

공공빅데이터를 활용한 기계학습 기반 뇌졸중 위험도 예측

Machine Learning-based Stroke Risk Prediction using Public Big Data

정선우¹ · 이민지² · 유선용^{1*}

¹전남대학교 ICT융합시스템공학과

²전남대학교 IoT인공지능융합전공

Sunwoo Jeong¹ · Minji Lee² · Sunyong Yoo^{1*}

¹Dept. of ICT Convergence System Engineering, Chonnam Nation University, Gwangju 61186, Korea

²IoT Artificial Intelligence, Chonnam National University, Gwangju 61186, Korea

[요 약]

본 논문은 빅데이터를 이용하여 심방세동 환자의 뇌졸중 발병을 예측하는 기계 학습 모델을 제시한다. 학습 데이터로는 국민 건강 보험공단에서 제공하는 대한민국 전수에 해당하는 심방세동 환자의 정보를 수집하였다. 수집된 정보는 인구사회학, 과거 병력, 건강검진을 포함한 68개 독립변수로 구성된다. 본 연구의 목표는 기존 심방세동 환자의 뇌졸중 위험도 예측에 사용되던 통계적 모델(CHADS₂, CHA₂DS₂-VASc)의 성능을 검증하고 기계 학습 모델을 적용하여 기존 모델보다 높은 정확도를 가지는 모델을 제시하는 것이다. 제안하는 모델의 정확도, AUROC (area under the receiver operating characteristic)를 검증한 결과 제안하는 기계 학습 기반의 모형이 심방세동 환자의 뇌졸중 위험도를 사용한 모델이 기존의 통계적 모델보다 높은 정확도, 민감도, 특이도를 가지는 것을 확인할 수 있었다.

[Abstract]

This paper presents a machine learning model that predicts stroke risks in atrial fibrillation patients using public big data. As the training data, 68 independent variables including demographic, medical history, health examination were collected from the Korean National Health Insurance Service. To predict stroke incidence in patients with atrial fibrillation, we applied deep neural network. We firstly verify the performance of conventional statistical models (CHADS₂, CHA₂DS₂-VASc). Then we compared proposed model with the statistical models for various hyperparameters. Accuracy and area under the receiver operating characteristic (AUROC) were mainly used as indicators for performance evaluation. As a result, the model using batch normalization showed the highest performance, which recorded better performance than the statistical model.

Key word : Atrial fibrillation, Stroke, CHA₂DS₂-VASc, Machine learning, Deep neural network.

<https://doi.org/10.12673/jant.2021.25.1.96>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 19 January; Revised 30 January 2021

Accepted (Publication) 15 February (28 February 2021)

*Corresponding Author ; Sunyong Yoo

Tel: +82-62-350-1761

E-mail: syyoo@jnu.ac.kr

I. 서론

세계적으로 뇌졸중은 사망원인 2위, 장애 원인 3위를 차지하고 있다 [1]. 전 세계 약 1500만 명이 매년 뇌졸중으로 고통받고 있으며, 그중 5백만 명이 사망하고 5백만 명에는 영구 장애가 남는다 [2]. 한국에서는 암, 심혈관질환에 이어 빈도수 3위의 사망 원인이 되며 2016년 기준 뇌혈관질환으로 10만 명당 45.8명이 사망하였다 [3]. 뇌졸중의 발병은 치매, 우울증, 불안과 같은 심리적 장애와 편측 마비, 안면 마비, 의사소통 장애, 식이 장애, 심각한 두통, 어지러움으로 인한 보행 및 평행 장애, 시각장애와 같은 심각한 신체적 장애를 일으킨다.

뇌졸중은 뇌혈관을 순환하는 혈액이 혈관의 좁은 영역에서 응고되면 뇌에 혈액 및 산소 공급이 차단되어 뇌세포가 손상되며 발생한다. 혈관성 원인에 의해 24시간 이상 상기 증상이 지속되기도 하며, 갑작스러운 발생으로 사망을 초래하기도 하는 국소적, 혹은 전반적인 뇌 기능의 장애를 보이는 임상징후이다. 또한, 원인에 따라 혈전의 조성이 다르며 약물의 선택이 달라질 수 있는 다양한 기전이 작용하는 복합질환으로 정의된다. 일반적으로 고지혈증, 고혈압, 당뇨, 흡연 등의 위험인자를 가진 사람에게서 발견되는 동맥경화가 주원인이 되며 그 외에도 다양한 유전적 원인으로 발생하는 허혈성 출혈성 뇌경색이나 심장 부정맥 및 심부전증에 의한 심인성 뇌경색 등 다양한 원인이 존재한다. 특히 심방세동은 뇌졸중을 5배 이상 증가시키며 이는 대부분 크기가 크고 심한 후유증을 동반하므로 철저한 예방이 필요하다 [4].

심방세동이 뇌졸중 발병에 주는 큰 영향으로 인해 심방세동 환자의 뇌졸중 발병 예측에 관한 많은 연구가 진행되어왔다. 기존 연구에서는 대부분 심방세동 환자의 뇌졸중 위험도를 평가하기 위해 위험 인자들을 고려하여 서로 다른 가중치를 부과한 CHA₂DS₂-VASc score를 이용했으나 본 연구는 더 높은 정확도와 넓은 범용화를 위해 위험 인자들을 종합적인 독립변수로 두어 딥러닝을 통해 뇌졸중 발병 예측을 진행할 것이다. 해당 독립변수에는 인구사회학적 요소와 진료역, 과거력 정보를 포함하며, 기존 통계적 방법보다 더 포괄적으로 위험 인자를 고려할 수 있게 하였다. 따라서 본 연구의 목적은 기존 통계적 방법보다 더욱 개선된 성능을 가지며, 심방세동 환자의 뇌졸중 발병에 관한 인구 사회학적, 과거력 등의 요소들을 종합적으로 고려하여 학습하는 딥러닝을 통한 뇌졸중 예측 기법 개발을 목적으로 한다.

II. 이론적 배경

2-1 CHADS₂ score

CHADS₂ Score는 심방세동이 있는 환자의 경우 뇌졸중 발생 위험이 훨씬 크다는 점에서 착안하여 세워진 고위험군 분류 기준이다. 총 5가지 (심부전 유무, 고혈압 유무, 나이, 당뇨

표 1. CHADS₂ score의 위험요소와 점수표

Table 1. Risk factors and their CHADS₂ scores.

| Letter | Risk factor | Score |
|----------------|--------------------------|-------|
| C | Congestive heart failure | 1 |
| H | Hypertension | 1 |
| A | Age(≥75) | 1 |
| D | Diabetes mellitus | 1 |
| S ₂ | Prior, TIA or stroke | 2 |

표 2. CHA₂DS₂-VASc score의 위험요소와 점수표

Table 2. Risk factors and their CHA₂DS₂-VASc scores.

| Letter | Risk factor | Score |
|----------------|--|-------|
| C | Congestive heart failure, LV dysfunction | 1 |
| H | Hypertension | 1 |
| A ₂ | Age(≥75) | 2 |
| D | Diabetes mellitus | 1 |
| S ₂ | Thrombo-embolism, TIA or stroke | 2 |
| V | Vascular disease | 1 |
| A | Age(65-74) | 1 |
| S | Sex category: female | 1 |

유무, 뇌졸중, 일과성 뇌 허혈증, 색전증의 유무)의 항목이 있으며, 각 항목에 점수를 부여해 합산하여 고위험군을 분류할 수 있다. 뇌졸중, 일과성 뇌 허혈증, 색전증의 유무는 한 항목으로 묶어 해당 사항이 있을 시 2점을 부여하고 나머지 4항목은 모두 1점을 부여한다. 표 1은 CHADS₂ Score의 세부 항목을 나타내고 있다.

2-2 CHA₂DS₂-VASc score

과거 연구에서는 심방세동 환자의 뇌졸중 발병 예측을 위해 주로 CHADS₂ score를 사용했다. 그러나 의학적 발전이 거듭됨에 따라 CHADS₂ score만으로 뇌졸중 고위험군을 정의하기엔 부족하다는 사실이 밝혀지며 이를 보완해 새로운 분류 기준이 세워졌다. CHA₂DS₂-VASc score는 기존 CHADS₂ score에 포함되지 않은 세 가지의 추가적인 항목과 기존의 항목을 세분화한 항목들이 추가되었다 [5]. 표 2는 CHA₂DS₂-VASc score의 세부 항목을 나타내고 있다.

2-3 인공신경망 모델링

딥러닝이란 인공 신경망을 기본으로 하는 학습 알고리즘 중 하나이다. 기계학습의 소분류로 포함될 수 있으며, 여러 개의 인공 신경망을 이용한 깊은 레이어 계층을 이용해 학습한다. 기

계학습의 경우 학습에 필요한 특징 (feature)를 선별해 줄 필요가 있지만, 딥러닝의 경우 이러한 특징 역시 알고리즘이 자체적으로 선별해 학습한다는 차이점이 있다. 지도 학습과 비지도 학습으로 나눌 수 있으며, 지도 학습은 알고리즘에 제공하는 데이터에 정답이 존재하는 경우이고 비지도 학습은 정답이 존재하지 않는 경우이다. 지도 학습은 주어진 데이터를 통해 연속된 값을 예측하는 회귀 문제와 주어진 데이터를 특정 카테고리로 분류하는 분류 문제에 주로 사용된다. 비지도 학습은 정답이 따로 존재하지 않기 때문에 데이터의 패턴이나 특징 등을 학습해 분류하는 군집 문제에 주로 사용된다. 본 연구에 사용된 데이터는 정답이 0과 1로 존재하는 데이터로써, 지도학습의 분류에 해당한다.

III. 연구 내용

3-1 활용 데이터

본 연구에서 사용된 자료원은 국민건강보험공단 데이터베이스의 심방세동 환자의 청구 및 건강검진 자료이다. 자료원의 수집 기간은 2011년~2018년까지이고, 관찰대상 수는 754,949명이다. 본 연구에서는 성별, 나이와 같은 인구 사회학적 요소와 혈관 질환, 비만, 간 질환, 폐혈전 색전증, 고혈압, 당뇨, 울혈성 심부전 등을 포함한 50가지 항목의 진료 및 과거력 내역을 포함한 총 68개의 독립변수를 사용하였고, 종속변수로는 뇌졸중 발생 여부를 나타내는 1개의 변수를 사용하였다.

3-2 제안하는 기계학습 기반 뇌졸중 위험도 예측 모델

심층 신경망 (DNN; deep neural network)은 입력층과 출력층 사이에 다수의 은닉층이 존재하는 인공 신경망이다 [6],[7]. 인공 신경망은 생물의 뇌 신경망을 모티브로 제작한 학습 알고리즘이다 [7].

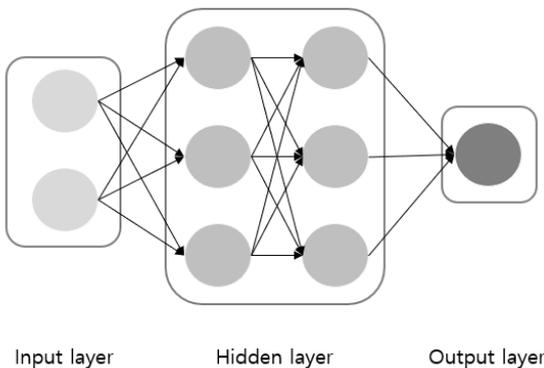


그림 1. 심층 신경망 개요

Fig. 1. Overview of deep neural network model.

그림 1은 히든 레이어가 2개인 인공 신경망을 나타낸 그림이다. 각 레이어에 존재하는 노드들은 가중치가 부과되는 가상의 선으로 연결되어 있다. 인간의 시냅스와 마찬가지로 이전 노드에서 값을 받아 다음 노드로 전달하게 되는데, 이때 입력받은 값의 총합을 하나의 출력값으로 변환하는 활성화 함수를 사용한다. 주로 사용되는 활성화 함수는 주로 sigmoid, ReLU, tanh 등이며 [7], 본 연구에서는 히든 레이어의 활성화 함수로는 ReLU, 출력 레이어의 활성화 함수로는 sigmoid를 사용하였다. Sigmoid 함수는 실수를 취해 0에서 1 사이의 값을 반환한다. 반환된 값이 어느 쪽에 더 가까운지 계산해 0과 1로 분류한다. 그러나 sigmoid 함수는 0과 1 사이의 값을 가져 gradient descent를 이용하면 계속해서 값이 감소하기 때문에 층이 깊어질수록 0으로 수렴하고 gradient vanishing이 발생해 정확도가 감소한다는 한계가 존재한다. 그러므로 은닉층에서는 이를 보완할 수 있는 ReLU를 사용한다 [6][8]. ReLU는 입력이 0보다 작으면 0을 반환하고 0보다 크면 그대로 입력된 값을 반환한다.

손실 함수는 모델이 예측한 결과와 실제 결과 간의 차이를 나타내는 함수로써 모델의 성능을 나타내는 하나의 지표 역할을 한다. 본 연구에서 사용한 모델의 손실 함수는 binary cross entropy 함수를 이용하였다.

가중치를 최적화하는 방법으로는 Adam optimizer을 사용한다. Adam은 확률적 경사 하강법 기법 중 하나로 Momentum 기법과 AdaGrad의 장점을 융합한 기법이다 [9]. 일반적으로 가장 많이 사용되는 알고리즘이며 학습 시 방향 설정과 스텝 사이즈 설정을 적절히 조절해준다.

3-3 성능 평가

모델 성능은 accuracy, precision, recall, F1 score, AUROC (area under receiver operating characteristic), AUPR (precision/recall curve)을 포함한 총 6가지 방법으로 평가한다. 정확도는 모델이 예측한 결과가 얼마나 실제 결과에 부합하는지를 측정하는 척도이다. 모델이 올바르게 예측한 데이터의 수를 전체 데이터 수로 나눈 값으로 가장 기본적인 성능 평가의 방법이다. Precision은 모델이 1이라고 예측한 데이터 중 실제 1인 것의 비율이며 recall은 실제 1인 데이터 중 모델이 1이라고 예측한 데이터의 비율이다. 이들의 조화평균을 구하면 F1 score를 구할 수 있다 [10].

정확도를 통해 성능을 평가하면 종속 변수의 비율이 한쪽으로 편중된 데이터의 경우 정확한 평가를 할 수 없다. 모델이 비율이 높은 쪽의 메이저 클래스로만 예측하더라도 정확도 자체는 높게 나오기 때문이다. 이런 문제점을 보완한 평가 방법이 바로 AUROC이다. ROC 커브 아래의 면적을 나타낸 값으로 AUROC가 높을수록 모델의 성능이 높다고 판단한다. ROC 커브는 모든 분류 임계값에서 모델의 성능을 나타낸 그래프로, x축을 FPR (false positive rate), y축을 TPR (true positive rate)로 둔다 [11],[12]. 한 x축을 recall, y축을 precision으로 두면 PR 커브를 그릴 수 있고 이 역시 커브 아래의 면적이 넓을수록 높은 성능을 나타낸다 [13].

3-4 연구 결과

모델의 성능을 높일 수 있도록 모델에 적용될 기법을 선정하기 위해 다른 조건들은 동일하게 유지한 채 기법을 다르게 적용한 3가지 모델을 만들어 비교하였다.

모델 1은 dropout 기법만, 모델 2는 batch normalization과 dropout 기법을 혼용, 모델 3은 batch normalization 기법만을 추가하여 각각의 모델을 구성하였다. 그 결과, 가장 성능이 좋은 batch normalization 기법만을 추가한 모델3을 선정하였다.

딥러닝 모델의 성능을 높이기 위해, 노드 수 및 파라미터 조절하여 성능을 비교하였다. 레이어 노드 수에 변화를 주어 실험해 본 결과, 최적의 결과로 첫 번째와 두 번째 은닉층은 64개, 출력 레이어의 노드 수는 1개로 선정하였다.

모델의 batch size를 선정하기 위해, 다른 조건들을 동일하게 유지한 채 batch size만 변화를 주어 실험해 본 결과는 표 3-5와 같다. 특히, 불균형한 데이터로 학습시킨 모델의 성능을 평가할 때, F1-score와 AUROC의 결과를 중심으로 평가하게 되는데 그 결과, batch size가 64일 때 가장 높은 F1-score와 AUROC 값을 가지는 것으로 나타났다.

본 논문에서 제안하고자 하는 딥러닝 방식의 성능을 입증하기 위해, 기존의 심방세동 환자의 뇌졸중 여부를 예측하기 위해 사용되던 통계적 방식과 비교 실험을 하였다. 통계적 모델의 경우에는 표 6과 같이 정확도와 AUROC 두 가지 방법으로 평가하였으며, 정확도는 CHA₂DS₂-VASc score의 총점이 n 이상일 경우 발병으로 간주하여 계산하였다. n 은 9부터 1까지 존재하며, 모든 경우의 정확도를 계산한 결과는 표 6과 같다. 또한, AUROC는 자체적으로 임계치를 계산하여 구하기 때문에 임의로 조절해 줄 필요는 없어 모든 데이터를 그대로 사용하여 0.6448이라는 결과를 산출하였다. 즉, 통계적 모델의 가장 높은 정확도인 0.92, AUROC는 0.6448이라는 기록과 딥러닝 모델의 정확도는 0.92, AUROC는 0.7이라는 기록을 비교한 결과 딥러닝 모델이 더 우수한 것으로 나타났다.

IV. 결론

표 3. Batch size 32의 성능 평가표

Table 3. Evaluation of batch size 32.

| | all | 1:1 | 1:5 |
|-----------|---------|---------|---------|
| Accuracy | 0.92 | 0.5 | 0.83 |
| Precision | 1 | 1 | 1 |
| Recall | 0.00033 | 0.00033 | 0.00033 |
| F1-score | 0.00066 | 0.00066 | 0.00066 |
| AUROC | 0.6980 | 0.6932 | 0.6994 |
| AUPR | 0.16 | 0.67 | 0.3 |

표 4. Batch size 64의 성능 평가표

Table 4. Evaluation of batch size 64.

| | all | 1:1 | 1:5 |
|-----------|--------|--------|--------|
| Accuracy | 0.92 | 0.5 | 0.83 |
| Precision | 0.54 | 1 | 0.65 |
| Recall | 0.0012 | 0.0012 | 0.0012 |
| F1-score | 0.0024 | 0.0024 | 0.0024 |
| AUROC | 0.6980 | 0.6960 | 0.6980 |
| AUPR | 0.16 | 0.67 | 0.3 |

표 5. Batch size 128의 성능 평가표

Table 5. Evaluation of batch size 128.

| | all | 1:1 | 1:5 |
|-----------|--------|--------|--------|
| Accuracy | 0.92 | 0.5 | 0.83 |
| Precision | 0.65 | 1 | 0.81 |
| Recall | 0.0011 | 0.0011 | 0.0011 |
| F1-score | 0.0021 | 0.0021 | 0.0021 |
| AUROC | 0.6960 | 0.6967 | 0.6956 |
| AUPR | 0.16 | 0.67 | 0.3 |

표 6. CHA₂DS₂-VASc의 성능 평가표

Table 6. Evaluation of the CHA₂DS₂-VASc score.

| n | Accuracy | AUROC |
|-----|----------|--------|
| 9 | 0.92 | - |
| 8 | 0.92 | - |
| 7 | 0.91 | - |
| 6 | 0.89 | - |
| 5 | 0.85 | - |
| 4 | 0.77 | - |
| 3 | 0.62 | - |
| 2 | 0.42 | - |
| 1 | 0.21 | - |
| all | - | 0.6448 |

본 연구에서는 심방세동 환자의 뇌졸중 여부를 예측하기 위해 딥러닝을 이용한 모델을 개발하였으며, 이는 기존의 통계적 모델보다 개선된 성능을 보여주고 있다. 그러나 실제 환자에게 적용하기엔 다소 낮은 성능을 보이는 것이 사실이므로 실제 도입을 위해서는 모델을 더 개선해야 할 필요가 있다. 개선 후 뇌졸중을 사전 예측해 발병으로 일어나는 후유증이나 사망의 사전 방지를 기대할 수 있다. 향후 추가적인 건강검진 데이터를 추가하고 다른 기법을 사용하는 등 모델 성능 개선 관련 연구를 지속할 계획이며, 기계학습 기법과의 비교 분석을 통해 딥러닝

모델의 우수성을 증명할 수 있기를 기대한다.

Acknowledgments

본 연구는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020R1C1C1006007).

References

[1] Global Health Estimates: Life expectancy and leading causes of death and disability [Internet]. Available: <https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates>

[2] World Health Organization (2005). WHO STEPS Stroke Manual [Internet]. Available: <http://whqlibdoc.who.int/chp/steps/Stroke/en/>

[3] Korean Statistical Information Service (KOSIS). Annual Report on the Cause of Death Statistics [Internet]. 2016. Available: <https://kosis.kr/eng/search/searchList.do>

[4] Stroke Risk in Atrial Fibrillation Working Group, "Independent predictors of stroke in patients with atrial fibrillation: a systematic review," *Neurology*, Vol. 69, No. 6, pp. 546-554, Aug. 2007.

[5] J. B. Olesen, C. Torp-Pedersen, M. L. Hansen, and G. Y. H. Lip, "The value of the CHA2DS2-VASc score for refining stroke risk stratification in patients with atrial fibrillation with a CHADS2 score 0 - 1: a nationwide cohort study," *Thrombosis and Haemostasis*, Vol. 107, No. 6, pp. 1172-1179, 2012.

[6] Y. Bengio, A. Courville, and P. Vincent, "Representation

learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 8, pp. 1798-1828, Aug. 2013.

[7] J. Schmidhuber, "Deep learning in neural networks: An Overview", *Neural Networks*, Vol. 61, pp. 85-117, Jan. 2015.

[8] M. M. Lau and K. Hann Lim, "Review of adaptive activation function in deep neural network," in *Proceedings of the 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, Sarawak: Malaysia, pp. 686-690, 2018.

[9] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, "On optimization methods for deep learning," in *Proceedings of the 28th International Conference on Machine Learning*, Bellevue: WA, pp. 265-272, Jun. 2011.

[10] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *International Journal of Machine Learning Technology*, Vol. 2, No. 1, pp. 37-63, 2011.

[11] J. A. Hanley and J. M. Barbara, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*. Vol. 143, No. 1, pp. 29-36, 1982.

[12] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, Vol. 30, No. 7, pp. 1145-1159, 1997.

[13] J. Keilwagen, I. grosse, and J. Grau, "Area under precision-recall curves for weighted and unweighted data", *PloS One*, Vol. 9, No. 3, Mar. 2014.

[14] J. Davis, and M. Goadrich, "The relationship between precision-recall and ROC curves.", in *Proceedings of the 23rd international conference on Machine learning*, New York: NY, pp. 233-240, Jun 2006.



정 선 우 (Sunwoo Jung)

2021년 2월 : 전남대학교 인공지능 융합전공 (이학사)
※ 관심분야 : 헬스케어, 메디컬 AI



이 민 지 (Minji Lee)

2021년 2월 : 전남대학교 인공지능 융합전공 (이학사)
※ 관심분야 : 메디컬 AI



유 선 용 (Sunyong Yoo)

2018년 8월 : 한국과학기술원 바이오및뇌공학과 (공학박사)
2019년 9월 ~ 현재 : 전남대학교 ICT융합시스템공학과 조교수
※ 관심분야 : 의료정보학, 인공지능, 빅데이터