

VAE 기반 데이터 불균형 개선을 통한 치매 조기 탐지 기법

이민지*, 이석훈**

Early Detection Technique for Dementia by Improving VAE-based Data Imbalance

Minji Lee*, Sukhoon Lee**

요약

인구 고령화로 인해 치매 질환이 중대한 문제로 떠오르고 있으며, 치매의 조기 탐지를 위하여 라이프로그 데이터를 활용한 연구들이 진행되었다. 하지만, 이러한 라이프로그는 데이터 수집의 어려움 및 프라이버시 등으로 인해 데이터 불균형 문제를 야기시킨다. 본 논문은 이러한 데이터 불균형 문제를 해결하고자 변분 오토인코더(VAE) 기반의 데이터 불균형 개선을 통한 치매 조기 탐지 기법을 제안한다. 제안 기법은 VAE를 사용하여 데이터를 증강시키고, XGBoost, DNN, TabNet, Wide & Deep Learning 모델을 사용하여 치매를 분류하는 모델을 구현한다. 제안된 기법의 성능을 평가하기 위해 AI-Hub에서 제공하는 치매 고위험군 웨어러블 라이프로그 데이터셋을 활용한다. 비교 평가 결과, 4가지 모델 중 Wide & Deep Learning 모델이 가장 높은 성능을 보였으며, 제안 기법이 치매 데이터에 대하여 효과적으로 데이터 불균형 문제를 해결하였음을 확인하였다.

Abstract

With the progression of population aging, dementia has emerged as a significant issue, and studies utilizing lifelog data have been conducted to enable the early detection of dementia. However, lifelog data often face challenges such as difficulties in data collection and privacy concerns, which lead to data imbalance issues. This paper proposes a method to address the data imbalance problem by improving data balance using a Variational Autoencoder (VAE) for early dementia detection. The proposed method augments data through VAE and constructs a classification model for dementia using XGBoost, DNN, TabNet, and Wide & Deep Learning models. To evaluate the performance of the proposed method, the wearable lifelog dataset for high-risk dementia groups provided by AI-Hub was utilized. The results of the comparative evaluation showed that the Wide & Deep Learning model achieved the highest performance among the four models, confirming that the proposed method effectively resolves the data imbalance issue in dementia datasets.

Keywords

augmentation technique, data imbalance, lifelog data, deep learning, dementia classification

* 국립군산대학교 소프트웨어학과 학사과정
- ORCID: <https://orcid.org/0000-0002-0788-721X>
** 국립군산대학교 소프트웨어학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0002-3390-5602>

• Received: Nov. 14, 2024, Revised: Jan. 15, 2025, Accepted: Jan. 18, 2025
• Corresponding authors: Sukhoon Lee
Dept. of Software Science & Engineering, Kunsan National University, Korea
Tel.: +82-63-469-8914, Email: leha82@kunsan.ac.kr

1. 서론

급속히 진행되는 인구 고령화로 인해 치매는 현대 사회에서 중대한 건강 문제로 떠오르고 있다[1]. 치매는 기억력, 사고력, 행동 및 일상생활 수행 능력의 저하를 특징으로 하는 신경퇴행성 질환으로, 연령, 유전적 요인, 생활 습관 등 다양한 요인에 의해 발생할 수 있다[2]. 치매의 증가는 개인의 삶의 질 저하뿐만 아니라 사회 전체에 막대한 경제적 부담을 초래하고 있다[3]. 이러한 개인적, 사회적 부담을 줄이고 치매 환자의 삶의 질을 향상시키기 위해서는 조기 진단과 적절한 관리가 필수적이다[4].

그러나 현재의 치매 진단 방법은 여러 가지 한계가 있다. 주로 사용되는 MRI나 PET 스캔과 같은 고가의 의료 장비는 비용 부담이 크고, 환자가 직접 병원을 방문해야 한다는 접근성 문제로 조기 진단에 장벽이 되고 있다[5].

이러한 상황에서 웨어러블 기기의 발전과 함께 주목받고 있는 라이프로그 데이터는 치매 조기 진단의 새로운 가능성을 제시하고 있다. 기존 연구들에 따르면, 웨어러블 센서로 수집한 일상적인 보행 데이터가 인지 장애 선별에 유용할 뿐만 아니라 치매 진행의 초기 경고 신호로 활용될 수 있다는 점이 확인되었다[6]. 또한 저가의 센서를 통해 수집된 일상 행동 데이터를 머신러닝 모델로 분석하여 치매 환자의 이상 행동을 탐지하는 연구도 진행되었다[7]. 이러한 연구 결과는 일상생활에서 수집 가능한 데이터를 통해 치매의 조기 징후를 효과적으로 감지할 수 있다는 가능성을 보여준다.

그러나 라이프로그 데이터를 활용한 연구에서는 데이터 불균형 문제가 중요한 과제로 대두되고 있다. 데이터 불균형이란 특정 클래스에 속하는 샘플 수가 다른 클래스에 비해 매우 적은 상황을 의미하며, 이로 인해 딥러닝 모델은 대다수 클래스에 편향되기 쉽고 특히 소수 클래스에 대한 예측 성능이 저하되는 문제가 발생할 수 있다[8].

이러한 불균형이 발생하는 주요 원인으로는 희귀한 이벤트의 발생, 데이터 수집의 어려움, 그리고 프라이버시 문제 등이 있으며, 특히 치매와 같은 의료 연구에서는 정상인 데이터에 비해 환자 데이터의 수집이 제한적일 수밖에 없어 데이터가 불균형

하게 나타났다[9].

이와 같은 데이터 불균형 문제를 해결하기 위해 다양한 기법들이 제안되어 왔다. 기존 연구들에 따르면, SMOTE(Synthetic Minority Over-sampling Technique) 기법을 의료 데이터에 적용하여 소수 클래스의 재현율을 크게 향상시킨 사례가 있으며[10], 골다공증 데이터를 대상으로 다양한 언더샘플링과 오버샘플링 방법의 효과를 분석한 연구도 진행되었다[11].

그러나 이러한 기존 샘플링 기법들은 라이프로그와 같은 복잡한 고차원 데이터에 적용할 경우 효과가 제한적일 수 있다. 이에 비해 변분 오토인코더(VAE, Variational Autoencoder)[12]는 입력 데이터의 확률 분포를 학습하여 이로부터 새로운 샘플을 생성할 수 있어 라이프로그 데이터의 특성을 보존하면서도 더 나은 샘플링 효과를 기대할 수 있다.

따라서, 이 논문은 데이터 불균형 개선을 통한 치매 조기 탐지 기법을 제안한다. 이를 위하여 VAE를 사용해 데이터를 증강한 후 딥러닝 모델을 학습한다. 이후 성능 평가를 통해 가장 우수한 모델을 선정하여 제안 기법에 활용한다.

제안 기법은 치매 진단을 위한 분류 모델의 정확도를 향상시킨다. 이는 환자의 접근성이 높은 라이프로그 데이터를 활용하므로 고비용의 장비를 활용한 전문의의 진단 이전에 보다 효율성 높은 방식으로 환자의 치매 여부를 조기에 탐지할 수 있음을 보인다.

이 논문의 구성은 다음과 같다. II장에서는 관련 연구를 소개하고, III장에서는 제안된 모델의 구조와 기법을 설명하며, IV장에서는 실험 설계와 결과를 분석한다. 마지막으로 V장에서 결론을 제시한다.

II. 관련 연구

이 장에서는 데이터 증강 및 불균형 데이터 처리 기법과 기계학습 기반 치매 진단 기법에 대한 관련 연구를 기술한다.

2.1 불균형 데이터 처리 기법

데이터 불균형 문제를 해결하기 위한 다양한 접근 방식이 연구되었다.

기본적인 방법인 언더샘플링[13]은 다수 클래스의 데이터를 줄여 균형을 맞추는 방식으로, 구현이 간단하나 정보 손실의 위험이 있다.

SMOTE와 앙상블 학습을 활용한 연구[14]와 연합 학습과 Balanced-MixUp을 결합한 연구[15]는 소수 클래스의 예측 성능을 향상시켰으나, 생성된 데이터의 특성이 원본과 차이가 있고 과적합 문제가 발생할 수 있다는 한계를 보였다.

최근에는 VAE를 활용한 데이터 증강 연구[16]가 수행되어 MNIST와 OASIS 데이터셋에서 성능 향상을 보였으나, 데이터 특성에 따라 증강 효과가 일관되지 않다는 제한점이 있었다.

2.2 기계학습 기반 치매 진단 기법

라이프로그 데이터를 활용한 치매 예측 연구들이 다수 진행되었다. AutoML을 활용한 연구[17], 랜덤 포레스트와 SHAP 분석을 결합한 연구[18], 인공지능 기술을 적용한 연구[19] 등이 수행되었으나, 이들 연구는 공통적으로 데이터 불균형 문제와 데이터의 다양성 부족이라는 한계점을 지니고 있다.

의료 영상 데이터를 활용한 연구들도 활발히 진행되었다. CNN을 활용한 MRI 기반 연구[20]는 90% 이상의 정확도를 달성했다.

시계열 데이터 처리에 특화된 모델들을 활용한 연구도 주목할 만하다. CNN-RNN 하이브리드 모델을 활용한 EEG 분석 연구[21]는 각각 90%, 88% 이상의 높은 정확도를 달성했다. 특히 웨어러블 기기의 수면과 활동 데이터를 통합 분석한 LSTM 연구[22]는 92.72%의 정확도를 보였으나, 제한된 샘플 수와 단기간 데이터 활용이라는 한계가 있었다.

앞선 연구들은 몇 가지 주요한 한계점을 보였다. 대부분의 연구가 목표변수를 인지 기능 정상군(CN, Cognitively Normal)과 인지 기능 장애 군(CI, Cognitively Impaired)으로 이분화하여 분류함으로써, 경도 인지 장애(MCI, Mild Cognitive Impairment)와 치매(Dem, Dementia)의 구별을 어렵게 만들어 조기 진단과 적절한 개입의 기회를 제한하였다.

또한 기계학습 기법을 적용한 연구들은 주로 데이터 불균형 문제를 해소하지 못했다는 한계가 있었다. 치매 환자 데이터는 일반적으로 불균형 문제

가 심각하게 나타난다. 치매 관련 데이터셋에서는 정상군에 비해 치매 환자군의 데이터 수가 현저히 적어 학습 과정에서 모델이 정상 데이터를 우선으로 학습하거나 치매 환자군을 잘 예측하지 못하는 경향을 보인다. 이러한 불균형 문제를 해결하지 않으면 모델의 성능이 왜곡될 가능성이 높아지며, 특히 치매 환자군의 재현율이 낮아질 수 있다.

이를 개선하기 위해, 본 연구에서는 VAE와 같은 데이터 증강 기법을 활용하여 데이터 불균형 문제를 완화하고자 하였다. VAE는 잠재 공간에서 샘플링한 데이터를 통해 새로운 치매 환자 데이터를 생성함으로써 부족한 데이터를 보충하는 역할을 수행한다.

딥러닝을 활용한 치매 진단 연구들은 대부분 MRI나 EEG와 같은 영상 데이터에 집중되어 있다는 한계점을 지닌다. 이러한 한계점들을 극복하고자 이 논문은 라이프로그 데이터를 활용하여 인지 기능 정상군, 경도 인지 장애, 치매를 각각 CN, MCI, Dem으로 정의하고 3개의 클래스를 분류하는 다중 클래스 분류 모델을 개발한다.

III. 치매 환자 데이터 분석

이 장은 VAE를 활용한 데이터 불균형 처리에 관해 기술한다. 사용한 데이터의 설명과 데이터 불균형의 문제점, VAE에 대해 설명하고, 제안 기법의 구조를 보인다. 이후 딥러닝 학습을 위한 모델 학습의 구조에 관해 기술한다.

3.1 치매 환자 라이프로그 데이터셋

이 논문은 치매 환자 데이터 분석을 위하여 한국 지능정보사회진흥원의 AI-Hub[23]에서 제공하는 '치매 고위험군 웨어러블 라이프로그' 데이터셋을 활용한다. 이 데이터셋은 55세 이상의 성인 300명을 대상 수집되었으며 전문의의 정밀 진단을 통해 참가자들은 CN 및 MCI, Dem 클래스로 분류한다. 이 데이터셋에서는 무증상 치매(AAD)를 CN에 포함한다.

이 논문은 활동 데이터와 수면 데이터셋을 바탕으로 분석을 진행하였으며 간이 인지검사 데이터(MMSE, Mini-Mental State Examination)는 사용하지

4 VAE 기반 데이터 불균형 개선을 통한 치매 조기 탐지 기법

않는다. 12,183건의 데이터의 형태는 모두 수치형 데이터로 구성되어 있다.

각 라이프로그는 개인의 하루 동안의 정보를 담고 있으며, 활동성 변수 28개와 수면 변수 33개를 포함한다. 구체적으로, 수면 데이터(수면 시작과 종료 시간, 수면 점수, 수면 효율, REM수면 시간, 수면의 깊이 등)와 활동 데이터(활동 시작과 종료 시간, 운동 시간, 활동 점수, 신진 대사량, 움직인 거리, 칼로리 소모량 등)가 5분 단위로 수집되어 24시간 동안 추적되었다.

이 논문은 활동 데이터에 포함된 식별자인 이메일 변수와 5분 활동 로그, 활동 시작 시간, 활동 종료 시간, 1분 MET 로그와 같은 비정량화 변수는 분석에서 제외하였으며 비활동 알람 횟수, 미착용 시간과 같은 대상자의 활동이나 상태를 반영하지 못하는 변수들도 분석에서 제외하였으며 사용된 변수는 22개로 표 1과 같다.

표 1. 활동 데이터와 관련된 특성

Table 1. Features related to activity data

Features	Description
activity_average_met	Average MET
activity_cal_active	Active calories
activity_cal_total	Total calories burned
activity_daily_movement	Daily movement distance
activity_high	High-intensity activity time
activity_inactive	Inactive time
activity_low	Low-intensity activity time
activity_medium	Medium-intensity activity time
activity_met_min_high	High-intensity activity MET
activity_met_min_inactive	Inactive MET
activity_met_min_low	Low-intensity activity MET
activity_met_min_medium	Medium-intensity activity MET
activity_rest	Rest time
activity_score	Activity score
activity_score_meet_daily_targets	Daily target achievement score
activity_score_move_every_hour	Hourly movement score
activity_score_recovery_time	Recovery time score
activity_score_stay_active	Activity maintenance score
activity_score_training_frequency	Training frequency score
activity_score_training_volume	Training volume score
activity_steps	Daily steps
activity_total	Total activity time (minutes)

위와 같이 수면 데이터에 포함된 수면 종료 시간, 시작 시간, 5분 당 심박동 로그, 수면 상태 로그, 본 수면 여부, 수면 식별 아이디 등의 변수들도 분석에서 제외하였으며 사용된 변수는 24개로 표 2와 같다.

표 2. 수면 데이터와 관련된 특성

Table 2. Features related to sleep data

Features	Description
sleep_awake	Awake time
sleep_breath_average	Average breaths per minute
sleep_deep	Deep sleep time
sleep_duration	Total sleep duration
sleep_efficiency	Sleep efficiency
sleep_hr_average	Average heart rate (BPM)
sleep_hr_lowest	Lowest heart rate (BPM)
sleep_light	Light sleep time
sleep_midpoint_at_delta	Sleep midpoint delta time
sleep_midpoint_time	Sleep midpoint time
sleep_onset_latency	Sleep onset latency
sleep_rem	REM sleep time
sleep_restless	Restlessness ratio
sleep_rmssd	Average heart rate variability (RMSSD)
sleep_score	Overall sleep score
sleep_score_alignment	Sleep timing score
sleep_score_deep	Deep sleep score
sleep_score_disturbances	Sleep disturbance score
sleep_score_efficiency	Sleep efficiency score
sleep_score_latency	Sleep latency score
sleep_score_rem	REM sleep score
sleep_score_total	Sleep contribution score
sleep_temperature_delta	Skin temperature delta
sleep_temperature_deviation	Skin temperature deviation

3.2 문제 정의

이 논문에서 사용하는 데이터셋은 심각한 불균형 문제를 가지고 있다. 표 3에서 볼 수 있듯이, CN 클래스가 7,737개로 가장 많은 데이터를 보유하고 있지만, MCI 클래스는 3,661개, Dem 클래스가 785개로 클래스 간 데이터 수의 차이가 크다.

특히 치매군의 경우 정상군 대비 약 10%에 불과한 데이터만을 보유하고 있어 극심한 데이터 불균형 상태이다.

표 3. 분류별 라이프로그 개수

Table 3. Number of lifelogs per category

Group	CN	MCI	Dem
number of person	111	51	12
number of record	7,737	3,661	785

이러한 데이터 불균형은 모델 학습 시 심각한 문제를 야기한다. 여러 모델을 학습한 결과, 훈련 손실(Train loss)은 지속적으로 감소하는 반면 검증 손실(Validation loss)은 오히려 증가하는 심각한 과적합 현상이 발생하였다. 이는 모델이 다수 클래스의 특징에 과도하게 적응하여 소수 클래스에 대한 일반화 성능이 저하되는 전형적인 불균형 데이터 학습의 문제점을 보여준다. 실제로 전처리를 하지 않은 원본 데이터를 학습하였을 때 Dem 클래스의 정밀도는 높았으나 재현율이 낮아 F1-score 성능이 저하되는 것을 확인할 수 있었다. 이는 실제 Dem 클래스 사례 중 모델이 정확하게 예측하지 못한 사례가 많다는 뜻으로, 환자들이 필요한 조기 진단과 치료를 받지 못할 수 있음을 시사한다.

이와 같은 데이터 불균형의 문제점과 기존 해결 방법들의 한계를 고려할 때 보다 효과적이고 새로운 접근 방식의 필요성이 대두된다.

IV. VAE 기반의 치매 조기 탐지 기법

4.1 데이터 불균형 개선 기법

VAE는 오토인코더를 확장한 형태로 데이터의 잠재 공간을 확률적으로 모델링하는 방식을 채택한다.

이를 통해 VAE는 기존 오토인코더가 단순히 데이터를 압축하는 것을 넘어 새로운 데이터를 생성하거나 재구성할 수 있는 능력을 갖추게 되며 인코더(Encoder), 잠재 공간(Latent space), 디코더(Decoder)로 구성된다[24].

인코더는 입력 데이터를 잠재 공간의 분포 파라미터로 변환하는 신경망으로, 입력 데이터를 저차원 벡터로 압축하여 잠재 공간에 매핑한다. 잠재 공간에서는 재파라미터화 트릭을 사용하여 인코더가 출력한 잠재 분포에서 미분 가능한 방식으로 샘플링을 수행하며 이는 신경망이 샘플링 과정에서도 역전파를 통해 학습할 수 있게 한다. 디코더는 샘플링

된 잠재 변수를 사용하여 원본 입력 데이터를 재구성하는 신경망으로, 잠재 공간에서 다시 데이터 공간으로의 매핑을 수행한다[25].

VAE의 손실 함수는 재구성 손실과 KL 발산으로 구성되며 식 (1), (2)과 같다. 재구성 손실은 원본 데이터와 재구성된 데이터 간의 차이를 측정하며, 주로 평균 제곱 오차(MSE) 또는 로그 가능도로 계산된다.

$$E[\log p(x|z)] \approx \frac{1}{L} \sum_{l=1}^L \log p(x|z^{(l)}) \quad (1)$$

$$D_{KL}[q(z|x)||p(z)] = \frac{1}{2} \sum (1 + \log \sigma_j^2 - \mu_j^2 - \sigma_j^2) \quad (2)$$

VAE는 GAN(Generative Adversarial Network)에 비해 더 안정적인 학습이 가능하며 확률 분포를 직접 추정할 수 있다는 장점이 있다[26]. 특히 VAE는 고차원 데이터의 복잡한 패턴을 효과적으로 학습하고, 잠재 공간에서 소수 클래스의 특성을 정확히 반영하는 새로운 샘플을 생성할 수 있어 불균형 데이터 문제 해결에 적합하다.

4.2 제안 기법

그림 1은 이 논문에서 제안하는 데이터 불균형 개선을 위한 VAE 기반 치매 조기 탐지 기법의 구조를 보여준다. 제안 기법은 데이터 전처리, VAE 기반 데이터 증강, 그리고 모델 학습 및 분류 과정으로 구성된다.

데이터 전처리 단계에서는 잘못된 패턴 증폭을 방지하기 위해 이상치(Outlier) 제거가 필수적이며, 이를 위해 Isolation Forest[27] 기법을 활용하였다. Isolation Forest는 데이터 포인트를 이진 트리 구조로 분리하여 이상치를 탐지하는 알고리즘으로, 라이프로그 데이터와 같이 명확한 이상치 라벨링이 어렵거나 정상 데이터만 존재하는 환경에서도 효과적으로 이상치를 탐지할 수 있는 장점이 있다.

이후, VAE를 사용하여 소수 클래스의 학습 데이터를 증강하였다. 증강된 데이터를 기반으로 Classifier를 학습하기 전에, 범주형 변수는 라벨 인코딩을, 수치형 변수는 StandardScaler[28]를 사용하여 표준화를 수행하였다.

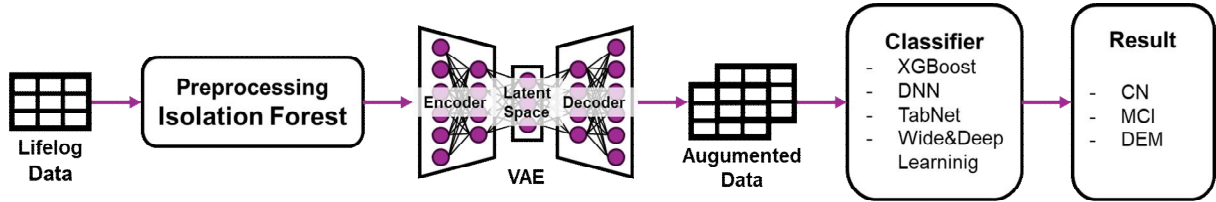


그림 1. 제안 기법
Fig. 1. Proposed method

제안하는 기법의 성능을 평가하기 위해 기계 학습 모델인 XGBoost와 딥러닝 모델인 DNN, TabNet, Wide & Deep Learning을 비교 대상으로 사용하였다.

XGBoost[29]는 그래디언트 부스팅 알고리즘 기반의 앙상블 모델로, 빠른 학습 속도와 높은 예측 정확도를 제공하며 특성 중요도 분석이 가능하다.

DNN[30]은 다층 퍼셉트론(Multi-layer perceptron)을 기반으로 비선형성을 효과적으로 학습하는 신경망 모델이다.

TabNet[31]은 순차적 주의 메커니즘을 활용해 테이블 데이터에서 중요한 특성을 학습하며 복잡한 상호작용을 다룰 수 있다.

Wide & Deep Learning[32]은 선형 모델(Wide)과 심층 신경망(Deep)을 결합해 선형 관계와 비선형 패턴을 동시에 학습하는 하이브리드 모델이다.

마지막으로, Classifier 학습 결과는 CN, MCI, Dem 클래스로 데이터를 분류하여 예측하였다.

V. 실험 및 평가

5.1 실험 방법

이 논문은 데이터 증강을 통해 불균형 데이터를 개선하고 이를 기반으로 치매 조기 탐지 성능을 향상시키는 것을 목적으로 한다. 실험은 표 4에 제시된 환경에서 수행되었다.

표 4. 실험 환경
Table 4. Experimental environment

Type	Spec
OS	Windows 10 Pro
CPU	12th Gen Intel(R) Core(TM) i5-12600
GPU	NVIDIA GeForce RTX 3050

데이터 전처리 과정에서는 이상치 제거를 위해 각 특성값의 상위 및 하위 10% 범위를 벗어나는 데이터를 제외하였다. 여러 임계값에 대해 실험한 결과, 10%로 설정했을 때 가장 우수한 성능을 보여 최종적으로 채택하였다. 이를 통해 CN 클래스는 7,737개에서 7,075개로, MCI 클래스는 3,661개에서 3,374개로, Dem 클래스는 785개에서 515개로 감소하였다. 증강된 데이터를 기존 훈련 데이터셋과 통합하여 최종 훈련 데이터셋을 구성하였다.

제안하는 VAE 모델의 구조는 그림 2와 같다. 입력 차원은 원본 데이터의 특성 수와 동일하게 설정하였으며, 잠재 공간의 차원은 500으로 구성하였다. 인코더와 디코더는 각각 512개와 256개의 뉴런을 가진 두 개의 완전 연결 층으로 이루어져 있으며, 과적합 방지를 위해 배치 정규화와 30% 드롭아웃을 적용하였다. 손실 함수는 재구성 오차와 KL 발산의 가중합으로 설정하였으며, 최적화에는 Adam 옵티마이저와 학습률 1e-4를 사용하였다.

학습된 VAE의 디코더를 사용하여 표준 정규 분포에서 샘플링한 잠재 벡터로 새로운 Dem 샘플을 생성하였다. 생성된 데이터의 재구성 오류는 0.0002로, 이는 입력 데이터와 모델이 재구성한 출력 데이터 간의 차이를 의미한다.

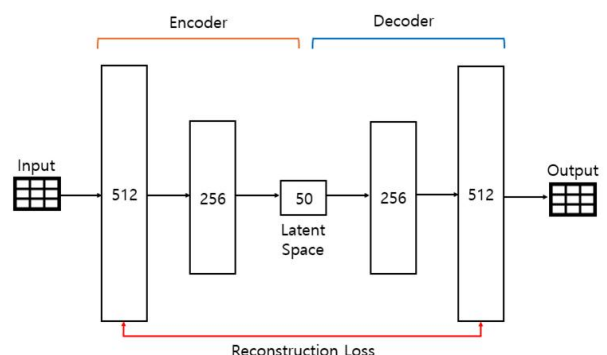


그림 2. 제안 기법의 VAE 구조
Fig. 2. VAE structure of a proposed method

낮은 재구성 오류는 모델이 입력 데이터의 특징을 잘 학습하여 출력 데이터가 입력 데이터와 거의 동일함을 나타낸다.

전처리 된 데이터는 학습, 검증, 실험 데이터셋으로 8:1:1의 비율로 분할하였으며 최종 학습 데이터는 표 5와 같다.

표 5. 최종 학습 데이터
Table 5. Final training data

Data	CN	MCI	Dem
Train	5660	2699	4412
Test	707	337	51
Valid	708	338	52

제안하는 기법의 성능을 평가하기 위해 사용한 모델들의 하이퍼파라미터는 다음과 같이 설정하였다. XGBoost는 다중 클래스 분류를 위해 softmax 함수, 최대 트리 깊이 6, 학습률 0.1로 설정하였다.

DNN은 512-256-128-64-32 노드로 구성된 5개의 완전 신경연결망 레이어로 설계되었으며, 각 층에 ReLU 활성화 함수와 L2 정규화, 배치 정규화, 50% 드롭아웃을 적용하였다.

TabNet은 특성 변환 차원과 주목(Attention) 차원을 각각 64로 설정하고, 5단계의 결정 과정을 거치도록 구성하였다.

Wide & Deep Learning의 Deep 컴포넌트는 256-128-64 노드 구조로 설계되었으며, ReLU 활성화 함수와 30% 드롭아웃을 적용하였다.

모든 딥러닝 모델은 교차 엔트로피 손실 함수와 Adam 옵티마이저를 사용하여 학습을 수행하였다.

5.2 실험 결과 및 비교 평가

치매 예측 모델의 성능을 평가하기 위해 정밀도(Precision), 재현율(Recall), F1-Score를 주요 평가 지표로 사용하였다. 이러한 지표들은 특히 치매와 같은 의료 진단 분야에서 모델의 성능을 다각도로 분석하는 데 중요한 의미가 있다.

정밀도는 모델이 특정 클래스로 예측한 사례 중 실제로 해당 클래스에 속하는 비율을 나타낸다. 예를 들어 Dem 클래스에 대한 정밀도의 경우, 모델이

Dem 클래스로 예측한 환자 중 실제로 치매 군인 환자의 비율을 의미하며 식 (3)과 같이 계산된다.

$$precision = \frac{TP}{(TP + FP)} \quad (3)$$

이때 TP는 실제 사례가 긍정(Positive)이고 모델도 긍정이라고 올바르게 예측한 사례의 수이며, FP는 실제 사례는 부정(Negative)인데 모델이 잘못 예측한 사례의 수를 의미한다.

재현율은 실제 특정 클래스에 속하는 사례 중 모델이 정확하게 예측한 비율을 의미한다. 즉, 실제 치매군 환자 중에서 모델이 치매 군으로 정확하게 식별한 비율로, 식 (4)와 같이 계산된다.

$$recall = \frac{TP}{(TP + FN)} \quad (4)$$

이때 TP는 정밀도와 마찬가지로 실제 사례가 긍정인데 모델이 긍정이라고 올바르게 예측한 사례의 수이며, FN은 실제 사례는 긍정인데 모델이 부정이라고 잘못 예측한 사례의 수를 의미한다.

F1-score는 정밀도와 재현율의 조화 평균으로, 두 지표 간의 균형을 평가하는 지표이며 식 (5)와 같이 계산된다.

$$F1 - score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (5)$$

치매의 조기 발견은 적절한 치료와 관리를 통해 질병의 진행을 늦출 수 있는 중요한 기회를 제공한다. 특히 CN 클래스를 MCI 클래스로 잘못 분류하는 것은 적절한 치료 기회를 놓치게 만들어 심각한 결과를 초래할 수 있다. 반면, 정상군이나 경도 인지 장애 군을 치매 군으로 잘못 분류하는 경우는 추가 검사를 통해 확인이 가능하다. 따라서 이 논문에서는 재현율을 가장 중요한 평가 지표로 고려하되, 불필요한 오진을 최소화하기 위해 정밀도와 F1-Score도 함께 평가하였다.

그림 3은 각 모델의 F1-score를 클래스별로 시각화하고 있으며, 평균(Avg) F1-score를 함께 나타내어 모델의 전체적인 성능을 평가하고 있다.

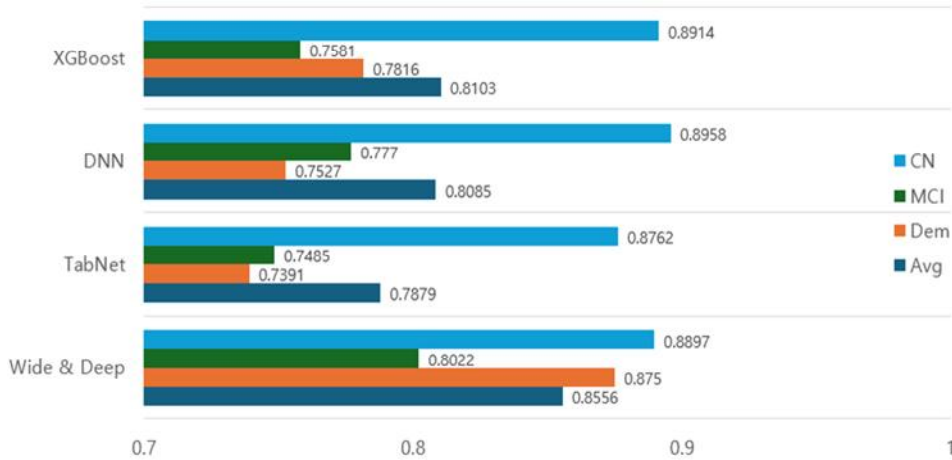


그림 3. 각 모델의 F1-score 클래스별 성능 비교
 Fig. 3. Performance by F1-score class for each model

XGBoost 모델은 CN 클래스에서 0.8914의 F1-score를 기록하여 가장 높은 성능을 보였으며, 이는 XGBoost가 CN 클래스 분류에서 특히 우수한 성능을 나타냄을 시사한다. 반면, Dem 클래스에서는 0.7816의 F1-score를 기록하여 CN 클래스에 비해 성능이 낮았으나, 여전히 의미 있는 성능을 보였다. 그러나 MCI 클래스에서는 0.7581의 F1-score로 CN과 Dem 클래스에 비해 상대적으로 낮은 성능을 나타냈다. 평균 F1-score는 0.8103으로, CN 클래스의 높은 성능이 전체 평균을 높이는 데 기여하였다.

Wide & Deep Learning 모델은 XGBoost 모델과 비교하면 모든 클래스에서 더 높은 성능을 보였다. Dem 클래스에서 F1-score가 0.8556으로 나타났으며, 이는 XGBoost 모델의 Dem 클래스 F1-score인 0.7816보다 현저히 높은 수치이다. 또한, MCI와 CN 클래스에서도 각각 0.8022와 0.8897의 F1-score를 기록하여 XGBoost 모델보다 전반적으로 우수한 성능을 보였다. 평균 F1-score 역시 0.875로 Wide & Deep Learning 모델이 데이터 불균형 문제에 대한 보다 효과적인 대처 능력을 보여주었다.

TabNet 모델은 Dem 클래스에서 0.7391의 F1-score를 기록하여 XGBoost 모델보다 낮은 성능을 나타냈다. MCI 클래스에서는 0.7485, CN 클래스에서는 0.8762의 F1-score를 보여 전반적으로 XGBoost보다 낮은 성능을 기록하였다. 평균 F1-score는 0.7879로, XGBoost 모델의 평균 F1-score인 0.8103보다 낮게 나타나, TabNet 모델이 XGBoost에 비해 상

대적으로 불리한 성능을 보인다.

DNN 모델은 Dem 클래스에서 0.7527, MCI 클래스에서 0.777, CN 클래스에서 0.8958의 F1-score를 기록하였다. XGBoost 모델과 비교했을 때 Dem 클래스와 MCI 클래스에서는 DNN 모델이 소폭 낮은 성능을 보였으나, CN 클래스에서는 0.8958로 XGBoost의 0.8914보다 약간 더 높은 성능을 보였다. 평균 F1-score는 0.8085로 XGBoost 모델의 0.8103과 유사한 성능을 보였다.

종합적으로 Wide & Deep Learning 모델이 전반적으로 가장 높은 F1-score를 기록하여 모든 클래스에서 균형 잡힌 성능을 보이며 데이터 불균형 문제에 대한 효과적인 대처 능력을 입증하였다.

표 6은 Wide & Deep Learning 모델의 데이터 증강 전후 성능을 비교한 결과를 보인다. 전체적인 성능 측정 결과 세 클래스의 평균 F1-score는 데이터 증강의 전후 각각 86.2%에서 85.6%로 하락하였다.

하지만 이 논문은 평가 지표에 대하여, 정상인인 CI를 판별하는 것 보다 경도 인지 장애인 MCI 혹은 치매인 Dem을 판별하는 것이 더 중요하며, 실제 존재하는 MCI나 Dem을 예측할 수 있는 수치인 재현율을 더 중요한 지표로 설정하였다.

그 결과, MCI와 Dem 클래스는 재현율이 각각 83.7%에서 88.4%, 76.4%에서 82.3%로 상승하여, CN 클래스가 92.2%에서 85.0%로 성능이 대폭 하락되었음에도 불구하고 클래스들의 평균 재현율은 84.1%에서 85.3%로 증가하였다.

표 6. 증강 데이터 wide & deep learning 성능 비교
Table 6. Augmented data wide & deep learning performance comparison

Class	Preprocessed data (without augmented)			Proposed method (using VAE)		
	precision	recall	F1-score	precision	recall	F1-score
CN	0.9107	0.9223	0.9165	0.9332	0.8501	0.8897
MCI	0.8398	0.8373	0.8385	0.7340	0.8843	0.8022
Dem	0.9070	0.7647	0.8298	0.9333	0.8235	0.8750
Avg.	0.8858	0.8414	0.8616	0.8668	0.8526	0.8556

이 논문에서 제안하는 방법은 기존의 방법과 비교하였을 때 재현율에 대해 더 높은 성능을 보인다. 이는 제안 방법이 숨겨진 치매 환자들을 더 정확히 탐지해냄으로써 치매 조기 진단에 유용하게 활용될 수 있을 것으로 판단된다.

VI. 결 론

이 논문은 라이프로그 데이터를 활용한 치매 조기 탐지를 위해 VAE 기반의 데이터 증강 기법과 딥러닝 모델을 제안하였다. 연구 결과, 제안된 VAE 기반 증강 기법이 소수 클래스인 Dem 클래스의 데이터를 효과적으로 보강하여 모델의 성능을 향상시킬 수 있음을 확인하였다.

이 논문의 주요 기여는 의료 분야에서 빈번히 발생하는 데이터 불균형 문제를 VAE 기반 데이터 증강 기법을 통해 효과적으로 해결한 점에 있다. 다양한 딥러닝 모델을 비교 분석한 결과 Wide & Deep Learning 모델에서 VAE 기반 증강 기법이 특히 효과적임을 입증하였다. 또한 고가의 의료 장비 없이 일상생활에서 수집된 라이프로그 데이터를 활용하여 치매를 조기에 진단할 수 있는 실용적인 대안을 제시함으로써 의료 접근성이 낮은 환경에서도 조기 진단과 맞춤형 중재의 가능성을 높일 수 있다는 의미가 있다.

이 논문은 인지 기능 저하의 단계를 더 정확하게 식별함으로써 치매 조기 진단의 신뢰성과 정확성을 높일 뿐만 아니라 데이터 불균형 문제 해결 방안을 적용하여 모델의 성능과 신뢰성을 동시에 향상시켰다. 마지막으로 라이프로그 데이터에 특화된 딥러닝

모델을 연구함으로써 기존 연구들과 차별화하였으며 라이프로그 데이터를 통한 치매 진단의 정확성과 실용성을 크게 향상시킬 것으로 기대된다.

향후 연구에서는 더 다양한 연령대와 활동 패턴을 포함하는 광범위한 라이프로그 데이터의 수집 및 분석이 필요하다. 또한 VAE와 다른 데이터 증강 기법을 결합하여 더욱 효과적인 데이터 불균형 해소 방안을 탐색할 필요가 있으며, 실제 임상 환경에서의 적용성을 높이기 위한 추가 연구가 요구된다.

References

- [1] World Health Organization (WHO), "Dementia: number of people affected to triple in next 30 years", <https://www.who.int/news/item/07-12-2017-dementia-number-of-people-affected-to-triple-in-next-30-years>. [accessed: Nov. 06, 2024]
- [2] G. Livingston, A. Sommerlad, and V. Orgeta, "Dementia prevention, intervention, and care: 2020 report of the Lancet Commission", *The Lancet*, Vol. 396, No. 10248, pp. 413-446, Aug. 2020. [https://doi.org/10.1016/S0140-6736\(20\)30367-6](https://doi.org/10.1016/S0140-6736(20)30367-6).
- [3] Alzheimer's Disease International (ADI), "From Plan to Impact V: World Health Organization (WHO) Global action plan on the public health response to dementia 2017-2025", <https://www.alzint.org>. [accessed: Nov. 06, 2024]
- [4] J. Rasmussen and H. Langerman, "Alzheimer's Disease: Why we need early diagnosis", *Degener Neurol Neuromuscul Dis.*, Vol. 9, pp. 123-130, Dec. 2019. <https://doi.org/10.2147/DNND.S228939>.
- [5] T. Shintani, T. Ohmori, and H. Fujita, "Comparison method of long-term daily life considering the manner of spending a day", *Proc. 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pp. 347-354, Nov. 2019. <https://doi.org/10.5220/0008343303470354>.
- [6] H. Xie, Y. Wang, S. Tao, S. Huang, C. Zhang, and Z. Lv, "Wearable Sensor-Based Daily Life

- Walking Assessment of Gait for Distinguishing Individuals With Amnesic Mild Cognitive Impairment", *Front. Aging Neurosci.*, Vol. 11, No. 285, pp. 1-13, Dec. 2019. <https://doi.org/10.3389/fnagi.2019.00285>
- [7] K. Kim, J. Jang, and H. Park, "Detecting Abnormal Behaviors in Dementia Patients Using Lifelog Data: A Machine Learning Approach", *Information*, Vol. 14, No. 8, pp. 433-445, Aug. 2023. <https://doi.org/10.3390/info14080433>.
- [8] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance", *Journal of Big Data*, Vol. 6, No. 1, pp. 27-46, Mar. 2019. <https://doi.org/10.1186/s40537-019-0192-5>.
- [9] H. He and E. A. Garcia, "Learning from imbalanced data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21, No. 9, pp. 1263-1284, Sep. 2009. <https://doi.org/10.1109/TKDE.2008.239>.
- [10] A. J. Mohammed, M. M. Hassan, and D. H. Kadir, "Improving Classification Performance for a Novel Imbalanced Medical Dataset using SMOTE Method", *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No. 3, pp. 3161-3172, Jun. 2020. <https://doi.org/10.30534/ijatcse/2020/104932020>.
- [11] M. Bach, A. Werner, and J. Żywiec, "The study of under- and over-sampling methods' utility in analysis of highly imbalanced data on osteoporosis", *Information Sciences*, Vol. 384, pp. 174-190, Nov. 2017. <https://doi.org/10.1016/j.ins.2016.09.038>.
- [12] B. Dai and D. Wipf, "Diagnosing and Enhancing VAE Models", *Proc. International Conference on Learning Representations*, New Orleans, USA, pp. 1-13, May 2019. <https://doi.org/10.48550/arXiv.1903.05789>.
- [13] T. Wongvorachan, S. He, and O. Bulut, "A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining", *Information*, Vol. 14, No. 1, pp. 54-68, Jan. 2023. <https://doi.org/10.3390/info14010054>.
- [14] J. Lee and K. Y. Lee, "An Anomalous Sequence Detection Method Based on An Extended LSTM Autoencoder", *The Journal of Society for e-Business Studies*, Vol. 26, No. 1, pp. 127-140, Mar. 2021. <https://doi.org/10.7838/jsebs.2021.26.1.127>.
- [15] R. A. Hamad, M. Kimura, and J. Lundström, "Balanced-MixUp for highly imbalanced medical image classification", *SN Computer Science*, Vol. 1, No. 4, pp. 204-218, Jul. 2020. https://doi.org/10.1007/978-3-030-87240-3_31.
- [16] K. M. Sunderland, D. Beaton, and J. Fraser, "The utility of multivariate outlier detection techniques for data quality evaluation in large studies: an application within the ONDRI project", *BMC Medical Research Methodology*, Vol. 19, No. 1, pp. 102-115, Apr. 2019. <https://doi.org/10.1186/s12874-019-0737-5>.
- [17] H. Choi, C. Yoon, and S. B. Lee, "Cognitive Impairment Prediction Model Using AutoML and Lifelog", *Journal of The Korea Society of Computer and Information*, Vol. 28, No. 11, pp. 53-63, Nov. 2023. <https://doi.org/10.9708/jksci.2023.28.11.053>.
- [18] M. Lee, J. Lee, and H. Lee, "Cognitive Dysfunction Prediction Model with Lifelog Dataset based on Random Forest and SHAP", *Journal of The Korea Society of Computer and Information*, Vol. 22, No. 1, pp. 1-8, Jan. 2024. <https://doi.org/10.14801/jkiit.2024.22.1.1>.
- [19] J.-Y. Lee and S. Y. Lee, "Development of an AI-Based Predictive Algorithm for Early Diagnosis of High-Risk Dementia Groups among the Elderly: Utilizing Health Lifelog Data", *Healthcare*, Vol. 12, No. 18, pp. 1872-1885, Aug. 2024. <https://doi.org/10.3390/healthcare12181872>.
- [20] M. Munir, S. A. Siddiqui, A. Dengel, and S.

- Ahmed, "DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series", Proc. IEEE Access, pp. 1991-2005, Oct. 2019. <https://doi.org/10.1109/ACCESS.2018.2886457>.
- [21] S. Chauhan and L. Vig, "Anomaly detection in ECG time signals via deep long short-term memory networks", Proc. IEEE International Conference on Data Science and Advanced Analytics, Paris, France, pp. 1-7, Oct. 2015. <https://doi.org/10.1109/DSAA.2015.7344872>
- [22] J. Kim and J. Lee, "Dementia Prediction using LSTM Model based on Life-Log", Journal of The Korea Society of Computer and Information, Vol. 22, No. 9, pp. 123-132, Sep. 2024. <http://dx.doi.org/10.14801/jkiit.2024.22.9.123>.
- [23] AI Hub, <https://www.aihub.or.kr/>. [accessed: Nov. 06, 2024]
- [24] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes", arXiv preprint arXiv:1312.6114, pp. 1-14, Dec. 2013. <https://doi.org/10.48550/arXiv.1312.6114>.
- [25] G. Roeder, Y. Wu, and D. P. Kingma, "Sticking the Landing: Simple, Lower-Variance Gradient Estimators for Variational Inference", arXiv preprint arXiv:1703.09194, pp. 1-10, Mar. 2017. <https://doi.org/10.48550/arXiv.1703.09194>.
- [26] A. Singh and T. Ogunfunmi, "An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications", Entropy, Vol. 24, No. 1, pp. 55-72, Jan. 2022. <https://doi.org/10.3390/e24010055>.
- [27] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation Forest", Proc. Eighth IEEE International Conference on Data Mining, Pisa, Italy, pp. 413-422, Dec. 2008. <https://doi.org/10.1109/ICDM.2008.17>.
- [28] F. Aldi, F. Hadi, and N. A. Rahmi, "StandardScaler's Potential in Enhancing Breast Cancer Accuracy Using Machine Learning", JAETS, Vol. 5, No. 1, pp. 401-413, Dec. 2023. <https://doi.org/10.37385/jaets.v5i1.3080>.
- [29] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, pp. 785-794, Aug. 2016. <https://doi.org/10.1145/2939672.2939785>.
- [30] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors", Nature, Vol. 323, No. 6088, pp. 533-536, Oct. 1986. <https://doi.org/10.1038/323533a0>.
- [31] S. O. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning", Proc. International Conference on Machine Learning, Long Beach, California, USA, pp. 1-15, Jul. 2019. <https://doi.org/10.48550/arXiv.1908.07442>.
- [32] H. T. Cheng, L. Koc, and J. Harmsen, "Wide & deep learning for recommender systems", Proc. 1st Workshop on Deep Learning for Recommender Systems, Boston MA USA, pp. 7-10, Sep. 2016. <https://doi.org/10.1145/2988450.2988454>.

저자소개

이 민 지 (Minji Lee)



2023년 3월 ~ 현재 :
국립군산대학교 소프트웨어학과
학사과정
관심분야 : 인공지능, 헬스케어,
데이터 분석

이 석 훈 (Sukhoon Lee)



2009년 2월 : 고려대학교
전자및정보공학부(학사)
2011년 2월 : 고려대학교
컴퓨터·전파통신공학과(공학석사)
2016년 2월 : 고려대학교
컴퓨터·전파통신공학과(공학박사)
2016년 3월 ~ 2017년 3월 :

아주대학교 의료정보학과 연구강사

2017년 4월 ~ 현재 : 국립군산대학교 소프트웨어학과
부교수

관심 분야 : 사물인터넷, 메타데이터 레지스트리, 데이터
품질, 연합 학습