

## 인공지능을 이용한 주진단 S코드의 낙상환자 예측모델 개발

박예지\*, 최은미\*\*, 방소현\*, 정진형\*\*\*

### Development of a Prediction Model for Fall Patients in the Main Diagnostic S Code Using Artificial Intelligence

Ye-Ji Park\*, Eun-Mee Choi\*\*, So-Hyeon Bang\*, Jin-Hyoung Jeong\*\*\*

**요약** 낙상사고는 세계적으로 매년 42만 건 이상 발생하는 치명적인 사고이다. 따라서, 낙상 환자를 연구하고자 낙상 환자의 손상외인코드와 주진단 S코드의 연관성을 찾고, 낙상 환자의 주진단 S코드 데이터를 가지고 손상외인코드를 예측할 수 있는 예측모델을 개발하였다. 본 연구에서는 강원특별자치도 강릉시에 있는 A 기관의 2020~2021년 2년간의 데이터를 받아 낙상에 관련된 손상외인코드 W00~W19까지 데이터만 추출하고, 낙상 손상외인코드 중 예측모형을 개발할 정도의 주진단 S코드를 가지고 있는 W01, W10, W13, W18 데이터를 가지고 예측모형 개발하였다. 데이터 중 80%는 훈련용 데이터, 20%는 테스트용 데이터로 분류하였다. 모형 개발은 MLP(Multi-Layer Perceptron)을 이용하여 6개의 변수(성별, 나이, 주진단S코드, 수술유무, 입원유무, 음주유무)를 입력층에 64개의 노드를 가진 2개의 은닉층, 출력층은 softmax 활성화 함수를 이용하여 손상외인코드 W01, W10, W13, W18 총 4개의 노드를 가진 출력층으로 구성하여 개발하였다. 학습결과 첫 번째 학습했을 때 31.2%의 정확도를 가졌지만, 30번째는 87.5%의 정확도를 나타냈고 이를 통해 낙상환자의 낙상외인코드와 주진단 S코드의 연관성을 확인할 수 있었다.

**Abstract** Falls are fatal accidents that occur more than 420,000 times a year worldwide. Therefore, to study patients with falls, we found the association between extrinsic injury codes and principal diagnosis S-codes of patients with falls, and developed a prediction model to predict extrinsic injury codes based on the data of principal diagnosis S-codes of patients with falls. In this study, we received two years of data from 2020 and 2021 from Institution A, located in Gangneung City, Gangwon Special Self-Governing Province, and extracted only the data from W00 to W19 of the extrinsic injury codes related to falls, and developed a prediction model using W01, W10, W13, and W18 of the extrinsic injury codes of falls, which had enough principal diagnosis S-codes to develop a prediction model. 80% of the data were categorized as training data and 20% as testing data. The model was developed using MLP (Multi-Layer Perceptron) with 6 variables (gender, age, principal diagnosis S-code, surgery, hospitalization, and alcohol consumption) in the input layer, 2 hidden layers with 64 nodes, and an output layer with 4 nodes for W01, W10, W13, and W18 exogenous damage codes using the softmax activation function. As a result of the training, the first training had an accuracy of 31.2%, but the 30th training had an accuracy of 87.5%, which confirmed the association between the fall extrinsic code and the main diagnosis S code of the fall patient.

**Key Words** : Falling, Damaged extraneous code, AI, MLP, Main Diagnostic S Code

This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE) in 2023 (2022RIS-005)

\*Dept. of Biomedical IT, Catholic Kwandong University

\*\*Dept. of Medical Management, Catholic Kwandong University

\*\*\*Corresponding Author : Department of Biomedical IT, Catholic Kwandong University (wsgud0201@cku.ac.kr)

Received December 12, 2023

Revised December 18, 2023

Accepted December 21, 2023

## 1. 서론

### 1.1 연구의 필요성

세계보건기구(World Health Organization)는 전 세계적으로 매년 42만 건 이상 발생하는 치명적인 낙상사고를 주요한 공중보건 문제로 선포했다[1]. 또한, 전 세계적으로 낙상은 사고 혹은 비고의적 부상으로 인한 사망원인 중 두 번째로 높다[2].

낙상은 의료기관에서 가장 빈번하게 발생하는 위해 사건 중 하나이고[3], 세계적으로 간호의 질을 평가하는 대표적인 지표이다[4]. 종합병원은 낙상의 잠재적인 위험이 복합적으로 도사리는 장소로[5] 낙상 발생률은 외국의 경우 재원일 수 1,000일당 3~5건으로 보고되고 있다[6].

우리나라는 2000년에 이미 전체 인구 중 65세 이상의 노인 인구가 7.2%로 고령화 사회로 진입하게 되었고, 2018년에는 14.3%로 고령 사회, 2025년에는 노인 인구가 20.3%를 차지함에 따라 초고령 사회에 도달할 것으로 예상하여지고 있다[7]. 노인 인구가 증가함에 따라 노인 건강에 관한 관심이 높아지고, 지역 사회나 요양시설, 병원을 중심으로 노인 낙상에 관한 연구가 활발히 진행되고 있다[8].

의료기관 내에서 낙상은 우발적인 낙상과 예측 가능한 생리적 낙상 그리고 예측 불가능한 생리적 낙상으로 분류된다[9]. 특히 노인의 질환과 활동 제한은 낙상과 관계가 깊고, 낙상을 하게 되면 손상을 쉽게 받으며, 젊은 연령에 비해 심하고 이로 인해 대부분 사망하게 된다[10].

질병관리청 국가손상정보포털에 올라온 2022 손상 유형 및 원인 통계 응급실손상환자심층조사 통계에 따르면 2022년 낙상 손상 환자 중 19.1%는 입원을 했고, 1.0%가 사망을 했다. 70세 이상군에서 입원을 (36.0%)과 사망률(2.2%)로 모두 전 연령 중에 가장 높다. 또한, 낙상으로 인한 손상 환자는 집(45.1%)에서 가장 많이 발생하고, 도로(25.0%)와 상업시설(9.5%) 순서로 발생했다. 그리고 낙상으로 다치는 부위는 외상성 뇌 손상이 52.0%로 가장 많았고 상지(22.2%), 기타 머리(21.1%), 하지(15.9%)를 다치는 경우가 많다고 통계로 나타났다[11].

국내에서는 노인의 낙상에 대한 연구는 많이 진행되었으나, 낙상에 영향을 미치는 요인이나, 노인의 낙상 경험에 따른 우울 및 삶의 질에 관한 연구가 주로 많이 차지하고 있다[12].

본 논문에서는 질병분류코드를 이용하여 예측모델을 개발하고자 한다. 질병분류코드란 질병을 유형별로 분류해서 알파벳과 숫자를 활용해 체계적으로 기호를 붙인 것으로 병원, 의원에서 의사가 발행하는 처방전에 기재가 되는데 환자가 어떠한 질환으로 분류되어 처방을 받는지 분류하는 기호이다.

### 1.2 연구 목적

본 연구는 선행 연구로 강원도 강릉시 소재의 A 기관 의료정보 관리 부서에서 2020년 1월 1일부터 2021년 12월 31일까지의 2년간 환자 데이터를 받아 낙상 코드별 발생 빈도수를 분석을 진행했다. 발생 빈도수를 확인 후 환자의 성별, 연령, 주진단 S코드, 수술유무, 입원유무, 음주유무 6가지를 낙상 외인코드와 손상 S코드의 상관관계 및 정확도를 예측했다. 이 정확도 예측을 통해서 지역 사회의 보건학적 차원의 손상 발생률과 사망률을 감소시킬 수 있는 예방적 방안 및 손상 그리고 외상환자 치료에 가이드라인을 제공하는 것을 목적으로 한다.

## 2. 연구방법

### 2.1 연구방법

#### 2.1.1 자료수집

본 연구를 위해 강원특별자치도 강릉시에 위치한 A 기관에서 2020년, 2021년의 2년간 '퇴원손상심층조사' 항목을 기준으로 기관의 Electric Medical Record(EMR) 시스템에 입력된 데이터 자료를 수집하였다. 퇴원손상심층조사는 2004년에 구축하여 2005년부터 질병관리본부에서 실시해 온 국가사업으로 전체 입원환자를 대상으로 조사한 행정자료이다. 퇴원손상심층조사에는 성별, 연령, 우편번호, 보험 유형 등 환자의 정보와 입원일, 퇴원일, 주진단 코드, 부진단 코드, 손상외인코드 등 환자의 진료 내용을 포함하고 있다.

2.1.2 자료 분석 방법

강원특별자치도 강릉시에 위치한 A 기관에서 수집한 2020년, 2021년의 2년간 데이터 중 전체 퇴원환자의 손상외인코드를 확인한 후 낙상 데이터 W00~W19를 1차 추출, 낙상코드 W00~W19까지의 데이터 중 미끄러짐, 걸림 및 헛디딤에 의한 동일 면상에서의 낙상인 W01, 계단에서의 낙상인 W10, 빌딩 또는 구조물에서의 낙상인 W13, 동일 면상에서의 기타 낙상인 W18에 해당하는 낙상환자 자료 187명의 데이터를 추출하였다.

2.2 변수정의

본 연구에서는 낙상환자의 성별, 연령, 재원일 수, 수술유무, 손상외인코드, 입원유무, 음주 여부 등을 예측모형 개발을 위한 변수로 사용하였다.

표 1. 데이터 변수 정의  
Table 1. Defining Data Variables

Variables	Definitions
Sex	either male or female in two categories in which organisms are divided
Age	length of time that a person has lived or a thing has existed
LOS	Length of Stay in hospital
Existence of surgery	Presence or absence of Surgery
Damaged extraneous code	Code that categorizes external causes of damage or illness suffered by a patient
Intensive care room hospitalization status	A ward that accommodates patients who are likely to recover and provides intensive care to patients within a 24-hour system
Drinking or not	a person who has drunk alcohol

2.3 예측모델개발 방법

2.3.1 예측모델개발 대상자

분석대상자의 낙상으로 발생한 W01, W10, W13, W18 환자를 대상으로 성별, 연령, 주진단 S코드, 수술유무, 입원유무, 음주유무 총 6가지 변수를 이용하여 모델을 제작하였다.

2.3.2 예측모델개발

주진단 S코드의 낙상환자 예측모델을 개발하기 위해 1차로 K-겹 교차검증 (K-fold cross validation) 방법, 2차로 MLP(Multi-Layer Perceptron) 방법으로 예측모델 개발을 진행하였다. K-겹 교차검증은 <그림 2>와 같이 5-겹 교차검증일 경우, 5개로 데이터를 나누고 1개씩 돌아가면서 테스트 데이터로 사용되는 것으로 <그림 2>을 바탕으로 예를 들면 Cross validation iteration 1의 경우 A, B, C, D 4개가 Train data가 되고 E가 Test data가 되지만, Cross validation iteration 2의 경우 A, B, C, E가 Train data로 이용되고 D가 Test data가 된다. 이후 Cross validation iteration 3에서는 C, Cross validation iteration 4에서는 B, Cross validation iteration 5에서는 A가 Test data로 이용되며 모든 데이터가 훈련 데이터와 테스트 데이터로 사용된다[13]. <그림 2>와 같은 테스트는 5-겹 교차검증이라 하며 <그림 3>과 같은 식을 나타낸다. 5-겹이 아닌 경우엔 <그림 1>처럼 k번 반복하여 나온 모든 결과값을 평균 내어 검증 결과값으로 사용한다. 따라서 모든 데이터를 Train 과 Test 데이터로 사용할 수 있고 데이터양이 적을 때 테스트 데이터를 어떻게 잡는지에 따라 성능이 다르며, 우연의 효과로 모델평가 지표에 편향이 생기게 되는데 k-fold validation은 모든 데이터가 최소 1번 이상 테스트 데이터로 사용되기 때문에 이러한 문제를 해결할 수 있다[14].

$$k\text{-fold cross validation} = \frac{1}{k} \sum_{i=1}^k \text{iteration}_i$$

그림 1. k-겹 교차검증

Fig. 1. k-fold cross validation

	A	B	C	D	E
Cross validation iteration 1	Train	Train	Train	Train	Test
Cross validation iteration 2	Train	Train	Train	Test	Train
Cross validation iteration 3	Train	Train	Test	Train	Train
Cross validation iteration 4	Train	Test	Train	Train	Train
Cross validation iteration 5	Test	Train	Train	Train	Train

그림 2. 5-겹 교차검증 예시

Fig. 2. 5-fold cross validation example

$$5\text{-fold cross validation} = \frac{1}{5} \sum_{i=1}^5 \text{iteration}_i$$

그림 3. 5-겹 교차검증

Fig. 3. 5-fold cross validation

또한, 2차 예측모델 개발에 사용된 MLP (Multi-Layer Perceptron)는 <그림 4>와 같이 Input layer, Hidden layer, Output layer로 이루어져 있으며, 은닉층을 한층 이상 가지고 있는 경우 다층 퍼셉트론이다. 본 논문에서는 뉴런의 활성화 함수로 은닉층은 ReLU (Rectified Linear Unit)를, 그리고 출력층은 4가지 W01, W10, W13, W18을 출력하기 위해 softmax 활성화 함수를 사용하였다. <그림 5>는 본 논문의 MLP(Multi-Layer Perceptron) 모델을 도식화한 것으로 입력층의 1~6은 각 1은 성별, 2 나이, 3 주진단 S코드, 4 수술유무, 5 입원유무, 6 음주유무 등을 나타내며, 은닉층은 64개의 노드를 가지고 두 개의 층으로 이루어져 있다. 마지막 출력층은 W01, W10, W13, W18로 4개의 노드로 이루어져 있다.

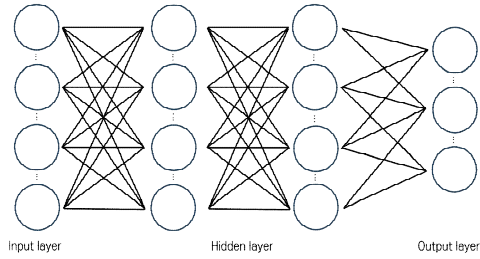


그림 4. 다층 퍼셉트론 예시

Fig. 4. Multi-Layer Perceptron example

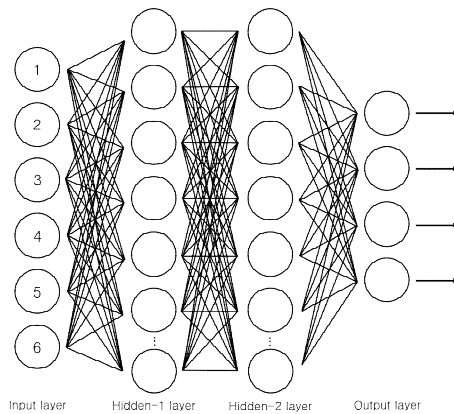


그림 5. MLP 모델 도식화

Fig. 5. MLP Model Schematic

### 3.3.3 예측모델평가

예측모델평가는 ‘model.evaluate’ 함수를 사용하여 모델의 loss와 accuracy를 측정했다. 입력 데이터에 대해 모델이 예측한 출력을 실제 정답과 비교하여 예측이 실제 값과 얼마나 차이가 나는지 손실(loss) 지표를 내며, loss 값이 적을수록 실제 레이블과 가깝다는 것을 나타내고 정확도(accuracy)는 모델이 올바른 예측한 비율을 나타내고 있다.

## 3. 연구결과

### 3.1 분석대상자의 특성

본 연구의 분석대상자는 159명이었다. 분석대상자의 특성은 성별은 남자 48.4%, 여자 51.6%로 나타났다.

으며, 연령대별로는 39세 이하 14.5%, 40~49세 11.3%, 50~59세 15.6%, 60~69세 16.4%, 70~79세 25.2%, 80세 이상 17%로 나타났다. 손상외인코드별로는 W01 25.2%, W10 13.2%, W13 11.3%, W18 50.3%였다. 수술유무는 수술하지 않은 사람이 71.7%, 수술한 사람이 28.3%였고, 입원유무는 입원하지 않은 사람 84.3%, 입원한 사람 15.7%로 나타났다. 마지막으로 술을 마신 사람 4.4%, 술을 마시지 않은 사람 95.6%의 특성을 보였다.

표 2. 분석대상자 특성  
Table 2. Characteristics of Analysis Subjects

Variables		N	%
Sex	Male	77	48.4
	Female	82	51.6
Age	~39	23	14.5
	40~49	18	11.3
	50~59	25	15.6
	60~69	26	16.4
	70~79	40	25.2
Damaged extraneous code	W01	40	25.2
	W10	21	13.2
	W13	18	11.3
	W18	80	50.3
Surgery status	No	114	71.7
	Yes	45	28.3
hospitalization status	No	134	84.3
	Yes	25	15.7
Drinking or not	No	152	95.6
	Yes	7	4.4

### 3.2 MLP 모델평가 결과

MLP 연구결과 입력층에 성별, 나이, 주진단 S코드, 수술유무, 입원유무, 음주유무 총 6개의 변수를 입력하고, 은닉층은 64개의 노드를 가지고 두 개의 층으로 설정 후, 마지막 출력층은 W01, W10, W13, W18로 4개의 노드로 구성하였다. 그 후, 30번의 학습을 시켰

을 때 첫 번째에서 손실 값이 1.3377, 정확도가 31.3%로 측정되었으나 학습이 늘어날수록 손실 값은 줄어들고 정확도는 올라가는 것을 확인할 수 있었다. 마지막 30번째에서는 손실 값이 0.3881, 정확도가 87.5%로 첫 번째와 비교했을 때 손실은 0.9496, 정확도는 56.3%의 차이를 확인할 수 있었다.

표 3. MLP 모델평가 결과  
Table 3. MLP Model Evaluation Results

Depth	Loss	Accuracy
1	1.3377	31.2%
2	1.0593	65.6%
3	0.9082	65.6%
4	0.7930	65.6%
5	0.7448	65.6%
6	0.6956	68.7%
7	0.6621	78.1%
8	0.6540	71.8%
9	0.6247	78.1%
10	0.5959	81.3%
11	0.5720	78.1%
12	0.5633	78.1%
13	0.5494	78.1%
14	0.5368	78.1%
15	0.5107	81.3%
16	0.5146	78.1%
17	0.4824	81.2%
18	0.4688	78.1%
19	0.4680	78.1%
20	0.4459	78.1%
21	0.4478	78.1%
22	0.4346	81.3%
23	0.4257	81.3%
24	0.4148	81.3%
25	0.4113	81.3%
26	0.4094	75.0%
27	0.3956	84.4%
28	0.3934	81.3%
29	0.3933	84.4%
30	0.3881	87.5%

#### 4. 결론

낙상환자에 관한 연구는 문헌 조사를 통해 낙상에 영향을 미치는 요인이나, 노인의 낙상 경험에 따른 우울 및 삶의 질에 관한 연구가 주로 많이 차지하고 있다는 점을 확인하였다. 따라서, 본 논문에서는 주진단 S코드와 손상외인코드를 가지고 예측모델을 만드는 연구를 진행하였다.

본 연구에서는 낙상 환자의 성별, 나이, 주진단 S코드, 수술유무, 입원유무, 음주유무 데이터로 손상외인코드를 예측할 수 있는 예측모델을 MLP를 이용하여 개발하였다. 이를 통해 낙상환자의 낙상외인코드와 주진단 S코드의 연관성을 확인할 수 있었다.

추후 W01, W10, W13, W18 4개의 손상외인코드가 아닌 추가 데이터를 얻어 W00~W19까지의 낙상에 관련된 모든 손상외인코드를 가지고 예측모델을 개발을 보다 정확한 예측을 연구할 계획이다.

#### REFERENCES

- [1] Seonghee Park, & Eunkyung Kim. (2016). Useful fall risk assessment tools for adult hospitalized patients: A systematic review and meta-analysis. *Korean J Health Promot*, 16(3), 180-191.
- [2] Donghyun Lim, Yubin Ji, & Kyungmin Kim. (2018). Risk stratification extraction and policy proposal through analysis of characteristics of fall fatalities. 2. *Journal of the Korean Disaster Prevention Society*, 18(1), 171-178.
- [3] Eun-Kyung Kim, Jae-Chang Lee, & Mi-Ran Eom. (2008). Predictors of fall risk in hospitalized patients. *Journal of Korean Academy of Nursing*, 38(5), 676-684.
- [4] Stephenson, T. (2005). The national patient safety agency. *Archives of disease in childhood*, 90(3), 226-228.
- [5] Oliver, D., Healey, F., & Haines, T. P. (2010). Preventing falls and fall-related injuries in hospitals. *Clinics in geriatric medicine*, 26(4), 645-692.
- [6] Cho Young-shin, Lee Young-ok, and Yoon Young-sun. (2019). Risk factors for falls among inpatients at a tertiary general hospital: A survival analysis. *Journal of Critical Care Nursing*, 12(1), 57-70.
- [7] Hee-seong Lee and Sun-ho Kwon. (2020). Problems and improvement plans for the elderly welfare system in a super-aging society. *Labor Law Journal*, 50, 1-29.
- [8] Park Ae Ja, Lim Nan Young, Kim Yoon Shin, Lee Yoon Kyoung, Song Jung Hee.(2011).Incidence and Predictors of Falls in Institutionalized Elderly.*Journal of Muscle and Joint Health*,18(1),50- 62.
- [9] Ja-hye Cheon, Hyun-ah Kim, Mi-jeong Kwak, Hyo-seon Kim, Seon-kyung Park, Moon-sook Kim, Ae-ri Choi, Ji-in Hwang and Yun-sook Kim. (2018). Clinical practice guidelines for fall risk factor assessment and fall prevention activities. *Korean Journal of Medical Quality Improvement*, 24(2), 41-61.
- [10] Younghye Park. (2004). Assessment of risk factors for falls among community-dwelling seniors. *Journal of Geriatric Nursing*, 6(2), 170-178.
- [11] Korea Disease Control and Prevention Agency, 2022 Injury Type and Cause Statistics, In-Depth Survey on Emergency Room Injury Patients, Publication Registration Number 11-1352159-000071-10, ISSN 2287-3481
- [12] Su-Jin Lee, Kyung-Won Baek, Nam-Soo Park, Min-Kyung Kim, & Sang-Nam Jeon. (2023). Relationship between fall occurrence location and death in elderly patients hospitalized due to falls. *Korean J Health Educ Promot*, 40(1).
- [13]Yong Eun Lee, Nak Joon Choi, Young Hoo Byun, Dae Won Kim, & Kyung Chun Kim (2021). Rubber O-ring defect detection system using K-fold cross validation and support vector machine. *Journal of The Korean Society of Visualization*, 19(1), 68-73.
- [14]Choi Byung-kwan, Ham Seung-woo, Kim Chok-hwan, Seo Jung-sook, Park Myung-hwa and Kang Sung-hong. (2018). Development of a model for predicting the number of hospital stays for acute stroke patients using artificial intelligence. *Digital Convergence Research*, 16 (1), 231-242.

---

저자약력

---

박 예 지 (Ye-Ji Park)

[정회원]



- 2023년 2월 : 가톨릭관동대학교 의료IT학과 졸업 (학사)
- 2023년 2월 ~ 현재 : 가톨릭관동대학교 일반대학원 석사과정 재학

<관심분야> 데이터분석, 헬스케어, 인공지능

최 은 미 (Eun-Mee Choi)

[정회원]



- 1993년3월 ~ 1995년8월 : 한양대학교 행정학석사(병원행정전공)
- 1999년9월~2003년2월 : 한양대학교 대학원 의학박사(산업의학전공)
- 2008년9월~현재 : 가톨릭관동대학교 의료경영학과 교수
- 2021년2월~현재 : 가톨릭관동대학교 의료융합대학장

<관심분야> 보건의료정보, 개인건강정보보호, 손상외인 질병 분류 사례

방 소 현 (So-Hyeon Bang)

[정회원]



- 2021년 2월 : 가톨릭관동대학교 의료IT학과 졸업 (학사)
- 2023년 2월 : 가톨릭관동대학교 일반대학원 졸업 (공학석사)
- 2023년 8월 ~ 현재 : 가톨릭관동대학교 일반대학원 박사과정 재학

<관심분야> 소프트웨어, 데이터분석, 영상처리

정 진 형 (Jin-Hyoung Jeong)

[정회원]



- 2012년 02월: 가톨릭관동대학교 의료공학과 졸업 (학사)
- 2014년 02월: 가톨릭관동대학교 일반대학원 졸업 (공학석사)
- 2017년 08월: 가톨릭관동대학교 일반대학원 졸업 (공학박사)
- 2021년 03월: 가톨릭관동대학교 의료IT학과 조교수

<관심분야> 의료 시스템, 데이터 분석, 통신, 인공지능