

DNN 기반 우울증 위험 예측 성능 향상을 위한 하이퍼파라미터 최적화 방법 연구

박상훈*¹, 하유진*², 김건우**

Hyperparameter Optimization Strategies for Improving DNN based Depression Risk Prediction

Sang Hoon Park*¹, Yu Jin Ha*², and Gun-Woo Kim**

본 논문은 2025년도 산업통상자원부 및 한국산업기술기획평가원(KEIT) 연구비 지원에 의한 연구임
(RS-2025-02633048)

요약

본 연구는 웨어러블 기기로 수집한 생체정보와 PHQ 9 설문 데이터를 활용한 DNN 기반 우울증 위험 예측 모델의 하이퍼파라미터를 최적화해 예측 성능을 높이는 방안을 제안한다. 데이터 전처리를 통해 품질을 향상하고 클래스 불균형을 보완하였으며, 전통적인 머신러닝 모델들과 DNN의 성능을 비교했다. 실험 결과, DNN 모델이 F1-score 0.8953, ROC-AUC 0.77로 가장 우수한 성능을 보였으며, 특히 MinMaxScaler, Nadam 옵티마이저, 6개 은닉층 구조의 조합이 최적의 설정으로 확인되었다. 이는 웨어러블 생체정보에 내재한 비선형적이고 계층적인 패턴을 DNN 모델이 비교 모델보다 효과적으로 학습함을 시사한다. 본 연구는 웨어러블 데이터를 활용한 우울증 조기 선별 및 지속 모니터링 시스템 구축의 실효성을 실증적으로 제시한다.

Abstract

This study proposes a method to improve the predictive performance of a DNN based depression risk model by optimizing its hyperparameters using biometric data from wearable devices and PHQ 9 survey responses. After rigorous preprocessing, including quality enhancement and class-imbalance correction with SMOTE, we compared several traditional machine-learning algorithms with a deep neural network (DNN). The DNN, optimized with MinMax scaling, six hidden layers, and the Nadam optimizer, achieved the best performance, recording an F1-score of 0.8953 and demonstrating superior ability to capture the non-linear and hierarchical patterns intrinsic to wearable biosignals. These findings provide empirical support for developing wearable-based systems for early depression screening and continuous monitoring.

Keywords

depression prediction, wearable devices, deep neural network, machine learning

* 경상국립대학교 AI융합공학부

- ORCID¹: <https://orcid.org/0009-0002-0430-9745>

- ORCID²: <https://orcid.org/0009-0004-8963-1913>

** 경상국립대학교 컴퓨터공학과 교수(교신저자)

- ORCID: <https://orcid.org/0000-0001-5643-4797>

· Received: May 14, 2025, Revised: Aug. 09, 2025, Accepted: Aug. 12, 2025

· Corresponding Author: Gun-Woo Kim

Dept. of Computer Science and Engineering, College of IT Engineering,
Gyeongsang National University, JinJu, Korea

Tel.: +82-55-772-3323, Email: gunwoo.kim@gnu.ac.kr

I. 서 론

현대사회에서 우울증 및 불안장애와 같은 정신건강 문제는 전 세계 인구의 10% 이상이 경험한다[1]. 이러한 정신건강 문제는 개인의 일상생활, 직업적 성취, 사회적 관계에 심각한 부정적 영향을 미친다. 세계보건기구(WHO)는 2030년까지 우울증이 전 세계적으로 가장 큰 질병 부담을 초래할 것으로 전망했다[2]. 이에 따라, 우울증의 조기 진단 및 개입의 필요성이 더 강조되고 있다[3]. 특히 우울증은 정신건강뿐만 아니라 신체 건강에도 영향을 미치며, 심혈관 및 대사 질환과의 연관성이 높아 적절한 관리가 되지 않으면 자살 위험이 증가할 수 있다[4].

최근 1인 가구 증가와 코로나19로 인한 사회적 거리 두기, 경제적 불안정은 사회의 고립감을 심화시키고 우울증 발병률과 진단율이 증가하고 있다 [5]. 이러한 상황에서 정신건강 관리 시스템의 필요성이 커지면서, 우울증을 조기에 예측할 수 있는 기술과 모델 개발에 대한 사회적 요구도 높아지고 있다. 우울증과 같은 정신질환은 주로 개인의 주관적 판단과 증상 보고에 의존하기 때문에 조기 발견이 어렵고, 정량적·지속적 모니터링도 쉽지 않다. 이를 보완하기 위해 최근 연구에서는 스마트워치 등 웨어러블 기기를 활용해 일상생활에서 지속적으로 생체신호를 수집하는 방법을 시도하고 있다. 스마트워치를 통해 심박수, 수면 패턴, 신체 활동 등의 데이터를 실시간으로 수집하고, 이들 생체정보와 우울증 간의 상관관계를 분석해 예측 모델을 개발하는 연구가 진행되고 있다[6].

스마트워치와 같은 웨어러블 기기의 사용률은 꾸준히 증가하고 있으며, 2024년에는 전체 사용률이 인구의 33%가 스마트워치를 사용할 것으로 예상된다[7]. 이는 스마트워치가 단순 피트니스 추적 시스템을 넘어 심박수, 혈압, 수면 패턴 등 다양한 생체 데이터 측정 기능을 제공함에 따라 건강 모니터링 및 의료 지원 도구로 활용도가 높아지고 있다는 점을 반영한다. 스마트워치에서 측정할 수 있는 생체 정보 중 심박수 변동, 수면 질 저하, 신체 활동 감소 등은 우울증과 밀접한 관련이 있는 주요 지표다. 따라서 스마트워치를 통해 이러한 지표들을 실시간으로 감지하고 분석하면 개인 맞춤형 우울증 관리

및 예방 전략을 효과적으로 수립할 수 있다.

본 연구는 웨어러블 기기로 수집한 생체정보와 PHQ-9 설문 데이터를 활용해 우울증 위험을 예측하는 DNN(Deep Neural Network) 모델의 하이퍼파라미터를 최적화하여 예측 성능을 높이는 방법을 제안한다. 기존 우울증 진단은 정적 설문 기반 방식에 의존해 실시간이나 개인 맞춤형 관리에 한계가 있었다. 이에 본 연구는 웨어러블 기기로 수집한 동적 생체 데이터를 활용해 PHQ-9 기반 우울증 위험 점수를 정량적으로 예측하고, 향후 조기 선별 및 실시간 모니터링 시스템 개발의 기반을 마련하는 것을 목표로 한다.

이를 위해 질병관리청에서 제공한 국민건강영양조사 데이터를 활용해 스마트워치에서 수집 가능한 건강 데이터를 시뮬레이션했다. 전체 실험 과정은 그림 1과 같으며, 데이터 전처리 단계에서 불균형한 우울증 데이터를 SMOTE 기법으로 증강하여 클래스 비율을 맞추고, 특징 생성 및 이상치 제거 과정을 거쳤다. 특히 DNN 모델의 성능을 최적화하기 위해 은닉층 개수와 유닛 수, 옵티마이저 종류 등을 조정하고 모델 구조를 체계적으로 개선했다. 또한 과적합을 방지하고 일반화 성능을 높이기 위해 배치 정규화, 드롭아웃, L2 정규화 등의 규제 기법을 적용했으며, 혼합 정밀도 학습, 그래디언트 클리핑, 학습률 스케줄링, 조기 종료 기법 등 고급 학습 전략을 종합적으로 활용했다. 이렇게 개선된 DNN 모델의 성능을 로지스틱 회귀, 서포트 벡터 머신, 랜덤 포레스트, XGBoost, LightGBM 등 전통적 머신러닝 기법과 정형 데이터를 이용하는 딥러닝 모델인 1D-CNN, TabNet 등을 비교하여 정확도, 정밀도, 재현율, F1 스코어 등의 지표로 평가했다.

II. 관련 연구

2.1 우울증 위험 척도

기존 연구에서는 우울증 초기 진단을 위해 자기 보고식 선별검사를 널리 활용해 왔다. 그중 PHQ-9은 9문항으로 구성되어 있어 피검사자가 3분 이내에 작성할 수 있을 정도로 간편하며, 임상 및 연구 환경에서 PHQ-9은 DSM(Diagnostic and Statistical

Manual of Mental Disorders) 진단 기준에 근거해 우울증 유무와 심각도를 객관적으로 평가할 수 있어 반복적 모니터링과 디지털 헬스케어 시스템 연계가 쉬워 표준 도구로 자리 잡고 있다. 본 연구에서는 PHQ-9 점수를 종속 변수(mh_PHQ_S)로 선정했다 [8][9].

2.2 웨어러블 디바이스를 통한 우울증 예측

기존 연구에서는 스마트워치 등 웨어러블 기기를 활용해 심박수, 수면 패턴, 활동량 자료를 수집하고 이를 바탕으로 우울증 위험을 예측하는 방법을 시도했다[10][11]. A. Shcherbina et al.은 웨어러블 기기의 생체신호 측정 정확도에 관하여 기술하였고[12], 2022년 FDA에서 규제 승인 기준을 마련하면서 웨어러블 기기의 의료적 활용에 대한 신뢰도를 높였다[13]. T. Shin et al.은 갤럭시 워치5를 활용하여 ECG, PPG Green, PPG IR 자료를 수집하였고, 그 후 STL·IQR·Wavelet Transform 기법을 사용하여 노이즈를 제거하고 특징을 추출했다[14][15]. Y. Tazawa et al.와 T. Mullick et al.의 연구에서도 웨어러블 및 모바일 센서를 활용하여 우울증 예측에 관한 다양한 머신러닝 접근법을 시도했다[10][16].

2.3 인공지능 기반 우울증 위험 척도 예측

기존 연구에서는 로지스틱 회귀, KNN(K-Nearest Neighborhood), 나이브 베이즈, 랜덤 포레스트, SVM(Support Vector Machine) 등의 머신러닝 알고리즘을 적용하여 생체신호와 PHQ-9 등 우울증 선별 검사 결과 간의 상관관계를 분석했다. 특히, PCA를 도입하여 고차원 데이터를 저차원 데이터로 축소하여 계산 효율성을 높이고 과적합 문제를 완화하는 방법을 적용하고 정확도, 정밀도, 재현율, F1-스코어 등의 성능 지표를 통해 비교 평가했다. 본 연구에서는 이와 유사한 접근법을 기반으로 DNN을 포함한 머신러닝 모델과 정형 데이터를 이용하는 딥러닝 모델들을 비교하여, 실제 스마트워치 기반 데이터에 최적화된 예측 모델을 구축했다. 또한, 제안한 DNN 모델과 AutoML 기반 모델의 성능을 비교하여 수동

하이퍼파라미터 튜닝의 효과를 검증했다[20].

III. 우울증 위험 척도 예측 모델 구축

3.1 데이터 전처리

본 연구에서는 우울증 위험 예측 모델 학습 전 모델 성능 향상을 위해 다음과 같은 전처리 과정을 적용했다. 우울증 위험 척도인 PHQ-9이 15세 이상의 청소년 및 성인 집단에서 신뢰도와 타당성이 검증되었다[16]. 이를 참고하여 15세 미만의 데이터를 제거했다. 종속 변수 mh_PHQ_S에 결측치가 있는 경우 해당 행을 삭제했으며, 나머지 연속형 변수의 결측값은 평균 대체법으로 처리했다. 수치형 변수의 이상치는 IQR(Interquartile Range)을 이용해 IQR의 1.5배 범위를 벗어나는 값을 제거했다. 문자열 형태의 범주형 변수는 레이블 인코딩을 적용해 수치형으로 변환을 진행했다. 종속 변수 mh_PHQ_S는 다중 분류를 위해 ‘우울증 비해당군’, ‘증상 의심군’, ‘고위험군’ 세 범주로 재구성하고, 각각 0, 1, 2의 정수로 인코딩하여 모델 학습에 사용했다. 연속형 변수는 단위와 범위가 서로 다르므로 StandardScaler와 MinMaxScaler를 이용해 정규화했다.

연속형 변수의 경우, 모델이 비선형 관계를 학습할 수 있도록 2차 다항식 항을 생성하여 데이터 증강을 수행했다. 이후 특징 선택을 통해 표 1과 같이 변수의 F-스코어를 기준으로 중요도가 높은 상위 11개의 변수를 선택했다[21].

표 1. F-스코어 기반 선택된 상위 11개 변수
Table 1. F-score-based selected variables

	F-score	p-value
sex	21.36	0.00
BP17_dg	2.45	0.09
BP16_1	22.58	0.00
BP16_2	10.88	0.00
BE3_31	2.72	0.07
HE_dbp	1.60	0.20
HE_dbp	1.53	0.22
sexBP17_dg	1.44	0.24
BE5_1HE_dbp	1.53	0.22
HE_rPLSHE_dbp	1.53	0.22
HE_dbp^2	1.50	0.22

Neural Network Visualization with Dropout

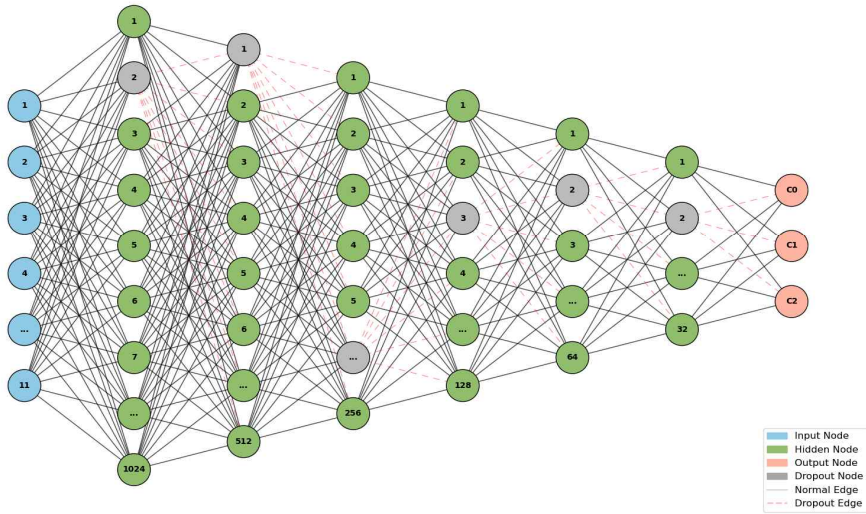


그림 2. 모델 구조
Fig. 2. Model architecture

원본 데이터셋은 우울증 고위험군 표본 현저히 적어 심각한 클래스 불균형을 보였다. 이에 SMOTE (Synthetic Minority Over-sampling Technique)를 적용하여 소수 클래스 데이터를 증강한 결과, 표 2와 같이 학습 데이터의 클래스별 표본 수를 1,817로 균등화해 총 5,613건으로 증강하였다. 표 3에 따르면 SMOTE 적용 후 DNN의 정확도, 정밀도, 재현율이 모두 개선되었으며, 이는 고위험군의 결정 경계가 분명해져 모델의 전반적인 일반화 성능이 향상된 것으로 해석된다.

표 2. SMOTE 적용에 따른 학습 데이터 수 변화
Table 2. Distribution changes after applying SMOTE

PHQ-9	Before SMOTE	After SMOTE
0~4(0)	1817	1871
5~9(1)	302	1871
10~27(2)	101	1871
Total	2220	5613

표 3. SMOTE 적용에 따른 DNN 성능 비교
Table 3. Performance comparison after applying SMOTE

	Before SMOTE	After SMOTE
Accuracy	0.833	0.893
Precision	0.739	0.879
Recall	0.834	0.892
F1-score	0.779	0.883

3.2 우울증 위험 척도 예측 모델

본 연구에서는 우울증 예측을 위해 전통적 머신러닝 기법들과 정형 데이터를 사용하는 딥러닝 모델들을 베이스로 선정한 DNN 모델과 성능을 비교하여 효율성을 검증했다. 비교에 사용한 머신러닝 모델로는 로지스틱 회귀, 랜덤 포레스트, XGBoost, SVM, AdaBoost, LightGBM을 선택했으며, 정형 데이터를 이용하는 딥러닝 모델들의 경우 1D-CNN, TabNet을 선정하였다.[22] 각 모델은 최적화를 수행한 후, 성능 비교를 진행했다.

DNN 모델은 다층 퍼셉트론의 형태로 설계하였으며 모델의 최종 구조는 다음 그림 2와 같다. 각 은닉층에서는 ReLU 활성화 함수를 사용하고, 배치 정규화를 적용해 학습 안정성을 확보했으며, Dropout을 통해 과적합을 방지했다. 또한 다양한 스케일링 기법과 최적화 알고리즘을 비교하여 최적의 조합을 선택했다[23].

모델의 일반화 성능을 높이기 위해 학습률 스케줄링과 조기 종료기법을 적용했다. 최종 모델 평가는 정확도, 정밀도, F1 스코어, ROC-AUC를 기준으로 수행했으며, 이를 통해 최적 모델을 선정한 뒤 추가적인 하이퍼파라미터 튜닝을 통해 우울증 위험 척도 예측 DNN 모델을 완성했다.

IV. 실험 및 성능 평가

4.1 실험 환경

본 연구의 실험 환경은 WSL2 기반 우분투 24.04.1 LTS 운영체제를 기반으로 하였으며, CUDA 12.6과 cuDNN 9.6.0을 통해 GPU를 지원했다. 구현된 모델은 파이썬 3.12와 텐서플로 2.18.0을 사용했다. 하드웨어 구성은 인텔 코어 i5-13600KF CPU, 엔비디아 지포스 RTX 4070 Ti GPU, 64 GB DDR5 RAM으로 모델 학습 및 실험을 진행하였다.

4.2 데이터

본 연구는 국민건강영양조사(KNHANES, Korea National Health and Nutrition Examination Survey)의 9기 1차 연도 자료를 활용하였다. 국민건강영양조사는 질병관리청이 수행하는 국가 단위 공공 데이터셋으로 대한민국 국민의 건강 상태, 영양 섭취, 생활 습관 등을 종합적으로 평가하기 위해 설계되었다. 제9기 1차 연도 자료는 2022년 한 해 동안 수집된 데이터로, 전국적으로 표본 가구를 대상으로 건강 설문, 신체 계측, 생화학 검사, 식이 조사 등을 실시해 수집되었다.

표 4. 실험 데이터셋

Table 4. Experimental data set

Variables	Name	Description
Target	mh_PHQ_S	PHQ-9 Score
Input	sex	SEX
Input	age	AGE
Input	HE_sbp	Systolic blood pressure
Input	HE_dbp	Diastolic blood pressure
Input	HE_rPLS	Pulse regularity
Input	HE_PLS_30	30-Second pulse count
Input	BP17_dg	Obstructive sleep apnea
Input	BP16_1	Average weekday sleep duration
Input	BP16_2	Average weekend sleep duration
Input	BE3_31	Days of walking
Input	BE5_1	Days of strength exercise

데이터셋은 총 6,265명의 응답자와 623개 변수로 구성되어 있다. 본 연구에서는 표 4에서 제시한 바와 같이 스마트워치 기반 측정이 가능한 변수를 선별하였다. 선별한 독립 변수는 총 11개이며 성별, 나이, 수축기 혈압, 이완기 혈압, 맥박 규칙성, 30초 간 맥박 수, 폐쇄성 수면 무호흡증 여부, 주중 평균 수면시간, 주말 평균 수면시간, 주간 걷기 일수, 주간 근력운동 일수이다. 종속 변수는 우울 정도를 나타내는 PHQ-9 점수이다.

4.3 하이퍼파라미터 최적화

본 연구에서는 DNN 모델의 성능 향상을 위해 은닉층 개수, 스케일러, 옵티마이저 순으로 하이퍼파라미터를 튜닝을 수행하였다. 먼저 은닉층 수에 따른 모델 탐색을 위해, 표 5와 같이 은닉층 및 뉴런 개수를 달리하고 스케일러는 MinMaxScaler, 옵티마이저는 RMSprop으로 고정하여 비교 실험을 진행하였다. 표 6의 결과, 은닉층을 6개로 설정하였을 때 정확도가 가장 높았다. 이는 은닉층 수가 6개로 설정하였을 때, 변수 간 비선형 상호작용을 충분히 학습하면서도 파라미터 수 증가를 억제해 일반화 오류를 최소화한 것으로 해석된다. 반면 은닉층의 수를 7개 이상으로 증가시키면 과적합 위험과 기울기 소실 문제가 발생하여 성능이 감소하는 경향이 관찰되었다.

표 5. 은닉층 개수 별 뉴런 수

Table 5. Neuron setting by hidden layer depth

Depth	Layers
4	[1024, 512, 256, 128]
5	[1024, 512, 256, 128, 64]
6	[1024, 512, 256, 128, 64, 32]
7	[1024, 512, 256, 128, 64, 32, 16]
8	[1024, 512, 256, 128, 64, 32, 16, 8]

표 6. 은닉층 수에 따른 성능 비교 실험

Table 6. Comparison of hidden layer counts

layer	scaler	optimizer	Accuracy
4	MinMax	RMSprop	0.8011
5	MinMax	RMSprop	0.8051
6	MinMax	RMSprop	0.8239
7	MinMax	RMSprop	0.8065
8	MinMax	RMSprop	0.7742

표 7의 스케일러 비교 결과 MinMaxScaler는 과반수의 옵티마이저 조합에서 StandardScaler보다 높은 정확도를 기록하였었다. 이는 입력값이 0-1의 구간으로 압축하여 ReLU 활성화의 분포가 균질하게 유지되어 역전과 과정에서 발생할 수 있는 기울기 소실과 폭주가 완화된 결과로 해석된다. 한편, AdamW의 경우 StandardScaler가 우수한 성능을 보여주었다. 이는 L2 정규화가 포함된 AdamW가 좁은 입력 분포에서 과도한 패널티를 부과해 학습 범위가 제한된 결과로 추정된다.

표 7. 스케일러 성능 비교 실험

Table 7. Performance comparison of scaler

optimizer	layer	scaler	accuracy
RMSprop	6	Standard	0.846
		MinMax	0.8596
Adam	6	Standard	0.8374
		MinMax	0.866
AdamW	6	Standard	0.8327
		MinMax	0.8065
Nadam	6	Standard	0.8566
		MinMax	0.8755

표 8의 옵티마이저 성능 비교 결과 Nadam이 가장 높은 정확도를 기록하였다. Nadam은 Adam의 적용 학습률에 Nesterov 모멘텀을 결합해 학습 초기 수렴 속도와 최종 손실을 동시에 개선하였다[24]. 앞서 은닉층 및 스케일러 실험 결과와 종합하여, 본 연구는 은닉층 6개의 구조에 MinMaxScaler와 Nadam의 구성을 채택했다.

표 8. 옵티마이저 성능 비교 실험

Table 8. Performance comparison of optimizer

optimizer	layer	scaler	accuracy
RMSprop	6	MinMax	0.8596
Adam	6	MinMax	0.866
AdamW	6	MinMax	0.8696
Nadam	6	MinMax	0.8938

최종 DNN 모델은 전처리 과정을 거친 학습 데이터에 MinMaxScaler를 적용하여 정규화 후, 은닉층 6개로 구성된 다층 퍼셉트론 구조를 설계하였다. 각 은닉층에는 드롭아웃 0.2를 적용하여 과적합을 억제하고, 옵티마이저로는 Nadam을 사용했다[25]. 학습률은 1e-3, 배치 크기는 32, 최대 에폭은 50으로 설

정했다. ReduceLRonPlateau 콜백을 적용하여 검증 손실이 일정 에폭 이상 개선되지 않는다면 학습률을 0.9배로 감축해 학습 안정성을 확보하였다. EarlyStopping 적용하여 검증 손실 향상이 정체되면 학습을 조기에 종료해 과적합을 방지했다.

4.4 머신러닝 모델 성능 비교 실험

표 9. 머신러닝 모델들 하이퍼파라미터 설정

Table 9. Hyperparameter settings for ML models

Model	Hyperparameter	Value
LogisticRegression	C	0.001
	penalty	l2
	solver	lbfgs
	max_iter	1000
RandomForest	max_depth	10
	max_features	sqrt
	min_samples_leaf	1
	min_samples_split	2
	min_weight_fraction_leaf	0.0
	n_estimators	50
SVM	C	10
	kernel	rbf
	gamma	auto
XGBoost	objective	multi:softprob
	gamma	None
	learning_rate	0.1
	max_depth	3
	min_child_weight	None
	n_estimators	50
	use_label_encoder	False
	eval_metric	logloss
AdaBoost	learning_rate	0.5
	n_estimators	300
LightGBM	boosting_type	gbdt
	colsample_bytree	1.0
	learning_rate	0.1
	importance_type	split
	max_depth	10
	min_child_samples	20
	min_child_weight	0.001
	min_split_gain	0.0
	n_estimators	50
	num_leaves	31
	reg_alpha	0.0
	reg_lambda	0.0
	subsample	1.0
	subsample_for_bin	200000
	subsample_freq	0

본 연구에서는 DNN 모델과의 성능 비교를 위해 로지스틱 회귀, 서포트 벡터 머신, 랜덤 포레스트, XGBoost, LightGBM 등 전통적인 머신러닝 기법을 적용하였다. 각 모델의 주요 하이퍼파라미터는 그리드 서치로 최적화하여 표 9에 제시했다.

표 10의 결과를 따르면, 로지스틱 회귀는 가장 낮은 성능을 기록하였다. 이는 모델이 선형 관계를 전제해 수치형 생체신호 간 비선형 복합 상호작용을 포착하지 못하였고, 과적합을 방지하기 위해 규제 계수 C를 0.001로 제한한 결과 결정 경계가 단 순화되어 성능이 저하된 것으로 해석된다.

랜덤 포레스트, XGBoost, LightGBM과 같은 트리 기반 앙상블 모델은 정확도를 비롯한 주요 지표에서 평균 0.84 내외로 로지스틱 회귀와 같은 선형 모델 대비 전반적으로 준수한 성능을 보여주었다. 랜덤 포레스트는 계산 복잡도를 억제하면서 배경 효과와 트리 다양성을 확보하여 안정적인 성능을 확보한 것으로 보인다. XGBoost는 학습률 0.1, 트리 수 50 설정으로 모델 복잡도를 제어하고, 다중 클래스 확률을 안정적으로 학습하였다. LightGBM은 XGBoost와 동일한 학습률과 트리 수를 적용하고, num_leaves는 31, max_depth를 10으로 설정하여 세밀한 분할을 허용하였다. 정규화 계수를 0으로 두어 모델 복잡도를 완화하였다. AdaBoost의 성능은 다른 트리 기반 앙상블보다 낮은 성능을 보였다. 이는 얇은 결정 트리를 0.5의 높은 학습률로 300회 반복 부스팅을 진행하였음에도 클래스 오차를 충분히 보정하지 못한 것으로 해석할 수 있다.

SVM의 경우 0.6864인 정밀도와 비교하여 0.6447의 떨어지는 재현율은 고위험군 탐지 민감도가 상대적으로 부족하였다. 이는 넓은 커널 반경과 높은 패널티가 평탄한 경계를 형성해, 복잡한 고위험군 분포를 포착하지 못한 결과로 해석된다. 이러한 결과는 본 연구의 데이터셋에 비선형적이고 계층적인 패턴이 내재되어 있어 트리 앙상블과 같은 비선형 모델이 선형모델보다 일관되게 우수한 성능을 보인 것으로 해석된다.

4.5 정형 데이터 딥러닝 모델 비교 실험

본 연구에서는 DNN 모델과의 성능 비교를 위해 1D CNN, TabNet 등 정형 데이터를 처리 가능한 딥러닝 기법들을 적용하였다. 그리드 서치를 통해 최적화한 하이퍼파라미터는 표 11에 제시하였다. 표 10의 결과를 따르면 1D-CNN은 정확도 0.6791, F1-스코어 0.6403으로 정확도 0.8391, F1-스코어 0.7838인 TabNet에 대비 성능이 낮았다. 이는 입력이 시계열 계열 원시 신호가 아니라 집계 통계치이므로, 1D-CNN의 커널 기반 지역 패턴 학습 효과가 제한된 것이 원인으로 해석된다.

또한, TabNet은 머신러닝 기반 트리 앙상블 계열 모델들과 유사한 수준의 성능을 달성하였지만, 이는 컬럼 마스킹 효과가 표본 수에 비해 상대적으로 작아, b_step을 3 이상으로 늘려도 성능 이득이 제한적이었던 것으로 해석된다.

표 10. 모델 성능 비교

Table 10. Performance comparison of baseline models

	Accuracy	Precision	Recall	F1 Score
LogisticRegression	0.6693	0.4480	0.6693	0.5367
RandomForest	0.8499	0.8471	0.8021	0.8471
XGBoost	0.8480	0.7653	0.8480	0.7982
SVM	0.7631	0.6839	0.6446	0.6599
AdaBoost	0.8194	0.6833	0.6238	0.6393
LightGBM	0.8391	0.7610	0.8391	0.7931
AutoML	0.8109	0.7104	0.8109	0.7450
1D CNN	0.6791	0.6763	0.6488	0.6403
TabNet	0.8391	0.7460	0.8391	0.7838
DNN	0.8938	0.8805	0.8938	0.8953

표 11. 딥러닝 모델들 하이퍼파라미터 설정
Table 11. Hyperparameter settings for DL models

Model	Hyperparameter	Value
1D-CNN	batch_size	64
	dropout_rate	0.2
	epochs	30
	filters	64
	kernel_size	3
TabNet	n_d	31
	n_a	59
	b_step	3
	gamma	1.2809
	lambda_sparse	0.0003

이러한 결과는 집계 통계치 중심의 입력 특성상, 1D-CNN의 지역 패턴과 TabNet의 컬럼 마스킹의 구조적 이점이 발휘되지 못한 것으로 보인다. 또한, 정형 데이터 전용 딥러닝 모델은 대규모 표본 또는 풍부한 범주 및 시계열 변수가 확보될 때 성능이 향상될 것으로 보인다. 이는 본 연구의 데이터셋과 같이 규모가 제한적이고 고차 비선형성이 강한 경우는 DNN이 정형 데이터 전용 딥러닝 모델보다 성능적 우위를 보일 수 있음을 의미한다.

4.6 AutoML 성능 비교 실험

본 연구에서는 자동화된 모델 탐색 기법과 수동 설계 모델 간 성능 차이 검증을 위해 Auto-sklearn 2.0을 활용한 AutoML 실험을 수행하였다. 제안한 DNN 모델과 동일한 데이터 전처리 절차를 적용하고 최적화 목표 함수를 `f1_weighted`로 설정하였다. Auto-sklearn이 도출한 최종 앙상블은 AdaBoost 54%, 가우시안 NB 24%, 랜덤 포레스트 10%를 주축으로 소규모 모델을 가장 평균한 구조였다. 표 9에 나타듯 수동으로 하이퍼파라미터를 조정된 DNN 모델이 정확도, 정밀도, 재현율, F1-스코어 전 지표에서 AutoML이 자동 탐색한 최적 모델을 웃돌았다. 이는 AutoML이 얇은 트리를 기반 AdaBoost와 조건부 독립 가정을 하는 가우시안 NB 위주로 구성되어 고차 비선형 패턴 학습에 한계를 보였고, 제한된 탐색 자원으로 도메인 지식을 반영한 세밀한 하이퍼파라미터 조정이 부족했기 때문으로 해석된다. 따라서

도메인 특화 지식에 기반한 직접적인 하이퍼파라미터 튜닝과 네트워크 구조 설계가 웨어러블 생체정보 기반 우울증 예측 과제에서 자동화 탐색 기법보다 우수한 성능을 제공함을 확인하였다.

4.7 우울증 위험 예측 모델 성능평가

본 연구에서는 전통적인 머신러닝 기법들과 선형 데이터 전용 딥러닝 모델들을 DNN 간의 성능을 비교 실험을 진행하였다. 표 10과 같이 정확도, 정밀도, 재현율, F1-스코어에서 모두 DNN이 최고 성능을 기록하였다. 트리 기반 앙상블 계열 모델들은 평균 성능 0.84 내외로 선형모델보다 우수했으나, 비선형 고차 상호작용 학습 능력에서는 DNN에 미치지 못했다. DNN 분류 세부 양상을 확인하기 위해 그림 3과 같이 혼돈행렬을 시각화하여 분석하였다. 분석한 결과 클래스 0(우울증 비해당군)과 클래스 2(우울증 고위험군)에서 높은 정확도로 분류됐지만, 클래스 1(우울증 의심군)에 속하는 과반수의 데이터가 클래스 0으로 오분류 되었다. 이는 클래스 1과 클래스 0간 생체신호 패턴이 일부 유사하거나 클래스 1 표본 수 부족이 원인으로 추정된다. 클래스 1의 판별력을 추가 검증하기 위해 ROC 커브 분석을 진행했다.

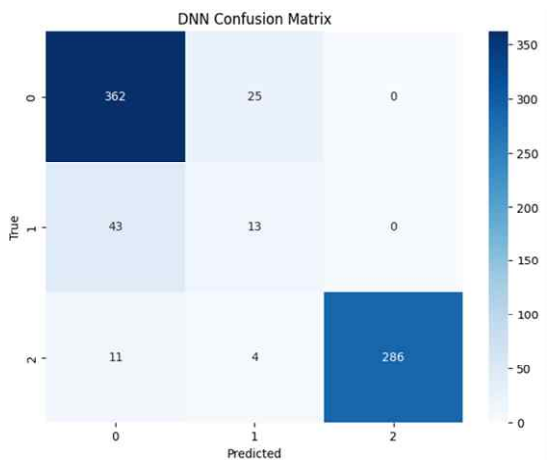


그림 3. DNN 모델의 다중 분류 결과 혼돈 행렬
Fig. 3. Confusion matrix of the DNN model

그림 4를 비교한 결과에 따르면 AUC 0.5를 기준으로 DNN이 전 구간에서 가장 큰 재현율을 유지하였고, 트리계열 모델들이 그다음으로 우수한 성능을 보이는 것을 확인할 수 있었다. 랜덤 포레스트는 초기 위양성률 구간에서 재현율 상승이 완만해 상대적으로 열세였다. 로지스틱 회귀의 경우 가장 낮은 위치를 보여주었다.

이와 같은 결과는 비선형적이고 계층적인 패턴이 두드러진 웨어러블 생체정보에서 DNN이 전통적인 머신러닝 모델과 선형 데이터 전용 딥러닝 모델들보다 일관되게 높은 성능이 보임을 시사한다. 한편, SMOTE로 클래스 불균형을 보정해도 클래스 1처럼 경계가 모호한 표본은 추가적인 표본 확충 또는 추가적인 학습이 요구된다. DNN은 높은 정밀도와 재현율을 동시에 확보했으므로, 우울증 위험군 사전 선별에 가장 적합한 모델로 예상된다.

V. 결론 및 향후 과제

본 연구에서는 웨어러블 기기를 통해 수집할 수 있는 생체정보와 PHQ-9 설문 데이터를 이용하여 우울증 위험 척도를 예측하는 모델을 제안했다. 제

안한 방법은 데이터의 특성을 고려한 결측치 처리, 이상치 제거, 특성 선택, 다항 특징 생성, 클래스 불균형 처리와 같은 데이터 전처리를 진행한 후 전통적 머신러닝 모델들과 정형 데이터 전용 딥러닝 모델들을 최적화된 DNN 모델과 성능 비교를 진행했다[26]. 이를 통해 우울증 위험 척도를 예측하는 최적의 DNN 모델을 제안했다.

제안된 DNN 모델은 하이퍼파라미터 튜닝, 학습률 감소, 조기 종료 등의 학습 방법을 적용하여 F1-스코어 0.883188을 달성했다. 이는 머신러닝 모델 중 가장 우수한 성능을 달성한 LightGBM 보다 0.00765 높은 수치로, DNN이 복잡한 비선형 관계를 학습할 수 있음을 보여준다. 이는 DNN이 우울증 위험 척도 예측 문제에서의 적용 가능성과 효용성을 입증한다.

본 연구에서 제안한 웨어러블 기기를 통해 수집 가능한 데이터를 활용한 우울증 위험 척도 예측 방법은, 스마트워치와 같은 웨어러블 기기의 사용률이 증가하는 현시점에서 조기 선별 및 예방 도구로서의 활용 가능성을 제시한다. 이는 개인의 생체 데이터를 기반으로 실시간 정신건강 모니터링 시스템 구축에 기여할 수 있다.

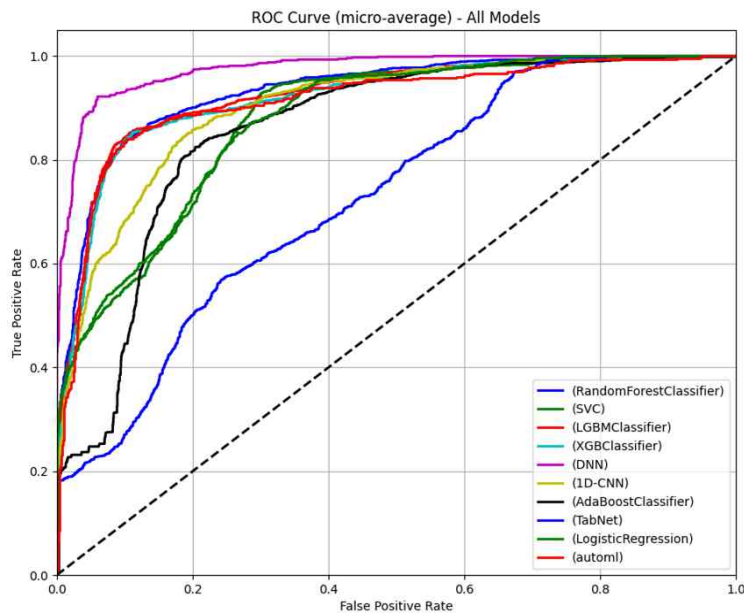


그림 4. 증상 의심군에 대한 ROC 커브 곡선
Fig. 4. ROC curve for class 1

그러나 본 연구는 국민건강영양조사의 데이터를 기반으로 수행되어, 다양한 국가나 인구 집단에 대한 환경에서 일반화하는 데 한계가 있을 수 있다. 향후 연구에서는 다양한 인구 집단을 포함한 데이터셋의 활용을 통해 모델의 보편성과 실용성을 강화할 필요가 있다. 또한, 혼동 행렬 분석 결과 중간 수준의 우울증에 해당하는 클래스에서 비교적 높은 오분류율이 관찰되었다. 이는 클래스 간 데이터 불균형 문제와 관련이 있을 수 있으며, 이를 개선하기 위해 추가적인 특징 공학과 함께 데이터 균형을 위한 보완 작업이 필요하다.

본 연구의 목표는 임상 현장과 정신건강 관리 시스템에 적용할 수 있는 실시간 우울증 위험 모니터링 시스템을 개발하는 것이다. 이를 위해 향후 연구에서는 실시간 자료 수집 및 분석 체계 구축, 실제 환경에서의 성능 검증, 사용자 맞춤형 인터페이스 설계 등 다양한 기술적 요소를 종합적으로 고려할 계획이다.

References

- [1] World Health Organization, "Depression and Other Common Mental Disorders: Global Health Estimates", <https://www.who.int/publications/i/item/depression-and-other-common-mental-disorders-global-health-estimates>. [accessed: Jun. 03, 2025].
- [2] World Health Organization, "Mental Health Atlas 2021", <https://www.who.int/publications/i/item/9789240036703>. [accessed: Jun. 03, 2025].
- [3] R. C. Kessler, E. J. Bromet, and K. G. Kahl, "The burden of depressive disorders in the 21st century", *Journal of Affective Disorders*, Vol. 229, pp. 19-26, Jan. 2018. <https://doi.org/10.1016/j.jad.2017.11.027>.
- [4] K. Smith, S. Jiang, R. Patel, and X. Liu, "Depression and cardiovascular disease: A bidirectional relationship", *The Lancet Psychiatry*, Vol. 7, No. 4, pp. 305-315, Apr. 2020. [https://doi.org/10.1016/S2215-0366\(19\)30432-9](https://doi.org/10.1016/S2215-0366(19)30432-9).
- [5] OECD, "COVID-19 and Mental Health: Policy Responses and Emerging Issues", <https://www.oecd.org/coronavirus/policy-responses/covid-19-and-mental-health-policy-responses-and-emerging-issues-2de8018b/>. [accessed: Jun. 03, 2025].
- [6] R. S. McGinnis, P. Siktberg, and N. Carlisle, "Wearable sensor-based behavioral and physiological monitoring for depression", *Journal of Medical Internet Research*, Vol. 23, No. 5, pp. e25776, May 2021. <https://doi.org/10.2196/25776>.
- [7] Statista, "Global smartwatch market: Adoption and usage trends", <https://www.statista.com/statistics/538237/global-smartwatch-unit-shipments/>. [accessed: Jun. 03, 2025].
- [8] C. Han, et al., "Validation of the Patient Health Questionnaire-9 (PHQ-9) Korean version in the elderly population", *Comprehensive Psychiatry*, Vol. 49, No. 2, pp. 218-223, Mar. 2008. <https://doi.org/10.1016/j.comppsy.2007.08.006>.
- [9] S.-J. Park, Y.-H. Kim, and M.-S. Lee, "Reliability and validity of the Korean version of PHQ-9", *Journal of Anxiety Disorders*, Vol. 15, No. 2, pp. 45-52, Jul. 2010. <https://doi.org/10.32835/anxiety.2010.15.2.45>.
- [10] Yuuki Tazawa "Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning", *Heliyon*, Volume 6, Issue 2, 2020, e03274, ISSN 2405-8440, <https://doi.org/10.1016/j.heliyon.2020.e03274>.
- [11] M.-J. Kim, "Building a cardiovascular disease prediction model for smartwatch users using machine learning: Based on the Korea National Health and Nutrition Examination Survey", *Biosensors*, Vol. 11, No. 7, pp. 228, Jul. 2021. <https://doi.org/10.3390/bios11070228>.
- [12] A. Shcherbina, C. M. Mattsson, D. Waggott, H. B. Salisbury, J. W. Christle, T. Hastie, M. T. Wheeler, M. J. Pletcher, and E. A. Ashley, "Accuracy in wearable optical heart rate monitors", *JAMA*, Vol. 317, No. 7, pp. 626-628, Feb. 2017.

- <https://doi.org/10.1001/jama.2017.0118>.
- [13] U.S. Food and Drug Administration, "Wearable health technologies and their regulatory approvals", <https://www.fda.gov/medical-devices/digital-health-center-excellence/>. [accessed: Jun. 03, 2025]
- [14] S. Park, H. Cho, and K. Lee, "Evaluation of wearable sensor-based monitoring for mental health: A systematic review", *Journal of Medical Systems*, Vol. 45, No. 7, pp. 110, Jul. 2021. <https://doi.org/10.1007/s10916-021-01756-7>.
- [15] T. Shin, D.-G. Lee, S.-W. Song, and W. Kim, "A study on integrated technologies of wearable devices and depression screening tests for predicting depression risk", *The Society of Convergence Knowledge Transactions*, Vol. 12, No. 1, pp. 65-75, Jan. 2024.
- [16] T. Mullick, P. Kulkarni, and S. Rao, "Predicting depression in adolescents using mobile and wearable sensors: Multimodal machine learning-based exploratory study", *JMIR Formative Research*, Vol. 6, No. 2, pp. e33629, Feb. 2022. <https://doi.org/10.2196/33629>.
- [17] A. Horwitz, T. Zimmerman, and S. Patel, "Utilizing daily mood diaries and wearable sensor data to predict depression and suicidal ideation", *Journal of Affective Disorders*, Vol. 310, pp. 72-80, Apr. 2022. <https://doi.org/10.1016/j.jad.2022.04.023>
- [18] R. Gupta, A. Sen, and C. Graham, "Smartwatch data in detecting mental health issues: A machine learning perspective", *Sensors*, Vol. 20, No. 10, pp. 2905, May 2020. <https://doi.org/10.3390/s20102905>.
- [19] D. Kim and J. Choi, "Deep learning-based approaches for mental health monitoring using wearable sensors", *IEEE Access*, Vol. 10, pp. 17842-17850, Feb. 2022. <https://doi.org/10.1109/ACCESS.2022.3151234>.
- [20] Y. Kwon and H. Park, "Automated machine learning for depression prediction: A systematic review", *IEEE Access*, Vol. 9, pp. 12345-12359, Mar. 2021. <https://doi.org/10.1109/ACCESS.2021.3053278>.
- [21] Y. Zhang, H. Li, and J. Wang, "Wearable device data analysis for mental health monitoring: A deep learning approach", *IEEE Journal of Biomedical and Health Informatics*, Vol. 25, No. 4, pp. 1234-1242, Apr. 2021. <https://doi.org/10.1109/JBHI.2020.3049364>.
- [22] S. O. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning", *Proc. of the AAAI Conference on Artificial Intelligence*, held virtually, Vol. 35, No. 8, pp. 6679-6687, Feb. 2021. <https://doi.org/10.1609/aaai.v35i8.16826>.
- [23] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *Proc. of the 32nd International Conference on Machine Learning (ICML 2015)*, Lille, France, pp. 448-456, Jul. 2015.
- [24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *Proc. of the 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, California, USA, May 2015.
- [25] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", *Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929-1958, Jun. 2014.
- [26] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning", *MIT Press*, Vol. 1, No. 2, pp. 1-800, Nov. 2016.

저자소개

박 상 훈 (Sang Hoon Park)



2023년 2월 : 경상국립대학교
컴퓨터공학과(공학사)
2024년 9월 ~ 현재 :
경상국립대학교 AI융합공학과
석사과정
관심분야 : 인공지능,
거대언어모델, XAI, 헬스케어

하 유 진 (Yu Jin Ha)



2023년 2월 : 경상국립대학교
컴퓨터공학부(공학사)
2023년 3월 ~ 현재 :
경상국립대학교 AI융합공학부
석사과정
관심분야 : 인공지능, 멀티모달,
컴퓨터 비전, 헬스케어

김 건 우 (Gun-Woo Kim)



2006년 12월 : 호주뉴캐슬대학교
컴퓨터공학과(공학사)
2007년 9월 : 호주뉴캐슬대학교
정보공학과(공학석사)
2017년 8월 : 한양대학교
컴퓨터공학과(공학박사)
2021년 9월 ~ 현재 :

경상국립대학교 컴퓨터공학과 부교수
관심분야 : 인공지능, 시멘틱 헬스케어, 데이터마이닝