

# 딥러닝 기반 80대·90대·100대 남녀 대상 폐암 진단 후 사망률 예측에 관한 연구

변경근\*, 이덕규\*\*, 이세영\*\*

## A Study on the Prediction of Mortality Rate after Lung Cancer Diagnosis for Men and Women in 80s, 90s, and 100s Based on Deep Learning

Kyung-Keun Byun\*, Doeg-Gyu Lee\*\*, Se-Young Lee\*\*

**요약** 최근 의학계에서도 딥러닝 기술을 이용한 질병의 치료결과 예측 연구가 활발하다. 그러나, 소규모 환자 데이터와 특정한 딥러닝 알고리즘을 선택·활용, 연구를 진행하여 특정 조건 아래에서 의미 있는 결과를 보여주었다. 본 연구에서는 연구 결과의 일반화를 위하여 환자 대상을 좀 더 확대·세분화하여 80대·90대·100대 남녀 대상으로 폐암 진단 후 사망률 예측 연구 결과를 도출하였다. 건강보험심사평가원의 대규모 진료 정보와 다종의 딥러닝 알고리즘을 제공하는 AutoML을 이용, 80대·90대·100대 남녀의 폐암 진단 후 84개월간의 사망률 예측을 위해 Decision Tree, Random Forest, Gradient Boosting, XGBoost, Logistic Regression 등 5개 알고리즘별 모델을 생성하고 이를 통해 예측 성능을 비교하고 사망률에 영향을 미치는 요인에 대한 분석도 추진하였다. 연구 결과, 80대와 90대에서 남성이 여성보다 사망 예측률이 더 높았으며 100대에서는 여성의 사망 예측률이 남성보다 높게 나타났다. 그리고 사망률에 가장 큰 영향을 미치는 요인으로는 치료기간으로 분석되었다.

**Abstract** Recently, research on predicting the treatment results of diseases using deep learning technology is also active in the medical community. However, small patient data and specific deep learning algorithms were selected and utilized, and research was conducted to show meaningful results under specific conditions. In this study, in order to generalize the research results, patients were further expanded and subdivided to derive the results of a study predicting mortality after lung cancer diagnosis for men and women in their 80s, 90s, and 100s. Using AutoML, which provides large-scale medical information and various deep learning algorithms from the Health Insurance Review and Assessment Service, five algorithms such as Decision Tree, Random Forest, Gradient Boosting, XGBoost, and Logistic Registration were created to predict mortality rates for 84 months after lung cancer diagnosis. As a result of the study, men in their 80s and 90s had a higher mortality prediction rate than women, and women in their 100s had a higher mortality prediction rate than men. And the factor that has the greatest influence on the mortality rate was analyzed as the treatment period.

**Key Words** : AutoML, Deep Learning Algorithms, Lung Cancer, Men and Women in 80s, 90s, and 100s, Prediction of Mortality Rate

### 1. 서론

최근 의학 분야에서 딥러닝 기법을 이용한 질병의

치료 결과와 진단 예측에 대한 연구가 활발하다. 본 논문에서는 기존 딥러닝 기반 80대·90대 노령자 대상 폐암 진단 후 사망률 예측에 관한 연구[1]를 더욱 확대·

\* Doctoral Degree, IT Policy and Management, Soongsil University(kkbyun@hanmail.net)

\*\* Doctoral Degree, IT Policy and Management, Soongsil University

Received March 01, 2023

Revised March 07, 2023

Accepted March 14, 2023

세분화하여 80대·90대·100대 남녀 대상 폐암 진단 후 84개월간의 사망률을 예측하고 사망률에 영향을 미치는 요인을 분석하고자 한다. 또한, 지난 연구[1]와 같이 다양한 알고리즘에 대한 비교·평가를 추진하였다. 본 연구에서는 건강보험심사평가원의 환자 진료 정보를 기반으로 추진하였다. 사망률이 높은 80대(남 : 21,559명, 여 : 9,206명)·90대(남 : 3,437명, 여 : 2,078명)·100대(남 : 126명, 여 : 114명) 남녀 노령자 총 36,520명을 대상으로 사망원인 1위[1]인 암 중에서도 사망률이 높은 폐암[2] 환자의 사망률을 Decision Tree, Random Forest, Gradient Boosting, XGBoost, Logistic Regression 등 5개 알고리즘 활용하여 예측·비교하고 사망률에 영향을 미치는 요인도 분석·제시하고자 한다. 그리고 효율적이고 신속한 연구 결과의 도출을 위해 AutoML을 활용하였다.

## 2. 관련 연구

### 2.1 국내 사망자 및 암 사망자 현황

통계청의 통계정보에 따르면, [그림 1]과 같이 '21년 총 사망자 수는 317,680명으로 전년 대비 12,732명(4.2%) 증가하였고 80세 이상의 사망자가 전체 사망에서 50.0%를 차지, 10년 전보다 15.2%p 증가하였다[2].

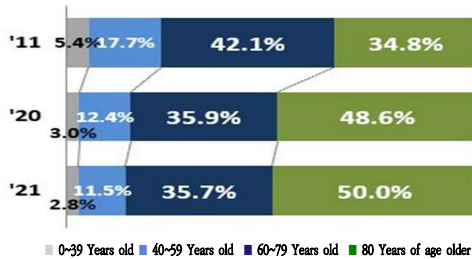


그림 1. 연령별 사망자 수 구성비  
Fig. 1. Composition ratio of death toll by age (Source: Statistics Korea)

한편, [그림 2]와 같이 주요 10대 사망원인은 악성 신생물(암), 심장 질환, 폐렴, 뇌혈관 질환, 고의적 자해(자살), 당뇨병, 알츠하이머병, 간 질환, 패혈증, 고혈압성 질환 순이다[2]. 사망원인 1위를 차지한 암중

[그림 3]과 같이 2021년 암종별 사망자 수 통계를 살펴보면, 폐암 사망자 수는 18,902명으로 암 중에서 사망률 1위를 차지하였다[3].

Unit : per 100,000 people in the study

Ranking	Cause of death	Mortality rate
1	Malignant neoplasm(cancer)	161.1
2	A heart disease	61.5
3	Pneumonia	44.4
4	Cerebrovascular disease	44.0
5	Suicide	26.0
6	Diabetes	17.5
7	Alzheimer's disease	15.6
8	A liver disease	13.9
9	Septicemia	12.5
10	A highly blood-synthes	12.1

그림 2. 사망원인 순위 추이  
Fig. 2. Changes in the ranking of causes of death (Source: Statistics Korea)

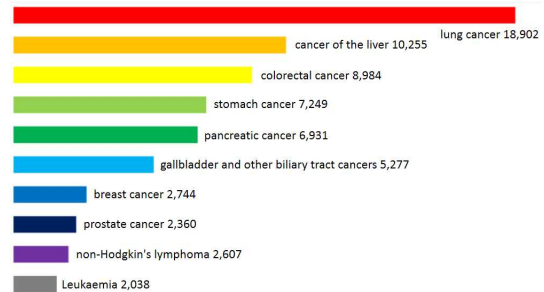


그림 3. 2021년 암종별 사망자수  
Fig. 3. Death toll by cancer species in 2021 (Source: National Cancer Information Center)

### 2.2 기존 연구

영국의 랭커스터 대학은 의료종사자들에게 보다 빠르고 효율적으로 환자를 진단하고 치료할 수 있도록 딥러닝 알고리즘(CNN)을 활용하여 폐렴, 폐암, 결핵(TB), 폐 혼탁 및 가장 최근의 COVID-19의 감지와 분류에 우수한 성능을 입증하였고[4]. 인도의 코넬루락스마야 교육 재단에서는 CNN(Convolutional Neural Network)과 INN(Inception Neural Network) 알고리즘을 이용한 딥러닝 기법으로 흉부 X-선을 이용하여 폐질환을 예측, 결핵, 폐렴, 폐암 및 천식 등의

정상 비정상 여부를 분류하였다[5]. 또한 중국의 쓰촨 대학에서는 컴퓨터 단층 촬영(CT) 이미지를 딥러닝 기술을 사용하여 결절을 식별하고 악성 종양을 평가에 도움이 되는 시스템을 개발하였다[6]. 국내에서는 서울대 의과대학에서 폐암 수술 환자 809명의 폐암 생존 예측 및 그 주요 요인을 평가하였다[7]. 한편, 이화여대에서 순환종양세포를 이용한 폐암 재발을 예측 [8], 말레이시아 푸트라재에서는 ‘암 게놈 지도(Cancer Genome Atlas)’ 슬라이드 이미지를 대상으로 CNN 기법(Inception v3)을 활용, 선암종, 편평세포암종 또는 정상 폐 조직을 정확하게 분류한 연구도 진행되었다[9]. 미국 메릴랜드대 등 4개 기관에서 공동 추진된 구강편평세포암 환자의 생존 및 국소 재발에 대한 예측 모델 개발에 머신러닝 및 딥러닝 모델을 제안하였다[10]. 또한, 마카오대에서 암 진단 및 예후 판단에 머신러닝이 어떻게 지원하는지 탐구한 사례도 있었다 [11]. 그리고 가천대에서 초음파 검사를 통해 갑상선 종양의 재발 예측이 가능한 딥러닝 학습 모델도 개발 되었으며[12], 하버드 의과대 등 3개 기관에서 비소세포폐암 CT 이미지를 기반으로 사망 위험 분류에 딥러닝 기술을 적용하였다[13]. 본 연구에서는 암 진단 및 예후 예측 관련 연구를 진행하되 대규모 환자 정보와 다양한 알고리즘을 기반으로 80대·90대·100대 남녀 대상 폐암 사망률과 그에 영향을 미치는 요인을 비교하고 분석하고자 한다.

### 3. 연구모형 및 방법

#### 3.1 연구모형

[그림 4]와 같이 국내 폐암 환자 현황 데이터와 딥러닝 기술을 제공하는 AutoML을 활용, 훈련용 데이터(데이터 전량의 80%)로 딥러닝 모델을 생성하고 검증용 데이터(데이터 전량의 20%)로 폐암 진단 후 84개월 동안의 사망률을 예측하고 사망률에 영향을 미치는 요인을 분석하는 연구모형을 설계하였다. 건강보험심사평가원의 2015년에 폐암 진단 후 2021년까지 84개월간의 국내 폐암 환자 현황 데이터중 80대·90대·100대 남녀 환자(36,520) 데이터를 선별하여 [표 1]과 같이 △환자식별번호 △치료결과 △성별 △나이 △주

병상 △수술여부 △치료기간 등 사망률 요인 분석에 필요한 변수를 정의하고 데이터를 정리하였다[1]. 주 병상 코드의 종류와 관련 설명은 [표 2]와 같다.

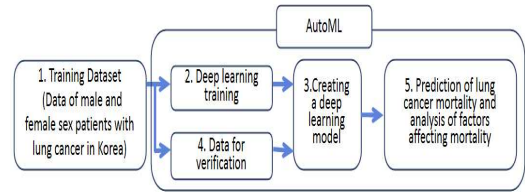


그림 4. 연구모형  
Fig. 4. Research model

표 1. 환자정보 데이터 구조

Table 1. Patient information data structure

Variable name	Sortation
Patient ID	13-digit number
Treatment Results	death, survival
Gender	male, female
Age	80S, 90S, 100S
The chief soldier code	C34, C340, C341, C342, C343, C348, C349
Operation status	surgery, non-surgery
Treatment Period	12 months, 24 months, 36 months, 48 months, 60 months, 72 months, 84 months

표 2. 주병상 코드 설명

Table 2. Description of the main lesion code

Code Name	Explanation
C34	malignant neoplasm of bronchial and lung
C340	malignant neoplasm of the main bronchus
C341	malignant neoplasm of the upper lobe, bronchial tubes, or lungs
C342	Malignant neoplasm of mesenchymal, bronchial, or lung
C343	malignant neoplasm of the lower lobe, bronchus, or lung
C348	Malignant neoplasm of overlapping lesions of the bronchial tubes and lungs
C349	Malignant neoplasm of unknown bronchial tubes or lungs

한편, 사망률 예측 및 폐암 진단 후 사망률에 영향을 미치는 요인 분석을 신속하고 효율적으로 수행하기 위해 딥러닝 기반의 AutoML(WiseProphet)을 활용하였다[1]. [표 3]과 같이 WiseProphet(http://prophet.wise.co.kr/)에서 제공하는  $\Delta$ Decision Tree  $\Delta$ Random Forest  $\Delta$ Gradient Boosting  $\Delta$ XGBoost  $\Delta$ Logistic Regression 등의 알고리즘별로 입력된 파라미터 값을 이용, 사망률을 예측·비교하였다[1].

표 3. 5개 알고리즘 및 입력 파라미터 값  
Table 3. 5 Algorithm and Input Parameter Values

Algorithm Type	Parameters(value)
Decision Tree	$\Delta$ criterion(gini) $\Delta$ max_depth(5) $\Delta$ min_samples(leaf 1)
Random Forest	$\Delta$ criterion(gini) $\Delta$ max_depth(5) $\Delta$ min_samples_leaf(1) $\Delta$ n_estimators(10) $\Delta$ random_state(None)
Gradient Boosting	$\Delta$ learning_rate(0.1) $\Delta$ max_depth(3) $\Delta$ subsample(1.0)
XGBoost	$\Delta$ eta(0.3) $\Delta$ gamma(0) $\Delta$ max_depth(3)
Logistic Regression	$\Delta$ C(1.0) $\Delta$ random_state(None)

그리고 사망률 예측 성능평가 지표는 [표 4]와 같이 정밀도와 재현율을 사용하였다[1]. 사망률에 영향을 미치는 요인 분석은 WiseProphet에서 제공되는 변수간 상관관계 분석 기능을 이용하였다[1].

표 4. 성능평가 지표  
Table 4. Performance evaluation index

Measurement formula			
$\Delta \text{ Precision} = \frac{TP}{(TP + FP)}$ * Percentage of predicted values where actual values occur			
$\Delta \text{ Recall} = \frac{TP}{(TP + FN)}$ * Proportion of probing accurately detected values classified in the mode			
Sortation		Prediction	
		Positive	Negative
Actuality	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

### 4. 성능 분석 결과

2015년에 폐암 진단 후 2021년까지 84개월간의 국내 폐암환자 현황 데이터(102,951 명)를 정리한 연령이 80대 이상(36,520) 환자 대상  $\Delta$ 치료결과  $\Delta$ 성별  $\Delta$ 나이  $\Delta$ 주병상  $\Delta$ 수술여부  $\Delta$ 치료기간 등의 일반 통계를 먼저 살펴보면 [그림 5]와 같이 사망자는 70% (25,485명)를 차지하였으며 생존자는 30%(11,035명)이다. 또한, [그림 6]과 같이 성비는 남성이 69%(25,122명) 여성이 31%(11,398명)를 차지하였다. [그림 7]과 같이 나이 비율은 80대가 84%(30,765명), 90대가 15%(5,515명), 100대가 1%(240명)를 차지하였다. 주 병상 비율은 [그림 8]과 같이 C349 65%(23,859명), C341 16%(5,893명), C343 12%(4,234명), C340 4%(1,313명), C342 3%(931명), C348 0.7%(256명), C34 0.09%(34명)이다. 한편, [그림 9]와 같이 수술을 받은 비율은 94%(34,359명), 수술을 받지 않은 비율은 6%(2,161명)이다.

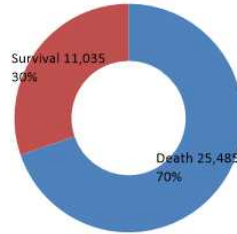


그림 5. 치료 결과 비율

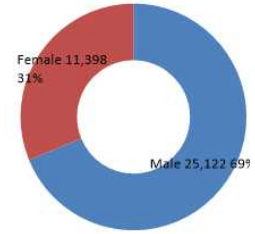


그림 6. 성비

Fig. 5. Treatment outcome ratio Fig. 6. Gender ratio

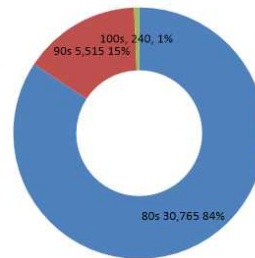


그림 7. 나이 비율

Fig. 7. The ratio of age

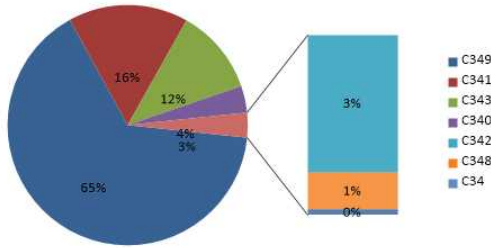


그림 8. 주병상 비율  
Fig. 8. Proportion of main lesion

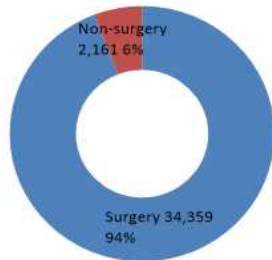


그림 9. 수술비수술 비율  
Fig. 9. Surgery/Non-surgery ratio



그림 10. 치료기간 비율  
Fig. 10. Treatment period ratio

치료기간을 살펴보면 [그림 10]과 같이 12개월내가 33%(12,122명), 24개월내가 18%(6,755명), 36개월내가 11%(3,870명), 48개월내가 8%(3,046명), 60개월내가 8%(2,920명), 72개월내가 10%(3,830명), 84개월내가 11%(3,977명)를 차지하였다. 이와 같은 환자 데이터를 기반으로 딥러닝 기법의 AutoML을 활용하여 나이대와 성별에 따라 사망률을 예측한 결과는 아래와 같다.

#### 4.1 80대 남자(21,559명) 사망률 예측 결과

[표 6]·[그림 11]과 같이 5개 알고리즘에서 예측한 정밀도의 평균 값은 0.79이며 평균 재현율은 0.89이었다. 정밀도는 D.T·G.B. 알고리즘이 우수했고 재현율은 R.F. 알고리즘이 우수했다. 한편 [표 7]과 같이 사망률에 영향을 미치는 요인 분석 결과, 치료기간이 가장 큰 영향을 미쳤으며 수술을 받아야만 하는 경우보다는 수술 이외의 치료를 받는 경우가 사망률에 영향력이 큰 것으로 나타났다. 그리고 주병상 중에는 C340(주기관지의 악성 신생물)이 가장 영향력이 큰 것으로 나타났다.

표 6. 80대 남성 사망률 예측 결과

Table 6. The prediction results of mortality rate for men in 80s

Algorithm Type		Precision	Recall
Decision Tree	D.T.	0.80	0.88
Random Forest	R.F.	0.77	0.93
Gradient Boosting	G.B.	0.80	0.88
XGBoost	X.B.	0.79	0.89
Logistic Regression	L.R.	0.79	0.88
Average		0.79	0.89

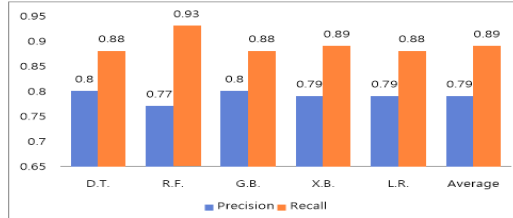


그림 11. 80대 남성 사망률 예측 결과

Fig. 11. The prediction results of mortality rate for men in 80s

표 7. 80대 남성 사망률에 영향을 미치는 요인 분석 결과

Table 7. Analysis of factors affecting mortality rate in men in 80s

Ranking	Variable	Importance
1	Treatment Period	0.93784
2	Operation status_surgery	0.02767
3	Operation status_non-surgery	0.02175
4	The chief soldier code_C340	0.00785
5	The chief soldier code_C342	0.00135
6	The chief soldier code_C341	0.0013
7	The chief soldier code_C348	0.00111
8	The chief soldier code_C343	0.00059
9	The chief soldier code_C349	0.00051
10	The chief soldier code_C34	0.00003

### 4.2. 80대 여성(9,206명) 사망률 예측 결과

[표 8]·[그림 12]와 같이 5개 알고리즘에서 예측한 정밀도의 평균값은 0.65이며 평균 재현율은 0.79였다. 정밀도는 L.R. 알고리즘이 우수했고 재현율은 R.F.·G.B.·X.B. 알고리즘이 우수했다. 한편, [표 9]와 같이 사망률에 영향을 미치는 요인 분석 결과, 치료기간이 가장 큰 영향을 미쳤으며 80대 여성 환자에 있어서는 수술 이외의 치료를 받는 경우가 수술을 받아야만 하는 경우 보다 사망률에 영향력이 큰 것으로 나타났다. 그리고 주병상 중에는 C348(기관지 및 폐의 중복병변)의 악성 신생물이 가장 영향력이 큰 것으로 나타났다.

표 8. 80대 여성 사망률 예측 결과  
Table 8. The prediction results of mortality rate for women in 80s

Algorithm Type		Precision	Recall
Decision Tree	D.T.	0.65	0.79
Random Forest	R.F.	0.65	0.80
Gradient Boosting	G.B.	0.65	0.80
XGBoost	X.B.	0.65	0.80
Logistic Regression	L.R.	0.66	0.78
Average		0.65	0.79

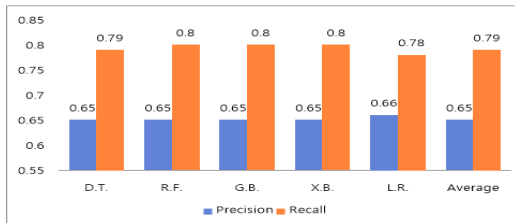


그림 12. 80대 여성 사망률 예측 결과  
Fig. 12. The prediction results of mortality rate for women in 80s

표 9. 80대 여성 사망률에 영향을 미치는 요인 분석 결과  
Table 9. Analysis of factors affecting mortality rate in women in 80s

Ranking	Variable	Importance
1	Treatment Period	0.8905
2	Surgery	0.03258
3	Non-surgery	0.03099
4	The chief soldier code_C348	0.01283
5	The chief soldier code_C343	0.00973
6	The chief soldier code_C349	0.00924
7	The chief soldier code_C340	0.00743
8	The chief soldier code_C34	0.00378
9	The chief soldier code_C341	0.00165
10	The chief soldier code_C342	0.00127

### 4.3 90대 남성(3,437명) 사망 예측률

[표 10]·[그림 13]과 같이 5개 알고리즘에서 예측한 정밀도의 평균값은 0.91이며 평균 재현율은 1.00이었다. 정밀도는 L.R.을 제외한 4개 알고리즘 모두 우수했고 재현율은 D.T.·R.F.·L.R. 알고리즘이 우수했다. 한편 [표 11]과 같이 사망률에 영향을 미치는 요인 분석 결과, 치료기간이 가장 큰 영향을 미쳤으며 주병상 C341(상엽, 기관지 또는 폐의 악성 신생물)이 두 번째 그리고 수술을 받아야만 하는 경우가 세 번째로 영향력이 큰 것으로 나타났다.

표 10. 90대 남성 사망률 예측 결과  
Table 10. The prediction results of mortality rate for men in 90s

Algorithm Type		Precision	Recall
Decision Tree	D.T.	0.91	1.00
Random Forest	R.F.	0.91	1.00
Gradient Boosting	G.B.	0.91	0.99
XGBoost	X.B.	0.91	0.99
Logistic Regression	L.R.	0.90	1.00
Average		0.91	1.00

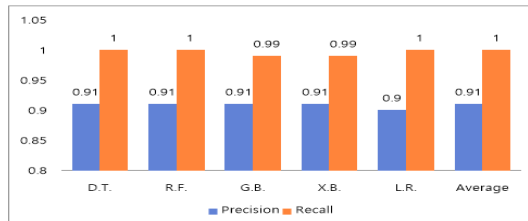


그림 13. 90대 남성 사망률 예측 결과  
Fig. 13. The prediction results of mortality rate for men in 90s

표 11. 90대 남성 사망률에 영향을 미치는 요인 분석 결과

Table 11. Analysis of factors affecting mortality rate in men in 90s

Ranking	Variable	Importance
1	Treatment Period	0.80247
2	The chief soldier code_C341	0.10018
3	Operation status_surgery	0.02904
4	The chief soldier code_C343	0.02818
5	Operation status_non-surgery	0.0178
6	The chief soldier code_C349	0.00999
7	The chief soldier code_C340	0.0077
8	The chief soldier code_C342	0.00453
9	The chief soldier code_C348	0.0001

#### 4.4. 90대 여성(2,078명) 사망 예측률

[표 12]·[그림 14]와 같이 5개 알고리즘에서 예측한 정밀도의 평균값은 0.89이며 평균 재현율은 0.93이었다. 정밀도는 D.T.·R.F. 알고리즘 우수했고 재현율은 G.B.·X.B. 알고리즘이 우수했다. 한편 [표 13]과 같이 사망률에 영향을 미치는 요인 분석 결과, 치료기간이 가장 큰 영향을 미쳤으며 주병상 C343(하엽, 기관지 또는 폐의 악성 신생물)이 두 번째 그리고 주병상 C340(주기관지의 악성 신생물)이 세 번째로 영향력이 큰 것으로 나타났다.

표 12. 90대 여성 사망률 예측 결과

Table 12. The prediction results of mortality rate for women in 90s

Algorithm Type		Precision	Recall
Decision Tree	D.T.	0.89	0.91
Random Forest	R.F.	0.89	0.91
Gradient Boosting	G.B.	0.86	0.95
XGBoost	X.B.	0.86	0.95
Logistic Regression	L.R.	0.86	0.94
Average		0.87	0.93

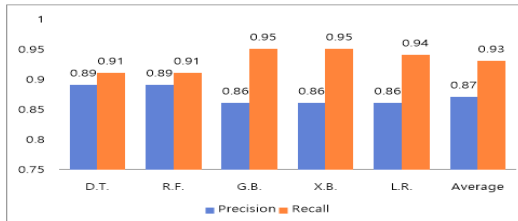


그림 14. 90대 여성 사망률 예측 결과

Fig. 14. The prediction results of mortality rate for women in 90s

표 13. 90대 여성 사망률에 영향을 미치는 요인 분석

Table 13. Analysis of factors affecting mortality rate in women in 90s

Ranking	Variable	Importance
1	Treatment Period	0.9006
2	The chief soldier code_C343	0.03218
3	The chief soldier code_C340	0.02638
4	Operation status_non-surgery	0.01338
5	Operation status_surgery	0.01251
6	The chief soldier code_C349	0.00792
7	The chief soldier code_C341	0.00589
8	The chief soldier code_C348	0.00105
9	The chief soldier code_C342	0.00008

#### 4.5 100대 남성(126 명) 사망 예측률

[표 14]·[그림 15]와 같이 5개 알고리즘에서 예측한 정밀도의 평균값은 0.96이며 평균 재현율은 1.00이었다. 정밀도는 5개 알고리즘 모두 우수했고 재현율도 모든 알고리즘이 우수했다. 한편 [표 15]와 같이 사망률에 영향을 미치는 요인 분석 결과, 치료기간이 가장 큰 영향을 미쳤으며 수술 이외의 치료를 받은 경우가 두 번째로 영향력이 있었으며 주병상 C343(하엽, 기관지 또는 폐의 악성 신생물)이 세 번째로 영향력이 큰 것으로 나타났다.

표 14. 100대 남성 사망률 예측 결과

Table 14. The prediction results of mortality rate for men in 100s

Algorithm Type		Precision	Recall
Decision Tree	D.T.	0.96	1.00
Random Forest	R.F.	0.96	1.00
Gradient Boosting	G.B.	0.96	1.00
XGBoost	X.B.	0.96	1.00
Logistic Regression	L.R.	0.96	1.00
Average		0.96	1.00

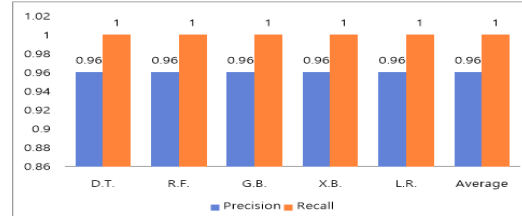


그림 15. 100대 남성 사망률 예측 결과

Fig. 15. The prediction results of mortality rate for men in 100s

표 15. 100대 남성 사망률에 영향을 미치는 요인 분석

Table 15. Analysis of factors affecting mortality rate in men in 100s

Ranking	Variable	Importance
1	Treatment Period	0.35879
2	Operation status_non-surgery	0.24903
3	The chief soldier code_C343	0.11422
4	The chief soldier code_C349	0.11273
5	Operation status_surgery	0.09626
6	The chief soldier code_C341	0.06897

#### 4.6. 100대 여성(114 명) 사망 예측률

[표 16]·[그림 16]과 같이 5개 알고리즘에서 예측한 정밀도의 평균값은 1.00이며 평균 재현율은 1.00이었다. 정밀도는 5개 알고리즘 모두 우수했고 재현율도 모든 알고리즘이 우수했다. 한편, [표 17]과 같이 사망률에 영향을 미치는 요인 분석 결과, 치료기간이 가장 큰 영향을 미쳤으며 주병상 C349(상세불명의 기관지 또는 폐의 악성 신생물)이 두 번째로 영향력이 있었으며 주병상 C341(상업, 기관지 또는 폐의 악성 신생물)이 세 번째로 영향력이 큰 것으로 나타났다.

표 16. 100대 여성 사망률 예측 결과

Table 16. The prediction results of mortality rate for women in 100s

Algorithm Type		Precision	Recall
Decision Tree	D.T.	1.00	1.00
Random Forest	R.F.	1.00	1.00
Gradient Boosting	G.B.	1.00	1.00
XGBoost	X.B.	1.00	1.00
Logistic Regression	L.R.	1.00	1.00
Average		1.00	1.00

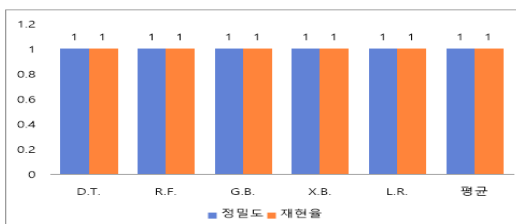


그림 16. 100대 여성 사망률 예측 결과

Fig. 16. The prediction results of mortality rate for women in 100s

표 17. 100대 여성 사망률에 영향을 미치는 요인 분석

Table 17. Analysis of factors affecting mortality rate in women in 100s

Ranking	Variable	Importance
1	Treatment Period	0.6178
2	The chief soldier code_C349	0.22024
3	The chief soldier code_C341	0.10429
4	Operation status_non-surgery	0.03276
5	Operation status_surgery	0.02491

## 5. 결 론

본 연구를 통해 딥러닝 기법을 활용하여 2015년 암 유병 진단 후 2021년까지 84개월 동안의 80대·90대·100대 남녀 폐암 환자의 데이터를 분석하여 5개 알고리즘별로 사망률을 예측·비교하고 사망률에 영향을 주는 요인을 확인해 보았다.

표 18. 80대·90대·100대 남녀 사망률 평균 예측 결과

Table 18. Results of predicting the average mortality rate of men and women in 80s, 90s, and 100s

구분		Average precision	Average recall
80s	Male	0.79	0.89
	Female	0.65	0.79
90s	Male	0.91	1.00
	Female	0.87	0.93
100s	Male	0.96	1.00
	Female	1.00	1.00

[표 18]과 같이 나이대별로 남녀 성별 평균 사망률을 예측한 정밀도 값과 재현율을 비교해보면 나이가 들수록 정밀도와 재현율이 높아졌다. 이는 고령일수록 관련 데이터의 양이 부족하여 나타난 현상으로 판단된다. 그리고 80대와 90대에서 남성이 여성보다 사망 예측률이 더 높았으나 100대에서는 여성의 사망 예측률이 남성보다 높게 나타났다. 한편, 학문적으로 다량의 환자 데이터를 기반으로 다양한 딥러닝 알고리즘을 제공하는 AutoML을 활용한 연구방법을 제시함으로써 타 분야에서도 활용할 수 있도록 효율적인 연구방법을 실증하였다. 한편, 이번 연구가 보건복지 당국의 정책 수립에 활용되기를 기대한다.



## REFERENCES

- [1] K. K. Byun, D. G. Lee, and Y. T. Shin, "A Study on the Prediction of Mortality Rate after Lung Cancer Diagnosis for the Elderly in their 80s and 90s Based on Deep Learning," Annual Spring Conference of KIPS, 29.1, pp. 452-455, 2022.
- [2] 2021 Cause of Death Statistics, National Statistical Office, 2022.
- [3] Major Cancer Mortality, National Cancer Information Center, 2022.
- [4] Alshmrani, Goram Mufarah M., et al. "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images." Alexandria Engineering Journal pp. 923-935, Nol. 64, 2023.
- [5] Bhattacharya, Debdatta, et al. "Classification of Healthy and Diseased Lungs by Pneumonia Using X-Rays and Gene Sequencing With Deep Learning Approaches." Smart Technologies in Data Science and Communication. Springer, Singapore, pp. 189-196, 2023.
- [6] Cheng, Nitao, et al. "Inferring cell-type-specific genes of lung cancer based on deep learning." Current Gene Therapy pp. 439-448, Vol. 22, No.5, 2022.
- [7] Jin-ah Sim., Young Ae Kim., Ju Han Kim., Jong Mog Lee., Moon Soo Kim., Young Mog Shim., Jae Ill Zo., Young Ho Yun., The major effects of health-related quality of life on 5-year survival prediction among lung cancer survivors: applications of machine learning, Scientific Reports, Vol 10, No. 10693, 2020.
- [8] Sehwa Moon, Janghwan Choi, Prediction of recurrence of lung cancer using deep learning-based circulatory tumor cells, The 2020 Spring Conference of the Korean Mechanical Society, pp 54-54, Aug. 2020.
- [9] Nicolas Coudray., Paolo Santiago Ocampo., Theodore Sakellaropoulos., Navneet Narula., Matija Snuderl., David Fenyö., Andre L Moreira., Narges Razavian., Aristotelis Tsirigos., Hyperparameter Tuning and Pipeline Optimization via Grid Search Method and Tree-Based AutoML in Breast Cancer Prediction, Nat Med, pp 1559-1567, Vol. 24, No. 10, Oct. 2018.
- [10] Ahmed S Sultan., Mohamed A Elgharib., Tiffany Tavares., Maryam Jessri., John R Basile., The use of artificial intelligence, machine learning and deep learning in oncologic histopathology, Journal Oral Pathol Med. 2020 Oct;49(9):849-856.
- [11] Shigao Huang., Jie Yang., Simon Fong., Qi Zhao., Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges, Cancer Lett, pp. 61-71, Vol. 28, No. 471, Feb. 2020.
- [12] Jieun Kil., Kwang Gi Kim., Young Jae Kim., Hye Ryoung Koo., Jeong Seon Park., Deep Learning in Thyroid Ultrasonography to Predict Tumor Recurrence in Thyroid Cancers, Journal Korean Soc Radiol, pp. 1164-1174, Vol. 81, No. 5, 2020.
- [13] Ahmed Hosny., Chintan Parmar., Thibaud P Coroller., Patrick Grossmann., Roman Zeleznik., Avnish Kumar., Johan Bussink., Robert J Gillies., Raymond H Mak., Hugo J W L Aerts., Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study, PLoS Med. e1002711, Vol. 30, No.15, Nov. 2018.

---

## 저자약력

---

### 변 경 근 (Kyung-Keun Byun)



- 1996년 홍익대 컴퓨터공학 학사
- 1998년 홍익대 전산학 석사
- 2021년 ~ 현재 송실대학교 IT정책경영학과 박사과정

〈관심분야〉 정보보호, 자동차 S/W 보안, ML

**이 덕 규 (Doeg-Gyu Lee)**



- 1987년 계명대 전자계산학과(학사)
- 2016년 카톨릭대 의료경영학과(석사)
- 2021년 ~ 현재 숭실대  
IT정책경영학과(박사과정)
- 1987년 ~ 현재 건강보험심사평가원  
실장

〈관심분야〉 보건의료 관련 IT정책 및 정보화분야

**이 세 영 (Se-Young Lee)**



- 1996년 서울대 컴퓨터공학과(공학  
사)
- 2011년 Univ. of Birmingham  
MBA(경영학 석사)
- 2013년 ~ 현재 숭실대학교 IT정  
책경영학과 박사과정

〈관심분야〉 전자정부, 정보보안, 마이데이터