을 밝혔다. 특히 패혈증과 관련된 사망률은 시간이 지연됨에 따라 증가하기 때문에 환자의 보이지 않는 미세한 증상들을 이용하여 조기에 패혈증을 예측하여 진단하는 것이 중요하다.

[9]는 Medicare claims data에서 4가지 질병(심부전, 치매, 만성 폐쇄성 폐질환 및 종양)을 지닌 65세 이상의 환자를 대상으로 나이브 베이저안, 지지벡터기계, k-NN, 인공신경망, 로지스틱 회귀분류, 랜덤 포레스트 등의 분류기를 이용하여 사망률을 예측하고 비교한다. 환자를 나타내는 지표는 질병이 발생한 시점부터 3개월 혹은 12개월까지의 순차적으로 표현되며, 병원을 방문하는 횟수나, 의사와 간호사와의 진단 건수가 포함된다. 생존, 사망 데이터 집합의 불균형을 맞추기 위해 up-sampling, down-sampling을 하며, 무작위로 뽑은 학습집합을 튜닝집합으로 매개변수를 최적화하고 검증집합으로 평가한다. 랜덤 포레스트의 경우 1,000개의 트리를 사용하였고, 인공신경망의 경우 2개의 은닉층을 사용하였으며, 지지벡터기계의 경우 rbf(Radial Basis Function) 커널을 사용하여 각 분류기의 모델을 최적화한다. 4가지 질병 중 종양을 지닌 환자의 사망률 예측과, Balanced Random Forest(BRF)를 이용한 사망률 예측이 가장 높은 성능을 보인다.

[10]은 MIMIC II Clinical Database에서 주어진 시간동안 측정된 환자의 정보들을 추출하여 HMM(Hidden Markov model)과 LDA(Latent Dirichlet allocation)를 병합한 모델을 구축한 후 지지벡터기계로 환자의 사망률을 예측하는 시스템을 제안한다. 기록 정보는 가족이 환자에 대한 지원과 정신적인 건강까지 일반적으로 관찰되지 않는 숨겨진 정보들을 포함한 시계열 데이터로서 다음 상태로 전이되는 정보가 담겨 있다. 데이터의 특성 상 하루 혹은 일주일 분량의 단기적인 정보보다 한 달 이상의 장기적인 정보가 기록된 환자의 사망률을 예측하는 것이 효과적이다.

[11]은 추출한 환자의 첫 입원 후 48시간동안 수집된 데이터를 12시간 단위로 쪼개어 각 시간별 기록을 단어주머니(bag-of-words) 모델을 이용하여 벡터화 한 후, [10]에서 사용한 LDA와 LSTM(Long Short-Term Memory)를 결합한 모델이 시계열 데이터를 활용한 예측 시스템에서 의미 있는 성능을 보인다. [12]는 48시간동안 수집된 데이터를 k-NN 알고리즘을 이용한 유사성 기반으로 예측한다. 예측된 결과의 오류를 Gradient Descent를 이용하여 오류 감소의 최대치를 찾아내 반복적으로 가중치를 조정함으로써 성능을 향상시켰다. 유사성 기반의 예측 시스템이 일반적인 사망률 예측보다 성능이 높고, 특히 시계열 데이터를 사용하는 것이 사용하지 않은 것 보다 월등한 성능을 보인다.

응급실 방문 노인 환자의 사망률 예측을 위해 이전에는 주로 활력징후(vital sign)에 의존한 통계적 기법과 규칙을 이용한 채점 시스템이 주로 이용되었다. 본 연구는 대표적인 기계 학습방법인 지지벡터기계와 Feed Forward Neural Network(FFNN)를 이용하여 응급실 방문 노인 환자의 사망률을 예측한다. 비시계열(non-time series) 데이터인 정적인 의료 데이

터로부터 추출한 수치 데이터와 단어 데이터를 함께 활용하는 것은 이전 연구에서는 시도되지 않은 방법이다. 환자의 주 증상명과 배양 검사 결과 정보는 문장으로 되어 있으며, 이를 Word2Vec을 사용하여 벡터로 변환하여 다른 수치 데이터들과 함께 사망률 예측에 사용하였다.

3. 사망률 예측 시스템

본 연구는 응급실을 방문한 65세 이상의 노인 환자의 사망률을 지지벡터기계와 인공신경망 모형을 설계하여 사망률 예측 시스템을 설계하며, Fig. 1은 시스템을 구축하는 과정을 나타낸다.

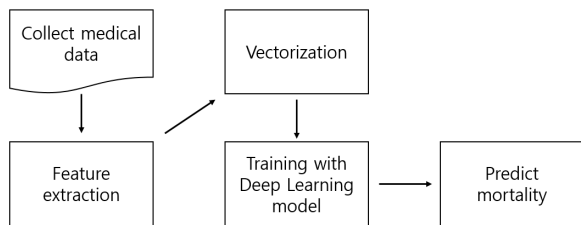


Fig. 1. The Process of Mortality Prediction System

먼저 과거 응급실을 방문한 노인 환자의 의료 데이터를 수집한다. 수집한 의료 데이터로부터 사망률 예측에 기여할 수 있는 적절한 자질들을 추출한 후, 추출된 자질들은 벡터 공간으로 사상시키고 심층망을 설계하여 사망률을 학습시킨다. 최종적으로 학습된 신경망을 이용하여 사망률을 예측한다.

3.1 의료 데이터 수집

응급실 방문 노인 환자의 의료 데이터 수집은 사람이 수동으로 EMR(Electric Medical Record, 전자의료기록)에 있는 정보를 입력함으로써 이뤄지고, 수집된 의료 정보로부터 사망률 예측에 기여할 수 있는 적절한 자질들을 선택하는 것이 중요하다. EMR에는 환자의 상태를 나타내는 다양한 의료 데이터가 있지만, 표준화가 아직 이뤄지지 않았기 때문에 병원마다 서로 다른 EMR 시스템을 구축하여 사용하고 있다. 게다가 개인정보보호, 의료산업기밀 등을 이유로 EMR 데이터를 자동으로 DB로 추출하는 시스템이 거의 도입되지 않아 모든 자료를 사람이 수동으로 수집하고 있다.

3.2 자질 선택 및 추출

본 연구는 환자의 연령, 성별, 증상 명 등 기본적인 정보와 과거 병력, 피검사, 균 배양검사 등 총 89가지의 자질을 이용하여 사망률 예측에 신경망 모형을 설계한다. 하지만 관련 연구에서처럼 지정한 기간 동안 환자의 정보를 기록한 시계열 데이터를 갖고 있지 않다.

수집된 환자는 총 1,661명이며 사망자는 86명이다. 이 중 사망률을 예측에 크게 기여하는 자질이 없는 환자는 제외시킨다. 예를 들어 응급실을 방문한 환자가 수축기 혈압, 이완

기 혈압, 호흡수, 심장 박동 수를 측정하기 어려운 상태인 경우 환자가 이미 사망한 상태이고 환자의 데이터가 많지 않은 상황에서 원활한 학습에 방해가 되기 때문에 집합에서 제외한다. 또, 호소하는 증상 명을 갖고 있지 않은 데이터도 제외한다. 마지막으로 사망자의 비율이 생존자의 비율보다 매우 작아 학습이 편중되는 것을 막기 위해 생존자의 비율을 무작위로 제거하여 전체 사망자 수와의 비율을 9대1로 맞추었다.

3.3 벡터화

3.2절에서 선택된 자질(n)은 환자(i)의 데이터의 특성에 따라 각각의 다른 방법으로 벡터 공간에 Equation (1)과 같이 맵핑시키며, 데이터의 특성(c)은 수치, 클래스, 단어 3가지로 구성된다.

$$V(i) = \langle w_{ic1}, w_{ic2}, \dots, w_{icn} \rangle \quad (1)$$

먼저, 데이터의 특성이 수치인 경우 관찰된 수치 중 가장 낮은 값과 가장 높은 값을 0과 1 사이로 변환하며, 이때 미등록 데이터의 가중치를 구분하기 위해 Equation (2)의 스무딩 기법을 적용한다.

$$w_{ic} = \frac{x - \min + 0.1}{\max - \min + 0.1} \quad (2)$$

데이터의 특성이 클래스인 경우 one-hot방식의 이중 가중치를 사용하여 Equation (3)과 같이 인코딩한다.

$$w_{ic} = \langle v_{i1}, \dots, v_{ij} \rangle, \quad v_{ij} \begin{cases} 1, & t_j \in \cup \\ 0, & t_j \notin \cup \end{cases} \quad (3)$$

데이터의 특성이 단어인 자질은 증상 명과 배양검사 결과로 나뉘며 Fig. 2는 특성이 단어인 자질을 벡터화 하는 과정이다.

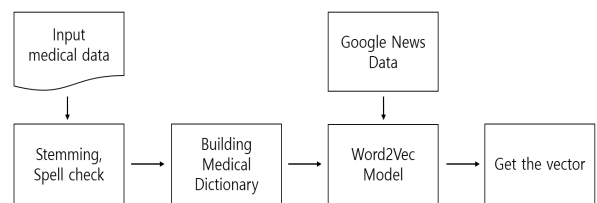


Fig. 2. The Process of Vectorization for Words

두 자질은 전문적인 의료 용어와 약어로 구성되는 경우가 많아 약어의 원형을 복원시켜 사전을 구축한다. 예를 들어서 “abd. pain”인 경우 “abdominal. pain”으로 치환하여 전처리 과정을 거친 후 용어 [“abdominal”, “pain”]에 대하여 사전을 구축한다. 이는 사람이 수동으로 구축한 데이터이기 때문에 대소문자 혹은 공백, 특수문자와 같은 실수로 인한 오류를 제거하는 과정이 필요하다.

Equation (4)는 구축된 사전을 이용하여 Equation (3)과 같이

one-hot 방식을 이용한 벡터와 구글 뉴스 데이터(GoogleNews-vectors-negative300)를 학습한 Word2Vec를 이용하여 Equation (5)와 같이 평균 벡터를 구한 후, 이 두 벡터를 결합하여 확장된 벡터의 수식이다. 구글 뉴스 데이터의 특성상 상대적으로 전문 의학 용어의 빈도수가 적게 출현하는 주증상명은 확장된 벡터를 사용하고 배양검사 결과는 확장 벡터를 사용하지 않는다.

$$w_{ic} = \langle v_{i1}, \dots, v_{ij}, w2v(v_i) \rangle, \quad v_{ij} \begin{cases} 1, & t_j \in \cup \\ 0, & t_j \notin \cup \end{cases} \quad (4)$$

$$w2v(v_i) = \frac{1}{j} \sum_j v_j^{w2v}, \quad v_j^{w2v} \begin{cases} w2v(t_j), & t_j \in w2v \\ w2v(0), & t_j \notin w2v \end{cases} \quad (5)$$

4. 실험 및 결과

우리는 사망률 예측을 기본적으로 회귀분석(regression) 문제로 보고 효과적으로 이항분류를 할 수 있는 지지벡터기계와 Feed Forward Neural Network(FFNN)를 사망률 예측 모형에 적용하여 두 시스템의 성능을 비교하였다. Fig. 3은 회귀분석을 위해 구현한 FFNN의 구조이다.

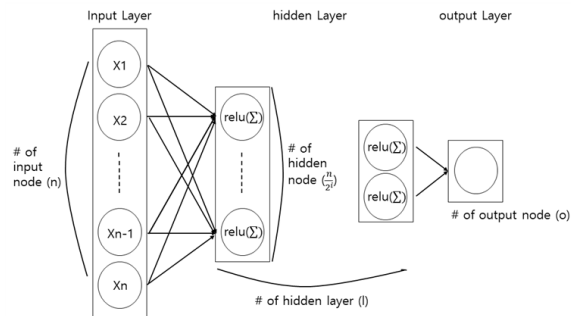


Fig. 3. Structure of Feed Forward Neural Network

입력층의 노드 수는 차원의 수와 같으며 다음 은닉층의 노드 수는 이전 개수의 50%를 사용한다. 0과 1사이의 값이 출력층으로부터 발생하며 0.5 이상인 경우 사망으로 판단한다.

Table 1은 실험에 사용된 지지벡터기계와 FFNN의 기본 파라미터 값을 나타낸다.

실험에 사용한 컴퓨터 사양은 Ubuntu 16.04 LTS, Intel(R) Core(TM) i5-7500 3.40GHz, 16GB DDR4, GeForce GTX 1080이였으며, 지지벡터기계와 FFNN은 각각 scikit-learn과 Tensorflow-gpu 1.4를 이용하여 구현하였다.

의학계에서는 진단율과 오진단율 모두를 낮추는 것이 중요한 평가 척도이며 AUC(Area Under Curve)로 시스템의 성능을 주로 평가한다. 실험은 5-fold 교차검증(Cross Validation)을 수행하여 시스템 성능을 병원내 사망률과 병원내 생존율을 기준으로 각각 다르게 측정하며, 정확률(P), 재현율(R), F1, 정확도(Acc), AUC로 평가한다.

Table 1. Basic Parameter Values in the Experiment

method	parameter	
SVM	Linear Kernel	
FFNN	# of input node	dimension(n)
	# of hidden node	$\frac{n}{2^l}$
	# of output node	1
	weight initialization	Xavier Initialization
	Activation Function	ReLU
	Optimizer	Gradient Descent
	Learning Rate	0.01

실험에는 총 758명의 데이터 중 681명이 생존자의 데이터이며 77명이 사망자의 데이터를 사용하였으며, Table 2는 5-fold 교차검증에 사용된 환자의 수를 나타낸다.

Table 2. The Number of Patients in the Set

set		# of alive	# of deaths	# of total
set 1	train	540	66	606
	test	141	11	152
set 2	train	549	57	606
	test	132	20	152
set 3	train	542	67	606
	test	139	13	152
set 4	train	547	59	606
	test	134	18	152
set 5	train	546	62	608
	test	135	15	150

실험에 사용되는 자질의 수는 89이고, 이를 이용하여 얻은 벡터는 836차원이다. 실험에서 사용되는 FFNN의 은닉층 수는 0, 1, 2 총 3가지의 경우로 나누어 실험한다. 은닉층의 수를 증가할수록 학습의 속도가 저하되고 특히 3개 이상의 은닉층을 사용하는 경우 과적합(over-fitting)이 발생한다. Table 3은 사망률을, Table 4는 생존율을 기준으로 한 5-fold 교차검증의 결과이고, Table 5는 Table 3과 Table 4의 결과를 평균하여 2-class 분류기로서의 성능을 평가한다.

Table 3. The Result Based on Mortality

model	P	R	F1	Acc	AUC	# of hidden layer	best epoch	training time (sec)
SVM	52.7	50.3	51.0	90.4	86.9	-	-	2.97
FFNN	57.5	38.2	45.8	91.0	89.4	0	20000	976.89
	58.1	47.7	52.0	91.3	88.6	1	10000	507.80
	51.1	45.9	48.1	90.0	88.0	2	10000	525.81

Table 4. The Result Based on Non-mortality

model	P	R	F1	Acc	AUC	# of hidden layer	best epoch
SVM	94.3	95.0	94.7	90.4	86.9	-	-
FFNN	93.0	97.2	95.0	90.9	89.6	0	20000
	94.3	96.3	95.3	91.4	88.9	1	10000
	93.9	95.7	94.8	90.6	88.7	2	5000

Table 5. The Result of 2-class Classification

model	# of hidden layer	P	R	F1	Acc	AUC
SVM	-	73.5	72.6	72.8	90.4	86.9
FFNN	0	75.2	67.7	70.4	91.0	89.5
	1	76.2	72.0	73.6	91.4	88.7
	2	72.5	70.8	71.5	90.3	88.3

Table 3과 Table 4에서 사망률 예측이 생존율 예측보다 많이 저조한 성능을 보였는데 이는 학습데이터가 9:1 정도로 생존자의 데이터에 편중된 결과로 분석된다. Table 3과 Table 5에서와 같이 지지벡터기계의 재현율은 FFNN보다 나은 성능을 보였으나, 전체적인 성능은 1개의 은닉층을 사용한 FFNN이 가장 좋은 성능을 보였고, 이 때 AUC가 88.7%이다.

Fig. 4는 Table 3의 사망률 예측 결과 중 지지벡터기계를 사용한 결과를 ROC 곡선으로 나타낸 것이고, Fig. 5는 1개의 은닉층을 사용한 FFNN의 결과이며, Fig. 6은 Kim et al. (2016) [7]의 결과이다.

Fig. 4와 Fig. 5의 ROC 곡선들을 비교하면 모든 5회의 실험에서 지지벡터기계보다 FFNN의 성능이 더 높았다.

ROC 곡선을 비교하면 [7]의 결과 중 가장 좋은 성능을 보인 조합의 AUC가 90.3%이었으며, 이는 본 실험보다 근소하게 나은 것처럼 보인다. 그러나 [7]의 결과는 학습데이터와 평가데이터가 동일한 닫힌 평가집합(closed test set)의 성능이다. 본 연구와 사용 데이터가 달라 직접적인 비교는 어려우나 제안 시스템을 닫힌 평가집합으로 평가하면 AUC가 100%이 나오는 것을 감안하면 본 연구가 좀 더 우수한 성능을 거뒀다고 할 수 있다.

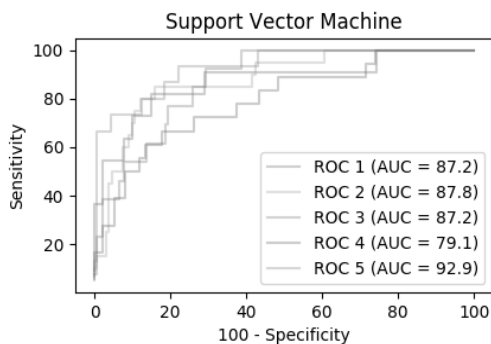


Fig. 4. The ROC Curves of Support Vector Machine

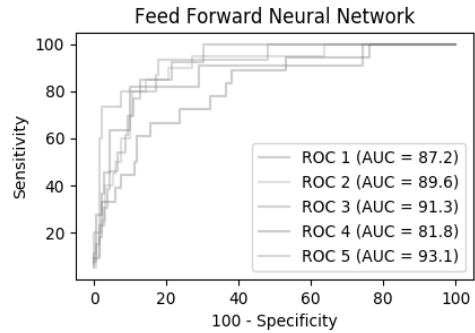


Fig. 5. The ROC Curves of Feed Forward Neural Network

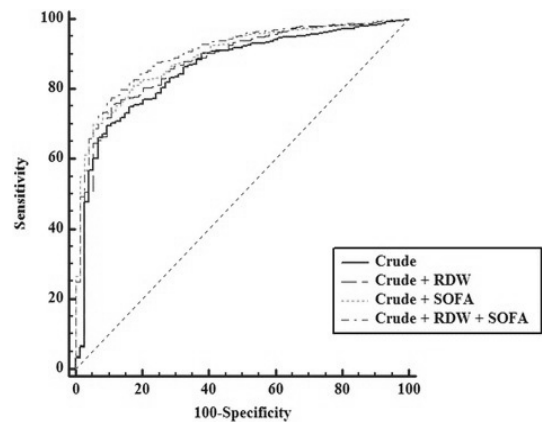


Fig. 6. The ROC Curves of Kim et al.(2016)

제안된 시스템의 88.7%의 AUC는 비록 사용된 데이터가 다르지만 [9]의 랜덤포레스트를 이용한 질병별 사망률 예측 결과인 82.6%보다 좋은 결과를 보였다.

5. 결론

본 연구에서는 응급실을 방문한 노인 환자의 사망률을 지지벡터기계와 인공신경망을 이용하여 각각 예측하였고, 그 결과 인공신경망을 이용한 예측이 지지벡터기계를 이용한 것보다 미세하게 좋은 성능을 보였고 한 개의 은닉층을 가진 FFNN이 가장 좋은 결과를 얻었다. 이 때 F1은 52.0%, AUC는 88.6%였다. 그러나 사망률 예측이 생존율 예측보다는 저조한 성능을 보였는데 이는 생존 환자의 데이터 비율이 사망 환자의 데이터 비율보다 월등히 높아서 발생하는 문제로 보인다. 향후 시계열 데이터, 의료 영상 데이터 등 사망률을 예측하는데 주요한 자질들이 추가된다면 다양한 신경망 모형을 설계하여 성능을 향상시킬 수 있을 것이다.

Deep Learning을 이용하여 응급실을 방문한 노인 환자의 사망률을 기존의 도구보다 정확하게 예측할 수 있다면 의학계뿐만 아니라 인공지능 산업 전반에 큰 영향을 끼칠 것이다. 제안된 시스템을 활용하여 환자의 상태에 따라 신속한 치료를 통해 환자의 생존율을 높일 수 있다면 의료 서비스 산업과 다른 인공지능 산업에 막대한 파급효과를 가져올 것이다.

References

[1] L. C. Mion, R. M. Palmer, G. J. Anetzberger, and S. 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.

[2] D. C. Roberts, M. P. McKay, and A. Shaffer, "Increasing rates of emergency department visits for elderly patients in the United States, 1993 to 2003," *Annals of Emergency Medicine*, Vol.51, No.6, pp.769-774, Jun. 2008.

[3] F. Aminzadeh, and W. B. Dalziel, "Older adults in the emergency department: a systematic review of patterns of use, adverse outcomes, and effectiveness of interventions," *Annals of Emergency Medicine*, Vol.39, No.3, pp.238-247, Mar. 2002.

[4] N. Samaras, T. Chevalley, D. Samaras, and G. Gold, "Older patients in the emergency department: a review," *Annals of Emergency Medicine*, Vol.56, No.3, pp.261-269, Sep. 2010.

[5] 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.

[6] M. Coslovsky, J. Takala, A. K. Exadaktylos, L. Martinolli, and T. M. Merz, "A clinical prediction model to identify patients at high risk of death in the emergency department," *Intensive Care Medicine*, Vol.41, No.6, pp.1029-1036, Jun. 2015.

[7] S. H. Kim, J. H. Yeon, K. N. Park, S. H. Oh, S. P. Choi, Y. M. Kim, H. J. Kim, and C. S. 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*, Vol.24, No.1, p.81, Jun. 2016.

[8] S. P. Shashikumar, M. D. Stanley, I. Sadiq, Q. Li, A. Holder, G. D. Clifford, and S. 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] M. Makar, M. Ghassemi, D. M. Cutler, and Z. 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.

[10] Y. Jo, N. Loghmanpour, and C. P. Rosé, "Time Series Analysis of Nursing Notes for Mortality Prediction via a State Transition Topic Model," in *Proceedings of International Conference on Information and Knowledge Management*, pp.1171-1180, Oct. 2015.

[11] Y. Jo, L. Lee, and S. Palaskar, "Combining LSTM and Latent Topic Modeling for Mortality Prediction," in *Proceedings of International Conference on Information and Knowledge Management*, *arXiv preprint arXiv:1709.02842*, Sep. 2017.

[12] M. A. Morid, O. R. L. Sheng, and S. Abdelrahman, "PPMF: A Patient-based Predictive Modeling Framework for Early ICU Mortality Prediction," in *Proceedings of International Conference on Information and Knowledge Management*, *arXiv preprint arXiv:1704.07499*, Apr. 2017.



박준혁

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

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

2017년 한국교통대학교 컴퓨터정보공학과 (학사)

2018년~현 재 한국교통대학교 컴퓨터정보공학과 석사과정

관심분야 : 자연언어처리, 기계학습, 의미분별, 인공지능



이성욱

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

e-mail : leesw@ut.ac.kr

1996년 서강대학교 전자계산학과(학사)

1998년 서강대학교 컴퓨터학과(석사)

2003년 서강대학교 컴퓨터학과(Ph.D.)

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