잡음제거 오토인코더(CNN AutoEncoder) 기반 졸음 및 이상치 탐지와 통계/군집/XAI 해석에 관한 연구

-척도 및 변수 개발을 중심으로-

2025.11

한양사이버대학교 일반대학원
주 저자 : 석사과정 서동관
교신 저자 : 박사과정 양건홍, 석사과정 박중헌,
석사과정 박민재

연구관련 코드 깃허브 주소입니다.

본 연구는 한양사이버대학교 일반대학원 기계/IT융합공학 학과 로봇공학특론 과제로 진행된 연구임을 밝혀 둡니다.

1. 논문 초록(국문)

-초록-

본 연구는 잡음제거 합성곱 오토인코더(Conv1D Denoising AutoEncoder, DAE)를 기반으로 운전자의 졸음 및 이상치 상태를 감지하고, 통계·군집 분석 및 설명 가능한 인공지능(XAI)을 결합하여 해석 가능한 졸음 탐지 프레임워크를 제안한다. 이를 위해 먼저 MediaPipe FaceMesh로부터 눈 및 얼굴 랜드마크를 추출하고, 고개 기울기 변화 비율(tilt_diff_ratio), 양안 개방 합(open_sum), 양안 개방 지속 길이(both_open_run), 편측 눈 감김 지속 길이(one_closed_run), 양안 감김 지속 길이(both_closed_run) 등 다섯 개의 시계열 변수를 새로 정의하였다. 정상 주행 영상으로만 Conv1D DAE를 학습한 뒤 재구성오차 분포로부터 시퀀스·프레임·초·3초 블록·비디오 수준의 다중 시간 스케일 졸음 척도와 임계값을 도출하고, 이를 이용해 3초 블록 단위 졸음 라벨을 비지도적으로 생성하였다. 이후 블록 단위 요약 피처와 AE 기반 이상 프레임 통계값을 입력으로 DecisionTree, RandomForest, LightGBM 등 분류 모델을 학습한 결과, 최종 DecisionTree 모델은 F1 $\approx$ 0.95, F2 $\approx$ 0.98, AUC $\approx$ 0.997, FN=0, FP=1의 우수한 성능을 보였다. K-means 군집 분석과 PCA/t-SNE 시각화에서는 졸음·정상 블록이 뚜렷이 분리되었고, Mann–Whitney U 및 2-표본 비율 검정에서도 p $\approx$ 0 수준의 통계적 유의성이 확인되었다. XAI 분석 결과, AE 재구성오차는 주로 자세 흔들림(tilt_diff_ratio), 양안 감김 run(both_closed_run), 눈 열림 정도(open_sum) 등에 의해 설명되었으며, 최종 분류기는 3초 블록 내 이상 프레임 개수(outlier_count)를 핵심 규칙으로 활용하는 것으로 나타났다. 본 연구는 졸음 행동을 직접 설명하는 변수와 다중 시간 스케일 척도를 개발하고, 비지도 AE–통계–군집–지도학습–XAI를 하나의 논리모형으로 통합했다는 점에서, 향후 운전자 상태 모니터링(DSM) 시스템의 설명 가능하고 실용적인 설계 방향을 제시한다.

-목차-

1. 배경

교통사고는 현대 사회에서 가장 큰 문제 중 하나이다. 특히 졸음운전은 교통사고의 주요 원인 중 하나로 높은 위험성을 내포하고 있다. 운전 중 잠시 졸음이 찾아온다면 그 결과는 치명적일 수 있다. 예를 들어, 시속 100km로 달리던 운전자가 단 3초 동안 졸면, 그것은 84m를 맹목적으로 질주하는 것과 같다. 이런 졸음운전의 위험성은 통계를 통해 더욱 명확하게 확인할 수 있다. 2019년 기준 국내 교통사고 사망자의 약 24.8%가 졸음운전에 의해 발생한 것으로 확인되었는데, 이는 매우 경각심이 필요한 상황이다. 이에 본 연구를 통해 이 문제를 해결해 보고자 하였다.

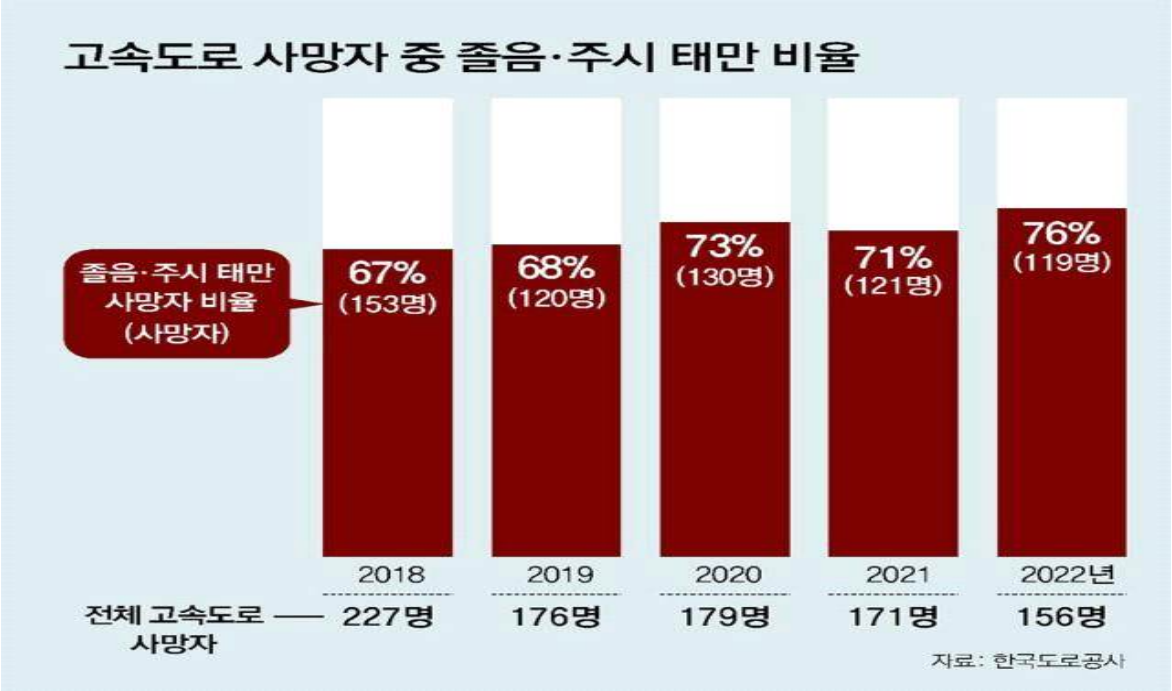


그림 1 졸음운전 고속도로 사망자 비율

이와 관련하여 최근 자율주행 및 운전자 보조 시스템(Advanced Driver Assistance Systems, ADAS)의 보급과 함께 운전자 상태 인식(Driver State Monitoring, DSM) 기술의 중요성이 급격히 증대되고 있다. 특히 졸음(drowsiness)과 피로(fatigue)는 여전히 교통사고의 주요 원인 중 하나로, 이를 사전에 감지하고 경고하는 기술은 안전·법규·책임성 측면에서 필수적인 핵심 요소로 인식되고 있다.

기존의 졸음 탐지 연구는 크게 세 가지 축으로 발전해 왔다.

첫째, 카메라 기반의 컴퓨터 비전(Computer Vision) 기법을 활용해 눈 깜빡임 빈도, 눈 감김 비율(PERCLOS), 눈 형태(EAR) 등을 추출하여 SVM, k-NN 등의 전통적 머신러닝 분류기에 입력하는 방식이다.

둘째, EEG, ECG, EOG 등 생체 신호(Physiological Signal)를 직접 측정하여 졸음

상태를 분류하는 방식이다.
셋째, 스티어링 휠 조향, 차선 이탈, 가속·제동 패턴 등 차량 데이터(Vehicle Data)를 분석하여 졸음을 간접적으로 추정하는 접근법이다.

최근에는 이러한 특징 추출 과정을 자동화한 CNN, RNN/LSTM 기반의 엔드투엔드(end-to-end) 딥러닝 모델도 활발히 연구되고 있다. 그러나 이들 연구는 공통적으로 지도학습(supervised learning) 기반 이진 분류기(졸음/정상)에 강하게 의존하며, 그로 인해 다음과 같은 구조적 한계를 가진다.
졸음/정상 라벨을 프레임 또는 짧은 구간 단위로 부여해야 하므로 라벨링 비용이 매우 크다.
실제 주행 데이터에서 졸음 상태는 전체 시간 중 극히 일부에 불과해, 심각한 클래스 불균형(class imbalance)이 발생한다.
모델이 "왜" 특정 구간을 졸음으로 판정했는지 설명하기 어려워, 안전·인증 및 책임 추적 측면에서 신뢰성 확보가 쉽지 않다.
무엇보다, "졸음 행동을 정량적으로 표현하는 척도(metric)와 변수(feature set)를 체계적으로 정의한 연구"는 상대적으로 부족하다.

이에 비해 AutoEncoder(AE) 기반의 이상 탐지(Anomaly Detection) 기법은 정상(normal) 상태 데이터만을 사용하여 모델을 학습하고, 새로운 데이터의 복원 오차(Reconstruction Error)를 통해 이상 상태를 탐지할 수 있다.
특히 잡음제거 오토인코더(Denoising AutoEncoder, DAE)는 입력에 노이즈를 주입해도 정상 패턴을 복원하도록 학습되므로, 실제 운전 환경의 조명 변화, 센서 노이즈, 일부 프레임 손실 등에도 강건한 장점을 가진다.

본 연구는 이러한 AE/DAE의 장점을 "단순 오차 기반 이상 탐지" 수준에 머무르지 않고, 졸음 행동을 설명하는 새로운 시계열 변수 5개를 개발하고, 이로부터 프레임·초·3초 블록·비디오 수준까지 이어지는 다중 시간 스케일 졸음 척도(metric)를 설계하며, 정상 데이터만으로 학습된 Conv1D Denoising AutoEncoder(DAE) 를 이용해 비지도(unsupervised)로 이상치를 탐지하고, 통계 검정·군집 분석·XAI·지도 분류모델까지 통합함으로써
"척도 및 변수 개발을 중심으로 한, 설명 가능한 졸음 및 이상치 탐지 프로그램 논리모형" 을 구축하는 것을 목표로 한다는 점에서 기존 연구와 차별성을 가진다.

2. 필요성

2.1 라벨링 및 데이터 불균형 문제

실제 도로 주행 환경에서 "이 시점이 정확히 졸음이다" 를 프레임 수준에서 정밀하게 라벨링하는 것은 거의 불가능에 가깝다. 졸음 유도를 위해 피실험자를 강제로 피로 상태에 두는 것도 윤리적·안전적 제약이 크다. 또한 대부분의 운전 시간은 정상 상태이며 졸음 상태는 전체 시간 중 극히 일부에 불과하다. 이로 인해 지도학습 기반의 이진 분류 모델은 다음과 같은 문제를 겪는다.

라벨링 비용이 매우 높아, 실용적인 데이터셋을 구축하기 어렵다.
훈련 데이터에서 정상:졸음 비율이 극단적으로 불균형해, 모델이 '정상'만 예측해도 높은 정확도를 보이는 현상이 발생한다.
졸음 상태 정의가 연구자·실험자에 따라 달라져, 모델의 기준과 임계값이 일관되지 않다.

2.2 시계열 패턴의 복잡성

졸음은 단일 프레임에서 "눈이 감겼다/떴다" 로만 나타나는 현상이 아니라,

- 눈 열림 정도 감소
- 눈 감김(run length) 증가
- 한쪽 눈만 감기는 비대칭 깜빡임
- 고개 기울기 증가 및 자세 흔들림
- 깜빡임 주기의 불규칙성 증가

등 시간에 따라 누적되는 행동 패턴으로 나타난다. 따라서 프레임 단위의 정적 특징만으로는 졸음을 안정적으로 판정하기 어렵고, 프레임 → 시퀀스 → 초 → 3초 블록 → 비디오로 이어지는 다중 시간 스케일 정보가 필수적이다.

2.3 설명 가능성(Explainability)의 요구

DSM 시스템은 운전자의 안전과 직결되므로, 단순히 "졸음이다/아니다" 라는 이진 결과만 제시하는 것은 충분하지 않다.
실제 차량 시스템에 적용하기 위해서는 다음과 같은 설명 가능성이 요구된다.

- 어느 구간에서 눈 감김(run length)이 길어졌는지,
- 고개 기울기(tilt)가 얼마나 증가했는지,
- AE 기준 이상 프레임이 얼마나, 몇 초 동안 지속됐는지,
- 어떤 변수들이 졸음 판정에 가장 큰 기여를 했는지,

이러한 요구를 만족하기 위해서는 명확한 척도(metric)와 그것을 구성하는 변수(feature)의 개발이 선행되어야 하며, 그 위에 XAI(Explainable AI)를 결합해 수학적·통계적 근거를 가진 졸음 판정이 가능해야 한다.

2.4 본 연구의 필요성

위와 같은 문제의식을 바탕으로, 본 연구는 다음을 목표로 한다.
❶ 정상 데이터만으로 학습 가능한 비지도/원클래스(one-class) 구조(AE/DAE) 를 채택하여 라벨링 비용과 데이터 불균형 문제를 완화한다.
❷ MediaPipe 기반 얼굴 랜드마크를 이용해 졸음 행동을 직접 설명하는 5개 시계열 변수를 새로 정의하고, 이를 AE 입력으로 사용한다.
❸ 프레임·초·3초 블록·비디오 수준에서 복원 오차·이상 프레임 비율을 정량화하여

졸음 척도를 구성한다.
❹ XAI 기법(SHAP, Permutation Importance, Ablation 등)을 통해 각 변수와 척도가 졸음 판정에 기여하는 정도를 수치화하고 시각화함으로써, 설명 가능한 DSM 모델을 제시한다.

3. 목적

본 연구의 궁극적인 목적은 잡음제거 CNN/Conv1D AutoEncoder를 이용해 운전자의 눈·자세 움직임 정상 패턴을 학습하고, 이를 기반으로 졸음 및 이상치 상태를 정량적으로 정의·탐지·설명할 수 있는 모델과 척도를 제시하는 것이다. 이를 위해 다음과 같은 세부 목표를 설정하였다.

3.1 변수 개발을 통한 정상 운전 패턴 모델링

MediaPipe로 추출한 눈·얼굴 랜드마크로부터
- tilt_diff_ratio (얼굴 기울기 변화 비율),
- open_sum (양쪽 눈 개방 정도 합),
- both_open_run (양 눈이 동시에 열린 상태의 run 길이),
- one_closed_run (한쪽 눈만 감긴 run 길이),
- both_closed_run (양 눈이 동시에 감긴 run 길이)

  의 5개 시계열 특징을 정의한다.

이들은 기존 EAR/PERCLOS 지표를 넘어, 졸음 행동의 시간적·행동학적 패턴을 직접 표현하도록 설계된 새로운 변수이다.
정상 구간(train_*.csv)만을 사용하여 Conv1D Denoising AutoEncoder를 학습하고, 시퀀스·프레임 단위 재구성 오차 분포를 기반으로 글로벌 임계값(threshold)을 도출한다.

3.2 다중 스케일 이상 탐지 및 자동 라벨링

프레임 단위 이상 라벨을 바탕으로 초, 분, 3초 비중첩 블록 단위 이상비율과 이상 프레임 개수를 산출한다.
특히 3초 블록 내 outlier_ratio와 outlier_count 조합을 이용해 졸음 블록(sleepy_label=1)을 자동 정의하고, 901개 블록 중 50개 졸음 후보 블록을 안정적으로 라벨링한다.
이를 통해 AE 기반 비지도 라벨링 → 블록 단위 지도학습으로 이어지는 하이브리드 구조를 구축한다.

3.3 블록·비디오 수준 최적 분류 모델 도출

- 블록 단위 요약 특징(5개 원시 변수의 mean/std/max,

- AE 재구성오차 mean/max, 이상 프레임 개수,
- 파생 비율 지표,
- KMeans 메타 피처)을 입력으로
- DecisionTree, RandomForest, SVM, LightGBM, XGBoost 등 여러 분류 모델을 학습한다.

F1/F2-score와 ROC-AUC를 중심으로 최적 모델과 임계값을 선정하고, FN 최소화(졸음 미탐지 방지)를 우선하는 기준으로 평가한다.

3.4 설명 가능한 모델(XAI) 구축

RandomForest 서러게이트 회귀를 통해 AE 재구성오차를 설명하고, SHAP을 사용해 어떤 변수(고개 흔들림, 눈 닫힘 run, 평균 눈 열림 등)가 AE 오차를 키우는지 시각화한다.

최종 블록-레벨 분류기에는 Permutation Importance와 Drop-one Ablation을 적용하여 outlier_count 등 핵심 변수의 기여도를 검증하고, 모델이 근본적으로 어떤 규칙을 학습했는지 이해 가능한 형태로 정리한다.

3.5 운영 가능한 임계값(Alarm Threshold) 설정

- 시퀀스 MSE 95퍼센타일,
- 3초 롤링 MSE 95퍼센타일,
- 초 단위 mean outlier ratio ROC/YoudenJ,
- 비디오별 이상 시퀀스 개수 95퍼센타일,
- 3초 블록 이상비율 등 다양한 후보 임계값을 비교·검증한다.

이를 통해, 예를 들어 "3초 블록 내 이상치 비율 ≥ 0.65 또는 이상 프레임 개수 ≥ 6이면 졸음 경고" 와 같이 실시간 시스템에서 사용할 수 있는 규칙형 임계값을 제안한다.

4. 방법

본 연구는 크게 아래의 절차로 수행되었다.

❶ 데이터 수집 및 전처리
❷ AutoEncoder 학습
❸ 임계값 탐색 및 자동 라벨링
❹ 분류 모델 학습
❺ XAI 및 통계 검정

4.1 데이터 수집

전면 카메라(720p, 30fps) 환경에서 정상 주행 세션 3개(train_0~2), 졸음 및 이상치

유도 세션 3개(test_0~2)를 수집하였다.
각 세션은 서로 다른 운전자 또는 주행 상황으로 구성되었으며, 총 6개 영상에서 정상 프레임 약 48,219개, 졸음 후보 프레임 약 33,204개를 확보하였다.

4.2 데이터 전처리 및 변수 생성

상위 파이프라인에서 MediaPipe FaceMesh를 이용해 얼굴 랜드마크를 추출한 뒤, 양쪽 눈 주변 좌표를 기반으로 다음 5개 시계열 변수를 계산하였다.

- tilt_diff_ratio : 얼굴 기울기 변화 비율(자세의 흔들림 정도)
- open_sum : 양쪽 눈 개방 정도 합
- both_open_run : 양 눈이 동시에 열린 상태의 run 길이
- one_closed_run : 한쪽 눈만 감긴 상태의 run 길이
- both_closed_run : 양 눈이 동시에 감긴 상태의 run 길이

이들은 졸음 시 고개가 점점 떨어지고, 눈이 덜 떠지며, 감긴 상태가 길어지는 특성을 정량적으로 표현하기 위해 설계된 변수이다.
각 변수에 대해 로그 변환 및 RobustScaler 기반 정규화를 적용하고, FPS=30 기준 시계열을 정렬해 AE 입력으로 사용하였다.

4.3 시계열 기반 Conv1D Denoising AutoEncoder(DAE) 설계 및 학습

본 연구의 실제 실험은 MediaPipe FaceMesh에서 계산된 5개 시계열 변수를 입력으로 하는 Conv1D Denoising AutoEncoder(DAE) 를 중심으로 수행되었다.
입력으로 사용된 5개 특성은 다음과 같다.

- tilt_diff_ratio
- open_sum
- both_open_run
- one_closed_run
- both_closed_run

(1) 윈도우링(Windowing)

프레임 시계열 전체를 AE에 직접 넣지 않고,
길이 L = 20 프레임(약 0.67초), stride S = 3 프레임으로 슬라이딩 윈도우를 구성해 (20, 5) 형태의 시퀀스 텐서를 만들었다.
이 윈도우링은 학습–추론–라벨링 전 과정에서 동일하게 적용된다.

(2) 모델 구조

인코더
GaussianNoise(std = 0.05)
Conv1D(64, kernel=3, activation='relu')

Conv1D(64, kernel=3, activation='relu')
Conv1D(32, kernel=3, activation='relu', name='latent')

디코더
Conv1D(64, kernel=3, activation='relu')
Conv1D(64, kernel=3, activation='relu')
Conv1D(F, kernel=3, activation='linear')
손실함수는 MSE, 최적화 함수는 Adam(lr=1e-3)을 사용하였다.

(3) 학습 전략
학습 데이터는 정상 구간(train_*.csv) 만 사용 validation split 및 EarlyStopping 적용
여러 구조 조합(윈도우 길이, latent 차원, noise std 등) 비교 후,
L=20, S=3, latent=32, depth=2, noise_std=0.05 구성이 재구성오차 분포 및 졸음 구간 분리도가 가장 우수하여 최종 모델로 선정

(4) 모델 저장 및 재현성 확보
학습된 모델 및 전처리 스케일러는 다음 파일로 저장하였다.
- best_dae.keras
- best_dae_weights.weights.h5
- best_dae.meta.json (L, latent, depth 등 구조 정보)
- scaler_robust.joblib

이를 통해 이후 실험에서도 동일한 환경에서 재현이 가능하도록 설정하였다.

4.4 시퀀스·프레임·초·블록 단위 라벨링 및 임계값 탐색

4.4.1. 시퀀스 MSE 기반 1차 임계값

train 세트 시퀀스 MSE의 95퍼센타일을 글로벌 임계값으로 설정.
이를 초과하는 시퀀스를 이상 시퀀스로 정의. seq → frame 변환 및 3초 롤링 기준
각 시퀀스 MSE를 겹침 평균하여 frame MSE를 계산한 뒤, FPS=30 기준
3초(90프레임) 롤링 평균을 구하여 frame_mse_roll3s를 정의.
train 데이터의 95퍼센타일을 3초 롤링 MSE 임계값으로 설정하고, 이를 초과하는 구간을 이상 구간으로 간주하였다.

4.4.2. 초/분 단위 요약 및 ROC 분석

- 초 단위: (video_id, sec)별 이상 프레임 비율(sec_outlier_ratio)과
- 개수(sec_outlier_count)를 계산하고,
- 정상 vs 비정상에 대한 ROC/AUC 및 YoudenJ/Max-F1 임계값을 도출하였다.

일부 시점에서는 AUC=1.0에 가까운 완전 분리가 확인되었다.

분 단위: minute별 평균 이상비율을 계산하여 per-minute trend를 시각화하고, 시간 경과에 따른 졸음 패턴 변화를 거시적으로 관찰하였다.

4.4.3. 3초 비중첩 블록 기준 졸음 블록 정의

FPS=30 기준 90프레임을 한 블록으로 정의하고, 각 블록에 대해
outlier_count, outlier_ratio, recon_err_mean, recon_err_max,
다섯 변수의 mean/std/max 등을 계산하였다.
outlier_ratio ≥ 0.65 AND outlier_count ≥ 6 조건을 만족하는 블록을 졸음 블록(sleepy_label=1)으로 정의하여, 총 901개 블록 중 50개를 졸음 블록, 851개를 정상 블록으로 자동 라벨링하였다.

4.4.4. 비디오 수준 스코어 및 임계값

각 비디오에서 "가장 졸린 3초 구간의 이상비율(score_block3_ratio)"을 스코어로 사용하여 train vs test 비디오를 ROC 분석한 결과, AUC=1.0을 달성하였다.
YoudenJ 기준 임계값 ≈0.72 부근에서 6개 비디오(train 3, test 3)가 완전 분리되어, "3초 구간 중 최대 이상 비율이 0.7 이상이면 졸음 비디오"라는 직관적인 운용 규칙이 가능함을 확인하였다.

4.5 분류기 학습 및 XAI 분석

3초 블록 단위에서 다음과 같이 피처를 구성하였다.
원시 특징 요약(5변수 × mean/std/max = 15개) AE 관련 지표(recon_err_mean, recon_err_max) 이상 프레임 통계(outlier_count) 파생 지표(tilt_err_ratio, eye_close_ratio, open_sum_cv, closed_sum_ratio 등) KMeans 군집 기반 메타 피처(km_dist_c0/1, km_dist_margin, km_cluster_aligned) 총 26차원의 피처 벡터를 구성하고, dependent(sleepy vs awake)를 타깃으로 DecisionTree, RandomForest, SVM, LGBM 등의 분류 모델을 학습하였다. 졸음 블록 비율이 약 5.5%에 불과하므로, train 세트에 대해 SMOTE를 적용하여 클래스 불균형을 완화하였다.
F1-score를 기준으로 하이퍼파라미터 탐색 및 F1-max threshold 튜닝을 수행한 결과, 테스트 셋에서
F1 ≈ 0.95
F2 ≈ 0.98
AUC ≈ 0.997
FN=0, FP=1
이라는 매우 높은 성능을 얻었고,
특히 DecisionTree 모델은 규칙 해석이 용이하여 XAI 및 운용 측면에서 유리하였다.
XAI 분석에서는 RandomForest 회귀 + SHAP을 통해 AE 재구성오차에 영향을 주는 핵심 변수(tilt_diff_ratio_std/max, both_closed_run_max, open_sum_mean 등)를

확인하였고, DecisionTree 분류기에 Permutation Importance 및 Drop-one Ablation을 적용하여 outlier_count가 졸음 판정의 절대적 핵심 변수임을 확인하였다.

4.6 통계 검정 및 군집 분석

AE 기반 라벨링의 타당성을 검증하기 위해,

- 프레임·블록·초 단위에서 Mann–Whitney U,
- 2-표본 비율 검정,
- Shapiro–Wilk 정규성 검정,
- Levene 등분산 검정 등을 수행하였다.

그 결과, 프레임 단위 이상 비율은 정상 $\approx$ 0.14, 비정상 $\approx$ 0.38로 크게 차이났고, Mann–Whitney U와 2-표본 비율 Z-검정에서 p $\approx$ 0 수준의 완전 분리가 나타났다.

또한 KMeans(k=2) + PCA/t-SNE 기반 군집 분석 결과, 정상·졸음 블록이 자연스럽게 두 클러스터로 분리되었으며,

ARI $\approx$ 0.76
NMI $\approx$ 0.57
Silhouette $\approx$ 0.71
등의 높은 군집 품질 지표를 확보하여, AE 기반 척도와 라벨이 실제 행동 패턴을 잘 반영한다는 비지도적 근거를 제공하였다.

5. 범위

본 연구는 소규모 탐색적 실험(exploratory study)으로 수행되었으며, 다음과 같은 범위와 전제를 가진다.

5.1. 데이터 규모 및 구성

영상 세션: 총 6개 (정상 3, 졸음 및 이상치 유도 3)
해상도: 720p, 30fps
프레임 수: 정상 약 4.8만, 비정상 약 3.3만
3초 블록 수: 총 901개(이 중 졸음 및 이상치 블록 50개)

5.2. 분석 단위

프레임 → 시퀀스(20프레임, stride 3) → 초 → 분(60초) → 3초 비중첩 블록(90프레임) → 비디오 단위까지 여러 시간 스케일을 모두 고려하였다.

5.3. 모델 입력 형태

본 논문의 실제 실험과 분석은 모두 MediaPipe FaceMesh로부터 얻은 랜드마크를 기반으로 계산한 5개 시계열 특징
tilt_diff_ratio
open_sum
both_open_run
one_closed_run
both_closed_run
을 입력으로 하는 Conv1D Denoising AutoEncoder 및 이로부터 파생된 블록 단위 요약 피처에 한정된다.

5.4. 학습/검증 전략

AutoEncoder는 정상(train) 데이터만을 사용해 학습하고, test 데이터는 재구성오차 및 이상도로 평가하였다.
블록-레벨 분류기는 block_df에서 80:20 stratified split을 사용하였고, train에 대해 SMOTE를 적용한 뒤 테스트셋에서 F1/F2/ROC-AUC 및 혼돈행렬로 평가하였다.

5.5. 적용 범위

본 연구는 제한된 피실험자·환경에서의 실험 결과이므로 곧바로 상용 DSM 시스템 전체에 일반화하기는 어렵다.
다만, "정상 시계열 패턴을 DAE로 학습하고, AE 오차 + 이상 프레임 비율 + 간단한 트리 기반 분류기로 졸음 블록을 설명 가능하게 탐지할 수 있다"는 프레임워크의 유효성을 보이는 데 초점을 둔다.

6. 이론적 고찰

졸음은 생체 리듬과 신경생리적 원리의 복합적인 상호작용에 의해 발생하는 현상으로, 운전 중 졸음은 치명적 위험을 초래하기 때문에 이를 과학적으로 이해하는 것은 졸음·이상치 감지 시스템 개발의 필수 요소가 된다. 졸음 연구의 기초는 Borbély(1982)가 제안한 수면 조절의 2-과정 모델(Two-Process Model) 에 기반한다. 이 모델에 따르면 인간의 수면 욕구는 수면 항상성(homeostatic sleep drive, 과정 S) 과 일주기 리듬(circadian rhythm, 과정 C) 이라는 두 가지 과정의 상호작용으로 설명된다(Borbély, 1982; Borbély & Achermann, 1999). 과정 S는 깨어 있는 시간이 길어질수록 증가하는 수면압을 의미하며, 이 과정에서 중요한 역할을 하는 것이 뇌에 축적되는 아데노신(adenosine) 이다. 아데노신은 신경 활동의 부산물로 깨어 있을수록 축적되어 뇌의 각성 시스템을 억제하고 수면 상태로 전환하는 신호를 보낸다(Porkka-Heiskanen et al., 1997; Porkka-Heiskanen, 1999). 우리가 흔히 마시는 카페인이 졸음을 억제하는 이유 역시 이 아데노신이

결합해야 하는 수용체를 차단하기 때문이다(Huang et al., 2005).
한편 과정 C는 뇌 시상하부의 시교차상핵(SCN) 을 중심으로 작동하는 생체 시계로, 하루 24시간 주기를 기준으로 각성과 수면의 시점을 조절한다. SCN은 빛이라는 강력한 외부 신호에 의해 조절되며, 밤이 되면 멜라토닌 분비를 증가시켜 신체가 자연스럽게 수면 모드로 전환되도록 한다(Arendt, 2022; Blume et al., 2019). 이 두 과정은 독립적으로 존재하지만 서로 긴밀하게 상호작용하며, 우리가 가장 졸린 시간을 만들어낸다. 예컨대 수면압이 충분히 높아진 밤 시간대에 멜라토닌이 분비되면 졸음은 극대화되고, 점심 이후 나타나는 이른바 '포스트 런치 딥(post-lunch dip)'은 일주기 리듬의 각성 신호가 일시적으로 약해지는 시점과 수면압이 누적되는 시점이 겹치기 때문에 발생한다(Monk, 2005; Bes et al., 2009).
졸음 상태가 되면 인지적·신체적 변화가 복합적으로 나타난다. 가장 두드러진 것은 인지 기능 저하와 반응 속도 지연이다. Dawson & Reid(1997)의 연구에 따르면 17시간 동안 깨어 있을 경우의 인지 기능은 혈중 알코올 농도 0.05% 수준과 유사하며, 24시간 미수면 상태는 0.10%에 해당하는 심각한 판단력 저하를 보인다. 이는 단순한 피로를 넘어 신경계가 외부 자극을 적절히 처리하지 못한다는 의미이며, 운전과 같은 고위험 작업에서는 특히 치명적이다.
더 나아가 심한 졸음 상태에서는 미세수면(microsleep) 현상이 발생할 수 있다. 미세수면은 0.5~15초 사이의 매우 짧은 수면 에피소드로, 사람이 눈을 뜨고 있어도 뇌가 외부 자극 처리를 잠시 멈춘 상태를 말한다(Sleep Foundation, 2023). EEG 측면에서는 각성 상태에서 나타나는 알파파가 사라지고, 수면 1단계(N1)의 특징인 세타파가 순간적으로 출현한다(Patel et al., 2024). 이때 운전자 의식은 사실상 '오프' 상태와 다름없기 때문에 고속도로 사고의 주요 원인 중 하나로 꼽힌다.
또한 졸음의 대표적인 행동 신호 중 하나인 하품은 단순한 피로 표현이 아니라 뇌 온도를 조절하기 위한 생리적 메커니즘이라는 가설이 제기되고 있다. Gallup & Gallup(2007)은 하품이 차가운 공기를 들이마시고 턱 근육을 크게 움직여 뇌 혈류를 식히는 효과를 만든다고 설명했다. 이외에도 졸음이 임박하면 눈꺼풀 근육의 긴장도가 감소하거나 심부 체온이 떨어지는 등 다양한 생리적 변화가 동반된다(AASM, 2014; Zisapel, 2018).

7. 관련 연구

운전자 졸음 감지 기술은 오랫동안 연구가 진행되어 왔으며, 최근에는 AI 기술의 발전과 함께 더욱 정교해지고 있다. 이러한 연구는 대체로 세 가지 범주로 나누어진다.

첫째는 컴퓨터 비전 기반 접근법,
둘째는 생체 신호 기반 접근법,
셋째는 차량 동역학 기반 접근법이다(Ramzan et al., 2019).

첫 번째 접근법인 컴퓨터 비전 기반 방법은 카메라 영상에서 졸음을 암시하는 시각적 특징을 추출하는 방식이다.

- 눈의 깜빡임 패턴, 눈꺼풀이 감겨 있는 비율(PERCLOS),
- 눈의 세로·가로 비율로 계산되는 EAR(Soukupová & Čech, 2016),
- 입 개방도, 머리 자세 변화(head nodding) 등이 대표적이다.

특히 PERCLOS는 NASA 연구(Dinges & Wierwille, 1994) 이후 가장 강력한 졸음 지표로 널리 인정받고 있다. 이 방법의 가장 큰 장점은 비침습적이라는 점이지만, 어두운 환경이나 안경·마스크 착용 시 성능이 저하될 수 있다는 한계도 존재한다.

두 번째 접근법은 생체 신호 기반 방식으로, EEG·ECG·EOG 등 다양한 생리신호로부터 졸음을 직접 검출한다. 특히 EEG는 졸음 상태일 때 나타나는 세타파 증가를 가장 민감하게 반영하며(Latreche et al., 2024), ECG 기반 HRV 분석을 통해 자율신경계의 변화를 감지하는 연구도 활발하다(Vicente et al., 2016). 이러한 방식은 정확도가 높지만 센서 부착이 필요하여 실용성이 떨어진다는 단점이 있다.

세 번째는 차량 데이터 기반 접근법으로, 운전자의 조향 패턴, 차선 유지 능력, 가속·감속 조작 등을 분석하여 졸음을 탐지한다.
예를 들어 Arefnezhad et al.(2019)은 스티어링 휠의 각도 변화를 분석하고, ANFIS 기반의 특징 선택과 SVM 분류기를 결합해 높은 정확도를 보고하였다. 이 방식은 추가 장비가 필요 없다는 장점이 있지만, 운전자의 운전 습관이나 부주의(inattention)와 졸음을 구분하기 어렵다는 문제가 있다.
모델 측면에서는 초기에는 SVM·k-NN 등을 중심으로 한 전통적 머신러닝 기반 연구가 많았다. 눈 깜빡임, EAR, PERCLOS 등의 특징을 명시적으로 추출하여 분류 모델에 입력하는 방식이다(Gwak et al., 2020). 그러나 최근에는 CNN-LSTM 기반의 end-to-end 딥러닝 모델이 주류로 자리 잡고 있다. CNN은 얼굴 영상에서 공간적 특징을 학습하고, LSTM은 프레임 간 시간적 변화를 포착하는 장점이 있어 눈 깜빡임이나 하품과 같은 연속적 패턴을 정교하게 감지할 수 있다(Liu et al., 2022).
이처럼 졸음은 생체 생리적 요인을 기반으로 발생하며, 다양한 행동·생리·차량 운전 패턴으로 표현된다. 따라서 졸음 감지 시스템은 단일 신호보다는 여러 특징을 종합적으로 고려해야 하며, 최근 연구들이 이러한 멀티모달 접근으로 발전하고 있다는 점은 매우 중요한 흐름이라 할 수 있다.

8. 실험 설계

8.1.가설 및 모형설정

본 실험은 EDA(탐색적 데이터 분석)와 특성공학을 토대로 하여 다음과 같이 5개의

귀무가설과 연구가설을 설정하였다.

졸음이 오는 동영상 관찰을 통해 다음과 같은 특징을 추출하였다.
첫째 잠이오는 경우 양안의 감김이 느려진다는 것을 발견하였다.
둘째 잠이오는 경우 양안의 감김이 지속된다는 것을 발견하였다.
셋째 잠이오는 경우 두눈의 위상차이로 한쪽 눈이 먼저 감긴다는 것을 발견하였다.
넷째 잠이오는 경우 양안이 개방되어 있는 시간이 길어짐을 발견하였다.
다섯째 잠이오는 경우 두눈의 기울기가 차이가 난다는 것을 발견하였다.

귀무가설:
첫째 양안의 감김이 느려지는 시간에 집단간 차이가 없다.
둘째 잠양안의 감김이 지속 시간에 집단간 시간의 차이가 없다.
셋째 두눈의 위상차이로 한쪽 눈이 먼저 감김에 집단간 차이가 없다.
넷째 잠이오는 경우 양안이 개방되어 있는 시간에 집단간 차이가 없다.
다섯째 잠이오는 경우 두눈의 기울기 차이에 집단간 차이가 없다.

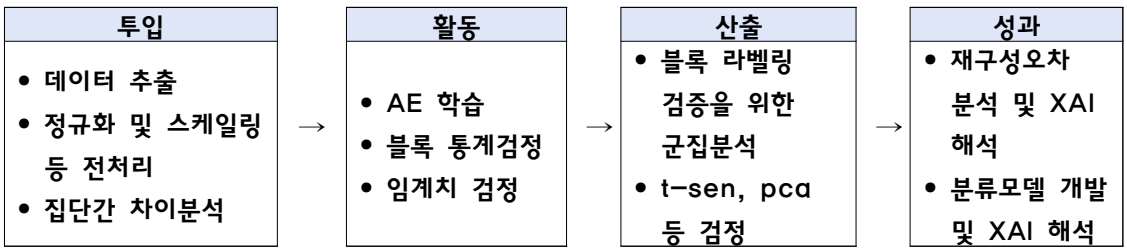
연구가설:
첫째 양안의 감김이 느려지는 시간에 집단간 차이가 있다.
둘째 잠양안의 감김의 지속 시간에 집단간 시간의 차이가 있다.
셋째 두눈의 위상차이로 한쪽 눈이 먼저 감긴에 집단간 차이가 있다.
넷째 잠이오는 경우 양안이 개방되어 있는 시간에 집단간 차이가 있다.
다섯째 잠이오는 경우 두눈의 기울기 차이에 집단간 차이가 있다.

위의 변수를 측정하기 위해 변수를 계량화하여 연속형 척도로 정해서 측정하였다.
1. 프래임당 두눈의 기울기 차이는 비율로 예를 들면 10%, 15%, 55% 등으로 측정하여 측정하여 저장하였다.
2. 프래임당 두눈이 1 또는 0 인 경우 1의 갯수로 예를 들면 두눈을 뜨고 있으면 1 + 1 = 2 를 저장하고 1 + 0 = 1, 0 + 0 = 0 으로 측정하여 저장하였다.
3. 프래임당 두눈이 1로 지속되면 카운트하고 연속해서 30이면 30으로 하고 두눈이 0이 되면 다시 시작하는걸로 해서 측정하여 저장하였다.
4. 프래임당 두눈이 1 과 0 으로 측정된 경우 0 이 지속된 시간 예를 들면 연속해서 0 이 한번이면 1 두번이면 2 .....30이면 30 등으로 저장 후 두눈이 1이되면 다시 시작하여 0이 지속되면 카운트하는 걸로 측정하여 저장하였다.
5. 프래임당 두눈이 0으로 지속되면 카운트하고 연속해서 40이면 40 으로 저장후 1이 되면 다시 시작하는 걸로 측정하여 저장하였다.
이 5가지 변수를 MediaPipe Face Mesh 기반 눈 ROI 추출로 측정하여 CSV로 저장

하였다.

관찰 대상은 정상적으로 정면을 응시하며 운전하는 3명 / 비정상 3명(졸음이 오는 어린이 2 명의 동영상과 실질적으로 정면을 응시한체 운전을 하지 않고 차량을 주차한 후 15분간 동승한 사람과 자연스럽게 이야기를 나누며 행동하는 1명의 동영상)
센서카메라(눈 영상)
데이터 전처리 MediaPipe Face Mesh 기반 눈 ROI 추출
학습 방법 CNN Denoising AutoEncoder
임계값 탐색법 Percentile / IQR
평가 지표 FPR, TPR, F1, p-value(MW, KS)
XAI 분석 SHAP, Feature MSE 변화량

모형설정: 잡음제거 오토인코더(CNN AutoEncoder) 기반 졸음 및 이상치 탐지와 통계/군집/XAI 해석 프로그램 논리모형



본 연구에서는 졸음 및 이상치 감지 모델의 전체 흐름을 프로그램 논리모형(Logic Model) 관점에서 구조화하여, 각 단계가 어떤 역할을 하고 어떻게 다음 단계로 논리적으로 이어지는지를 분명하게 제시하고자 한다.

본 연구에 있어 모형구축을 위해 프로그램논리모형을 사용하였다.
일반적으로 프로그램이론은 프로그램이 어떻게 작동하여 의도한 결과를 창출하는가에 대한 가정이며(Bickman, 1987; Chen,1990; Donalds on, 2003), 논리모형은 프로그램이론을 시각적으로 표현하는 방법으로 정의된다(Mc Laughlin, & Jordan, 1999; W.K. Kellogg Foundation, 2004; Kaplan, & Garrett, 2005; 김동립·이삼용, p.270에서 재인용).

프로그램이론이란 프로그램이 활동을 통해 의도한 결과를 산출하는 원리를 논리적으로 제시한 진술을 말한다. 즉 프로그램이 기대하는 효과와 그 효과를 생산하기 위하여 의도적으로 시행하는 활동의 연결관계를 자세하고 명료하게 기술한 것이다.

프로그램이론은 프로그램평가를 시행하기 위해 기초적으로 선행되어야 하는 작업인 동시에, 그 자체로 프로그램이 가지는 가정을 구체화하- 73고 검토하는 평가 연구의 독립적인 한 과정으로서 최근 관심을 모으고 있다.

논리모델(logic model)은 프로그램이론을 구축하는데 유용한 도구로 많은 연구에서 사용되고 있다. 이러한 논리모델은 프로그램이론에서 다루어야 할 기본적인 요소들을 제시하고 이를 시각화하여 간결하게 표현할 수 있도록 하기에 프로그램이론을 논리적으로 조직화하는데 편리하면서 유용하다(김지혜, 20 04 p.9에서 재인용).

논리모형은 프로그램(정책, 사업)의 각 구성 요소 간 상호작용을 분석하여, 그 결과로 만들어진 산출물은 무엇이며 각 요소들이 어떻게 결과를 만들게 되는지 보여준다.
프로그램논리모형의 구성요소로는
첫째 상황 또는 문제의 인식이 있으며,
둘째 투입요소로는 모든 종류의 자원이나 원료를 말하고,
셋째 활동요소로는 투입된 자원을 활용하여 목적달성을 위해 프로그램을 실행하는 것을 말하며,
넷째 산출요소 로 활동으로 얻어진 결과물을 사정하는 것이다.
그리고 마지막으로 성과요소로는 프로그램 활동의 결과로 얻어진 편익 또는 변화를 말한다(최영출, 2011, pp.20-22).

이 논리모델의 구성요소들은 학자들마다 분류방식(최소 3단계 또는 5단계이상으로 분류하기도 함)이 틀리지만, 기본적 접근방식은 투입, 활동, 산출, 성과가 필수적 요소라 할 수 있다.
프로그램 논리모형은 특정한 산식 혹은 가설을 수립하여 이를 검증하는 접근방식이 아니라 프로그램자체가 하나의 가설로 간주된다(박태정, 2014, pp.37-38). 이 같은 논리모형을 활용한 국내연구는 2007년 이후 본격화 되었다.

이를 위해 본 연구에서 모델의 흐름을
투입(Inputs)–활동(Activities)–산출(Outputs)–성과(Outcomes) 네 영역으로 구분하여 서술한다.

1) 투입(Inputs): 데이터 추출–전처리–집단 간 차이 분석
먼저, 본 모델의 논리는 데이터의 품질과 구조를 얼마나 잘 준비했는가에서 출발한다.
따라서 투입 단계에서는 다음과 같은 세 가지 축이 핵심이 된다.

❶ 데이터 추출

졸음 및 정상 상태를 구분하기 위해, 원시 영상으로부터 눈 깜빡임 관련 시계열 특징(예: EAR, tilt_diff_ratio, open_sum, run-length 계열 등)을 프레임 단위로 추출한다.

이 과정에서 각 영상 파일별로 프레임 인덱스, 시간 정보, 그룹 라벨(정상/졸음) 등이 함께 정리되며, 이후 모든 분석의 기반이 되는 표 형태의 원천 데이터셋이 구축된다.

❷ 정규화 및 스케일링 등 전처리

추출된 특징들은 분포, 단위, 스케일이 서로 다르기 때문에 그대로 AE에 입력할 경우 특정 변수만 과도하게 영향을 미치거나 이상값에 민감하게 반응할 위험이 있다.

이를 방지하기 위해 로그 변환, 비선형 변환, 정규화·스케일링(RobustScaler 등) 을 단계적으로 적용하여, 영상 간·피처 간 스케일 차이를 줄이고 이상값(극단값)의 영향을 완화하며AE가 안정적으로 패턴을 학습할 수 있는 입력 공간을 조성한다.

❸ 집단 간 차이 분석(기초 통계 검정)

전처리된 데이터를 바탕으로, 정상(train)과 졸음(test) 구간 사이에 각 특징이 실제로 유의미한 차이를 보이는지 통계 검정을 수행한다(예: Welch t-test, Mann–Whitney U test 등).이는 두 가지 의미를 가진다.

첫째, "AE가 복원오차를 통해 구분해야 하는 구조적 차이가 실제로 존재하는지"를 사전 점검하는 단계이다.

둘째, 이후 블록 단위 라벨링과 분류모델 설계 시, 어떤 변수들이 의미 있는 후보 피처인지에 대한 초기 가설(hypothesis)을 제공한다.

이와 같이 투입 단계는 단순히 "데이터를 준비한다"는 수준을 넘어, AE 및 이후 모든 분석이 신뢰할 수 있는 기반 위에서 진행되도록 하는 품질 보증 단계로 기능한다.

2) 활동(Activities): AE 학습–블록 통계 검정–임계치 검정–블록 라벨링

투입 단계에서 정제된 데이터를 바탕으로, 활동 단계에서는 실질적인 모델링·통계적 의사결정이 이루어진다.

이 단계는 다시 네 가지 하위 활동으로 구성된다.

❶ AE(오토인코더) 학습

Conv1D Denoising AutoEncoder(DAE)를 이용하여 정상 운전자 데이터만으로 복원 모델을 학습한다.

모델은 입력 시계열(예: L=20프레임 × 5개 피처)을 압축(latent space)했다가 다시

복원하는 과정을 통해, 정상 패턴의 구조를 내부 표현으로 학습하게 된다.
이때 핵심은 "정상 패턴에 익숙한 모델은 정상 구간을 잘 복원하고, 이상 구간은 상대적으로 잘 복원하지 못한다"는 점이다.
이 가정이 이후 복원오차(reconstruction error)를 이용한 이상치 탐지의 논리적 토대가 된다.

❷ 블록 단위 통계 검정(Per-block statistical analysis)
프레임/초 단위 복원오차와 이상치 비율을 바탕으로, 일정 길이의 시간 블록(예: 3초, 30프레임 묶음 등)을 정의하고 각 블록에 대해 평균 복원오차 이상치 프레임 비율 run-length 요약 통계 등을 계산한다.
이렇게 만들어진 블록 수준 통계량에 대해, 정상 vs 졸음 블록 간 차이를 검정함으로써 "어떤 지표 조합이 졸음 상태를 가장 잘 설명하는가"를 탐색한다.

❸ 임계치 검정(Threshold validation)
AE 출력으로부터 얻은 복원오차를 이용해 임계값(threshold) 후보를 설정하고, ROC 곡선, AUC, Youden J 통계, 최대 F1 점수 등을 기반으로 FP/FN 간의 균형을 고려한 최적 임계값을 평가한다.
이 과정은 "어디까지를 정상으로, 어디부터를 이상으로 볼 것인가"에 대한 모델의 규범적 기준을 수치로 정립하는 단계이다.

❹ 블록 라벨링(Block-level labeling)
앞서 정의된 임계값과 블록 통계량을 이용하여, 각 블록에 대해
"정상 블록", "잠재적 졸음/이상 블록" 으로 라벨을 부여한다.
예를 들어 3초 블록 내 이상치 프레임 비율이 특정 임계값(예: 45% 또는 최적화된 값)을 초과하면 '이상 블록'으로, 그 미만이면 '정상 블록'으로 분류하는 식이다.
이 블록 라벨은 이후 군집분석·분류모델 학습·XAI 해석에서 정답 라벨(ground truth proxy) 의 역할을 수행한다.
즉, 활동 단계는 AE로 패턴을 학습하고, 통계적 검정을 통해 임계값을 선정한 뒤, 시간 블록에 의미 있는 라벨을 부여하는 전체 의사결정 파이프라인에 해당한다.

3) 산출(Outputs): 블록 라벨링 검증을 위한 군집분석·t-SNE/PCA 검정
활동 단계의 결과로 생성된 가장 중요한 산출물은 "블록 단위 라벨"이다.
그러나 이 라벨이 실제 데이터 구조와 잘 맞는지 검증하지 않으면, 이후 분류모델이나 XAI 분석이 왜곡될 수 있다.
따라서 산출 단계에서는 라벨의 타당성을 다각도로 검증하는 절차를 수행한다.
군집분석(K-means 등)의 경우 블록 수준 피처(복원오차 평균, 이상치 비율, 원시 피처의 요약통계 등)를 입력으로 하여 K-means(k=2) 등의 군집분석을 수행한다.

이때 한 클러스터는 "데이터 특성상 정상에 가까운 블록들" 다른 클러스터는 "이상 패턴이 두드러지는 블록들"로 분리되는지 확인하고, 이 결과를 기존의 블록 라벨(정상/이상)과 비교함으로써 "AE+임계치 기반 라벨링이 데이터 고유의 구조와 얼마나 일관적인가"를 검증한다.
차원 축소 기반 시각적 검정(t-SNE, PCA)의 경우 고차원 블록 피처들을 PCA나 t-SNE로 2D/3D 공간에 투영하고, 색상으로 블록 라벨(정상/이상)을, 위치로 피처 공간에서의 관계를 시각화한다.
만약 정상 블록과 이상 블록이 저차원 공간에서도 어느 정도 분리되어 나타난다면, 이는 블록 라벨링이 실제 데이터 분포를 잘 반영하고 있다는 시각적·직관적 근거가 된다.
반대로 두 집단이 완전히 뒤섞여 있다면 임계값 설정이나 피처 정의를 재검토해야 한다는 신호가 된다.

이와 같이 산출 단계는 단순히 라벨이 생성되었다는 결과를 넘어, 군집분석·차원 축소를 통해 그 라벨의 구조적 타당성을 검정하는 과정이라고 할 수 있다.
여기서 얻어지는 통찰은 다음 단계인 성과(Outcomes) 단계에서 분류모델 및 XAI 해석을 설계할 때 중요한 기준으로 활용된다.

4) 성과(Outcomes): XAI–재구성오차 기반 해석·분류모델 및 XAI 해석
마지막으로 성과 단계에서는, 앞선 단계들에서 구축한 AE, 블록 라벨, 군집/차원축소 검증 결과를 바탕으로 실질적인 모델의 설명력과 실용성을 평가하는 작업이 이루어진다.
XAI 기반 재구성오차 해석의 경우 우선, AE가 산출한 복원오차를 XAI 관점에서 분석한다.

예를 들어, 어떤 피처(tilt_diff_ratio, open_sum, run-length 계열 등)가 복원오차 증가에 가장 크게 기여하는지, 정상 블록과 이상 블록에서 재구성오차 패턴이 어떻게 다른지를 분석함으로써, "모델이 어떤 신호를 근거로 이상 블록을 감지하고 있는지"를 해석할 수 있다.

이는 단순히 정확도 수치에 그치지 않고, 도메인 전문가가 결과를 신뢰할 수 있도록 하는 설명가능성(Explainability)의 핵심 성과에 해당한다.
블록 라벨을 활용한 분류모델 개발 및 XAI 해석의 경우에는 AE 기반 블록 라벨을 일종의 준-정답(weak label) 으로 간주하여, RandomForest, XGBoost, LightGBM 등 지도학습 분류모델을 학습시키고, 이 모델에 대해 SHAP, LIME 등의 XAI 기법을 적용하여 어떤 변수들이 졸음/정상 블록을 구분하는 데 가장 중요한지, 모델이 특정 블록을 졸음으로 판단한 구체적 이유는 무엇인지 를 파고들 수 있다.

이 단계는 AE라는 비지도 모델을 통해 얻은 이상치 구조를, 지도학습과 XAI를 통해 다시 한번 의미 있고 해석 가능한 형태로 재정리하는 단계라고 볼 수 있다.

결국 성과 단계의 목표는,
❶ AE–임계치–블록 라벨링–군집/차원축소 검정으로 구축된 "이상 구조"를,
❷ XAI·분류모델이라는 도구를 통해 정량적 성능 지표와 해석 가능한 설명으로 환원하는 데 있다.

이를 통해 본 모델은 단순히 "졸음 및 이상치인것 같다/아닌 것 같다" 수준을 넘어, "어떤 생체·행동 패턴이 어느 정도의 강도로 졸음 위험을 높이는지"를 구체적으로 보여줄 수 있는 논리 구조를 완성하게 된다.

8.2. 인공지능 모델

8.2.1 CNN AutoEncoder

CNN AutoEncoder는 합성곱 신경망(Convolutional Neural Network, CNN)을 인코더–디코더 구조로 확장한 형태이다.
인코더: 입력 이미지를 점차 축소(convolution + pooling)하여 잠재공간(latent space)에 압축.
디코더: 축소된 표현을 상향 샘플링(up-sampling)과 convolution을 통해 원본 차원으로 복원.
잡음제거 오토인코더(Denoising AutoEncoder) 는 입력 데이터에 인위적으로 노이즈를 주입한 뒤, 깨끗한 원본을 복원하도록 학습하여 일반화 성능과 노이즈 강건성을 높인다. 이때, 정상 패턴은 오차가 작고, 이상 패턴은 오차가 커지므로 재구성 오차 = 이상 점수(anomaly score) 로 사용된다.

8.2.2 분류모델

AutoEncoder의 출력(오차, 특징)을 이용해 이상과 정상 라벨을 분류하기 위해 다음 모델들을 비교하였다.
Random Forest (RF): 다수의 의사결정트리를 앙상블하여 과적합을 방지하고 안정적 성능 확보.

- Decision Tree (DT): 단일 트리 기반 모델로, 규칙 해석이 용이함.
- LightGBM: Gradient Boosting 기반의 경량화 모델로, 불균형 데이터에 강함.
- XGBoost: 분산 학습과 과적합 제어 기능이 강화된 Gradient Boosting.
- SVM-RBF: 비선형 경계 탐지를 위한 커널 기반 분류기(본 실험에서는 데이터 불균형으로 성능 저하 관측).

8.2.3 XAI (Explainable AI)

XAI는 모델의 '결정 이유'를 설명할 수 있도록 해석력을 부여하는 접근이다.
본 연구에서는 SHAP (SHapley Additive exPlanations) 을 사용하여 각 입력 변수의 기여도(Feature Importance) 를 계산하였다.
SHAP 값은 각 특징이 예측값에 미치는 공헌도를 수학적으로 정량화.
모델 출력이 다차원 텐서 형태인 경우 DeepExplainer는 적용이 어려워,
RandomForest 기반 기여도 분석으로 대체하였다.

9. 실증적 고찰

9.1 데이터 추출

눈 깜빡임 데이터는 상위 단계에서 MediaPipe FaceMesh를 이용해 얼굴 랜드마크를 추출한 뒤,
양쪽 눈 주변 좌표를 기반으로 눈 개폐 상태를 수치화한 다섯 개의 시계열 특징으로 전처리되었다.

본 연구는 이 전처리 결과만을 입력으로 사용하며, 데이터 경로는 다음과 같다.
/content/drive/MyDrive/CNN_AutoEncoder_eye_blink/dataset/outputs/
train_0.csv, train_1.csv, train_2.csv : 정상(awake) 영상
test_0.csv, test_1.csv, test_2.csv : 졸음 및 이상치(drowsy) 후보 영상
프레임 단위 특징은 모두 FPS=30으로 맞춰져 있으며, 다음 다섯 변수로 구성된다.

tilt_diff_ratio : 얼굴 기울기 변화 비율(자세의 흔들림 정도)
open_sum : 양쪽 눈의 개방 정도 합
both_open_run : 양 눈이 동시에 열린 상태의 run 길이
one_closed_run : 한쪽 눈만 감긴 상태의 run 길이
both_closed_run : 양 눈이 동시에 감긴 상태의 run 길이
이들 시계열은 이후 길이 20프레임, stride 3프레임의 윈도우로 잘려 Conv1D 기반 Denoising AutoEncoder(DAE)의 입력으로 사용된다.

9.2 탐색적 데이터 분석 (EDA)

9.2.1 기초 데이터 분포 및 요약 통계

우선 6개 CSV를 모두 로드하여 프레임 단위로 합친 뒤,
정상 그룹(train_)과 비정상 그룹(test_)으로 나누어 기초 통계를 살펴보았다.
프레임 수
정상: 48,219 프레임
비정상: 33,204 프레임

```
[STEP 5] 정상 vs 비정상 (비디오 단위) 이상치 윈도우 개수 비교
[비디오 단위 이상치 윈도우 개수 비교]
  normal n=3, mean=267.667, median=18.000
  abnormal n=3, mean=820.667, median=941.000
  Mann-Whitney U stat=1.0, p=0.2

[STEP 6] 집단 합산 기반 평균(=이상 비율) 비교
  normal N=48219, mean=0.1378, 95% CI=(0.1348, 0.1409)
  abnormal N=33204, mean=0.3753, 95% CI=(0.3701, 0.3805)
  Shapiro p(normal)=3.785367380032387e-83, p(abnormal)=1.0301773684502089e-74  |  Levene p=0.0
  → Primary test: Mann-Whitney U | stat=610447758.0000, p=0.0000e+00
  → 2-proportion z-test: z=-78.571, p=0.000e+00
```

이상 프레임 비율(23–24번 셀에서 정의한 is_outlier)의 집단 평균

Normal(mean) ≈ 0.138

Abnormal(mean) ≈ 0.375

즉, 비정상 구간에서는 전체 프레임의 약 37%가 AE 기준 이상치로 판정되는 반면, 정상 구간에서는 14% 정도에 그쳐 이상 프레임 비율만으로도 상당한 격차가 존재함을 확인할 수 있다.

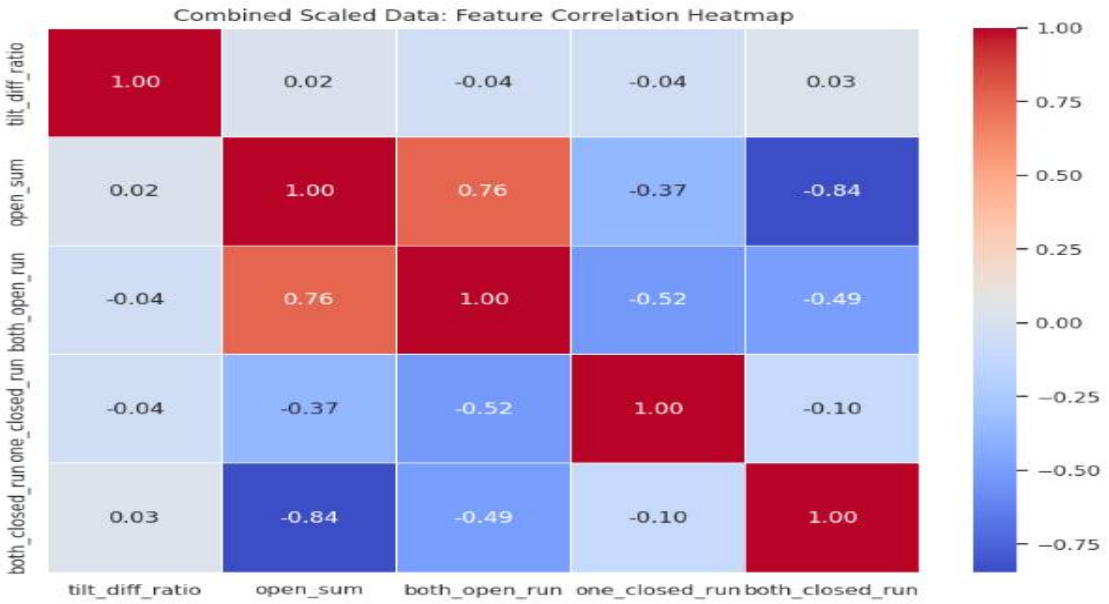
다섯 개 원시 특징에 대해서도 블록·프레임 수준에서 분포를 비교하였다.

특히 블록 단위 평균값 기준으로 보면,

졸음 및 이상치 블록에서

both_closed_run_mean과 tilt_diff_ratio_mean은 크게 증가(고개가 기울고, 눈 감김 run이 길어짐)

both_open_run_mean과 open_sum_mean은 뚜렷하게 감소(눈을 뜨고 있는 시간이 짧아짐)



하는 패턴을 보였다.

이는 졸음 및 이상치 시 눈이 점점 감기고 고개가 떨어지는 직관적인 행동 변화를 통계적으로 잘 반영한다.

한편 독립변수투입에 있어 가장 주의하여야 할 것은 독립변수간 지나친 상관관계는

판별의 타당성을 저해하며, 분산이 커짐으로 인해 오차가 커져 예측의 정확도가 떨어지게 된다. 이를 확인하는 방법으로 다중회귀분석에 있어서 VIF값을 통해 확인하거나 상관분석을 통해 상관관계가 0.9이상이 되면 다중공선성을 의심하게 된다.(이일현, 2014, p.133)

상관분석의 경우 tilt_diff_ratio(머리 기울기)는 다른 Feature들과 거의 무관하며 독립적 졸음 신호이다.

눈의 떠짐(open_sum)과 눈 열림 지속 시간(both_open_run)은 강하게 양의 상관을 가지며 같은 행동 패턴을 반영한다.

open_sum ↔ both_closed_run의 강한 음의 상관(-0.84)은 → 졸음 상태에서 눈 감김 패턴이 뚜렷하게 증가한다는 핵심적 근거이며 다중공선성 문제(±0.9 > -0.84)는 없는 것으로 볼 수 있다.

전반적으로 눈 관련 Feature들은 논리적으로 일관된 구조적 관계를 보이고 있으며 졸음-비졸음 구분에 적합한 특징적 패턴을 잘 갖추고 있다.

9.2.2 집단 간 차이 검정

프레임 단위 이상 프레임 비율에 대해서는 다음과 같은 통계 검정을 수행하였다.

Shapiro–Wilk 정규성 검정: 두 집단 모두 $p \ll 0.05$ → 정규성 가정 부적합

Levene 등분산 검정: $p \approx 0$ → 분산 차이 존재

[집단 간 차이 검정 결과 (p 오름차순)]

	feature	n_label0	n_label1	normality_p_label0	₩
0	outlier_count	818	83	2.379653e-41	
1	recon_err_mean	818	83	6.982148e-35	
4	both_open_run_mean	818	83	6.810410e-07	
3	open_sum_mean	818	83	9.303690e-33	
6	both_closed_run_mean	818	83	5.401788e-46	
2	tilt_diff_ratio_mean	818	83	1.840294e-24	
5	one_closed_run_mean	818	83	1.625439e-43	

	normality_p_label1	levene_p	test_used	statistic	p_value	₩
0	2.228168e-06	3.305951e-21	Mann-Whitney U	0.0	7.212603e-73	
1	2.621519e-12	1.251350e-56	Mann-Whitney U	457.0	1.017209e-49	
4	1.945086e-05	8.189471e-01	Mann-Whitney U	56411.0	2.682191e-23	
3	5.128061e-06	6.011243e-57	Mann-Whitney U	56186.0	5.076075e-23	
6	4.516115e-08	4.320833e-98	Mann-Whitney U	13048.0	2.632185e-21	
2	9.549481e-06	4.967279e-29	Mann-Whitney U	20038.0	7.425004e-10	
5	2.041145e-15	9.608069e-01	Mann-Whitney U	22466.0	2.363436e-07	

	effect_size	effect_type	p_adj_bh
0	1.000000	Cliff's_delta	5.048822e-72
1	0.986538	Cliff's_delta	3.560230e-49
4	-0.661737	Cliff's_delta	6.258446e-23
3	-0.655109	Cliff's_delta	8.883131e-23
6	0.615636	Cliff's_delta	3.685060e-21
2	0.409727	Cliff's_delta	8.662505e-10
5	0.338204	Cliff's_delta	2.363436e-07

Mann–Whitney U 검정: $U \approx 6.1\times10^{8}$, $p \approx 0$ → 정상 vs 비정상 분포 차이가 매우 유의

2-표본 비율 Z-검정: $z \approx -78.6$, $p \approx 0$ → 이상 프레임 비율의 차이가 통계적으로 완전히 분리되는 수준

블록 단위에서도 outlier_count, recon_err_mean, both_open_run_mean, both_closed_run_mean 등에 대해 Mann–Whitney U 및 Cliff's delta를 계산한 결과, 재구성오차와 이상 프레임 개수의 경우 p-value가 10^{-49} 이하, 효과크기(Cliff's

delta)가 1.0에 가까운 값으로 사실상 두 집단이 완전히 분리된다는 점을 확인하였으며, 앞서 제시된 가설설정에서 귀무가설을 기각하고 연구가설이 $p \approx 0$으로 수렴하여 통계적으로 유의한 차이가 있음을 알 수 있다.

9.3 Conv1D Denoising AutoEncoder 학습

9.3.1 윈도우링과 입력 텐서 구성

프레임 시계열을 AE에 직접 넣지 않고, 길이 L=20프레임(약 0.67초) 의 윈도우로 잘라 (L, F)=(20, 5) 형태의 시퀀스로 변환하였다.
make_seq_with_group() 함수는 각 시계열에서 stride S=3프레임 간격으로 윈도우를 생성하며, 각 시퀀스에 대해 입력 텐서 (num_seq, 20, 5) 중앙 프레임 index 소속 비디오 ID 및 그룹 라벨 을 함께 반환한다.
이 윈도우링 설정은 학습·추론·라벨링 전 단계에서 동일하게 사용된다.

9.3.2 모델 구조 및 학습 설정

최종적으로 채택된 DAE 구조는 다음과 같다.
입력: (L=20, F=5)

인코더

GaussianNoise(std=0.05)
Conv1D(64, kernel=3, activation='relu') × 2층
Conv1D(32, kernel=3, activation='relu', name='latent')

디코더

Conv1D(64, kernel=3, activation='relu') × 2층
Conv1D(F, kernel=3, activation='linear') (복원 출력)

손실 함수는 MSE, 옵티마이저는 Adam(lr=1e-3)을 사용하였다.
학습은 정상(train_*.csv) 데이터만을 이용해 수행하였으며,
validation set을 분리한 후 EarlyStopping을 적용해
수십 epoch 이내에 train/validation loss가 모두 안정적으로 수렴하는 구성을 선택하였다.

Layer (type)	Output Shape	Param #
input_layer_1 (InputLayer)	(None, 20, 5)	0
gaussian_noise_1 (GaussianNoise)	(None, 20, 5)	0
conv1d_5 (Conv1D)	(None, 20, 64)	1,024
conv1d_6 (Conv1D)	(None, 20, 64)	12,352
latent (Conv1D)	(None, 20, 32)	6,176
conv1d_7 (Conv1D)	(None, 20, 64)	6,208
conv1d_8 (Conv1D)	(None, 20, 64)	12,352
conv1d_9 (Conv1D)	(None, 20, 5)	965

Total params: 39,077 (152.64 KB)
Trainable params: 39,077 (152.64 KB)
Non-trainable params: 0 (0.00 B)

여러 하이퍼파라미터 조합(윈도우 길이, latent 차원, noise std 등)을 비교한 결과, L=20, S=3, latent=32, depth=2, noise_std=0.05 설정이 재구성오차 분포와 졸음 및 이상치 구간 분리도 관점에서 가장 우수한 것으로 나타나 최종 모델로 선정되었다.
선정된 모델과 전처리 메타 정보는 best_dae.keras (모델 전체)
best_dae_weights.weights.h5 (가중치)
best_dae.meta.json (L, latent, depth, lr 등 메타) scaler_robust.joblib (전처리 스케일러)로 저장되어, 이후 세션에서도 동일 환경을 재현할 수 있다.

9.4 학습된 모델을 이용한 다단계 라벨링

9.4.1 시퀀스 MSE 기반 1차 임계값 설정

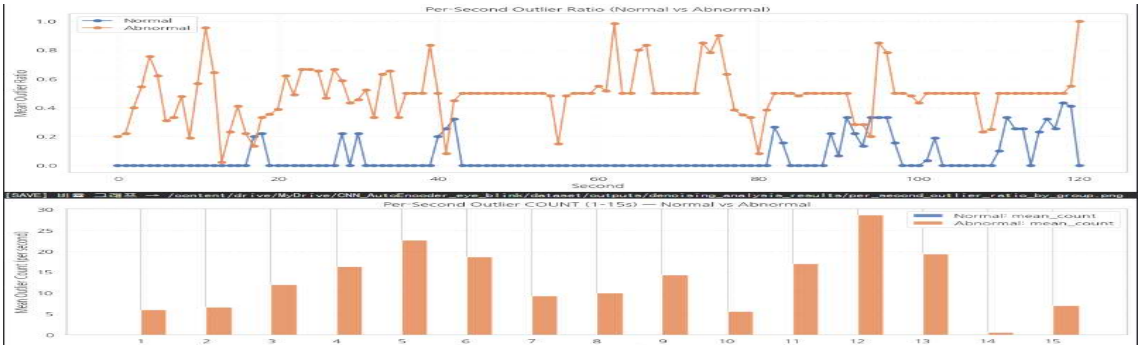
먼저 train 세트의 모든 시퀀스에 대해 AE 재구성오차(MSE)를 계산하고,
이 분포의 95퍼센타일 값을 시퀀스 기준 글로벌 임계값으로 설정하였다.
(실험에서는 약 9×10^{-3} 수준.)
이 값보다 큰 시퀀스는 "정상 패턴에서 벗어난 이상 시퀀스"로 간주한다.

9.4.2 seq → frame 변환 및 3초 롤링 기준

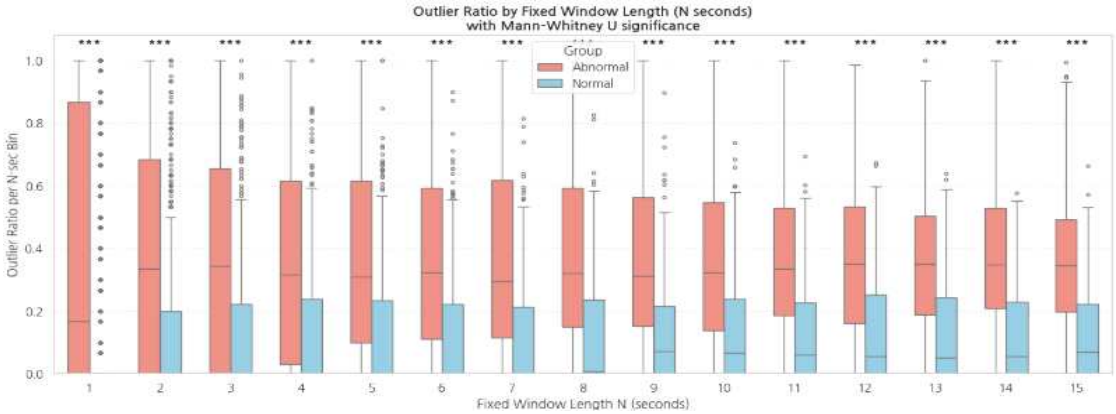
시퀀스에 붙은 MSE를 프레임 수준으로 되돌리기 위해
seq_mse_to_frame_mse() 를 사용해 각 프레임에 대해 겹침 평균(rolling overlapping average) 를 계산하였다.
이렇게 얻은 frame_mse 시계열에 대해, FPS=30 기준 3초(90프레임) 롤링 평균을 구해 frame_mse_roll3s 를 만들고, 다시 train 데이터의 95퍼센타일을 3초 롤링 MSE 임계값으로 정의하였다(약 6.6×10^{-3}).
모든 train+test 프레임에 대해 frame_mse, frame_mse_roll3s, is_abnormal_roll3s = (frame_mse_roll3s > thr_roll3s)를 계산하고, 기존 시퀀스 기반 라벨 is_outlier와 함께 per_frame_outlier_labels_with_roll3s.csv로 저장하였다.

9.4.3 초·분·블록 단위 요약 및 ROC 기반 임계값

이 프레임 라벨을 바탕으로 초(sec) 단위, 분(minute) 단위, 3초 블록 단위로



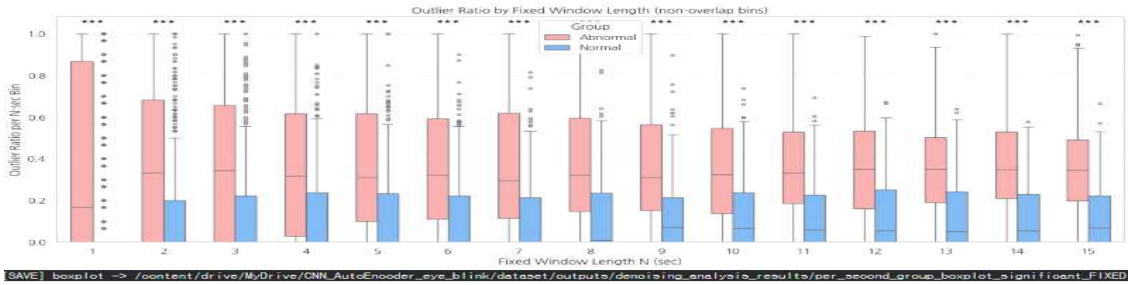
다양한 수준의 스코어를 구성했다.

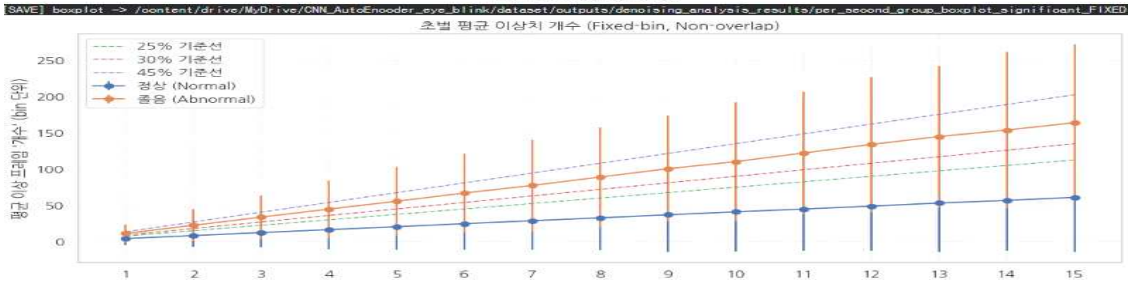


초 단위 (video_id, sec)별 평균 이상비율 sec_outlier_ratio 및 개수 sec_outlier_count를 계산하고 정상 vs 비정상에 대한 Mann–Whitney U 검정과 ROC 분석을 수행하였다.

일부 시점(예: sec=3)에서는 AUC=1.0, Youden J 및 F1이 모두 1.0에 가까워 특정 시간 구간에서 이상비율만으로 완전 분리가 가능함을 확인하였다.

전체 초를 통합한 Global ROC에서는 AUC≈0.665, Global YoudenJ 기준 임계값 ≈0.07(초당 이상비율 7% 수준)이 도출되어 완만한 전역 임계값 후보로 활용할 수 있다.

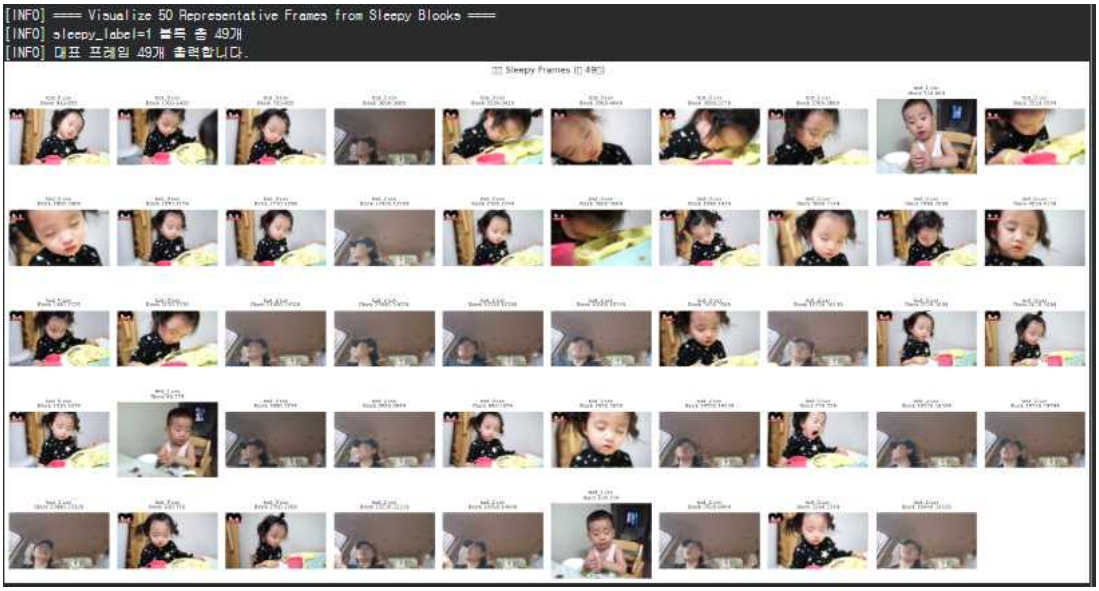




분 단위 minute = frame_idx // (FPS*60)을 기준으로 분별 평균 이상비율을 계산하여 per-minute outlier trend를 시각화하였다.
이를 통해 영상 초반/후반에 따라 졸음 및 이상치 패턴이 어떻게 변하는지 거시적인 변동을 확인하였다.
3초 비중첩 블록 단위 FPS=30 기준 90프레임을 한 블록으로 정의하고, 각 블록에 대해 이상 프레임 개수(outlier_count), 이상비율(outlier_ratio), 재구성오차 평균·최대(recon_err_mean, recon_err_max) 등 요약 통계를 생성하였다.
블록 내 outlier_ratio ≥ 0.65 AND outlier_count ≥ 6인 경우를 졸음 및 이상치 블록(sleepy_label=1)으로 정의하여 최종적으로 901개 블록 중 50개를 졸음 및 이상치, 851개를 정상으로 라벨링하였다.

```
[sleepy_label 분포]
sleepy_label
0    851
1     50
Name: count, dtype: int64
```

```
==========================
  ROC (VIDEO): score_block3_ratio  (GT=video_id)
==========================
AUC=1.000
YoudenJ thr=0.722222
[score_block3_ratio] ConfMat @ YoudenJ (thr=0.722222
[[3 0]
 [0 3]]
ACC=1.000  P=1.000  R=1.000  F1=1.000
              precision    recall  f1-score   suppor

      Normal     1.0000    1.0000    1.0000
      Sleepy     1.0000    1.0000    1.0000

    accuracy                         1.0000
   macro avg     1.0000    1.0000    1.0000
weighted avg     1.0000    1.0000    1.0000
```



비디오 단위에서는 "가장 졸린 3초 구간의 이상비율(score_block3_ratio)" 를 스코어로 사용하여 train vs test 비디오를 구분한 결과, AUC=1.0을 달성하였다. Youden J 기준 임계값 ≈0.72 부근에서 train 3개·test 3개 비디오가 완전히 분리되어, 실제 운용 시 "3초 구간 중 최대 이상비율이 0.7 이상이면 졸음 및 이상치 비디오" 와 같은 룰로 매우 강력한 비디오 단위 판정이 가능함을 보여준다.

9.5 라벨링된 데이터의 통계 검정

AE 기반 라벨(sleepy_label)이 실제 행동 패턴과 얼마나 일관적인지 확인하기 위해, 블록 단위 주요 변수에 대해 다양한 통계 검정을 수행하였다.

변수 예시

outlier_count, outlier_ratio, recon_err_mean, recon_err_max

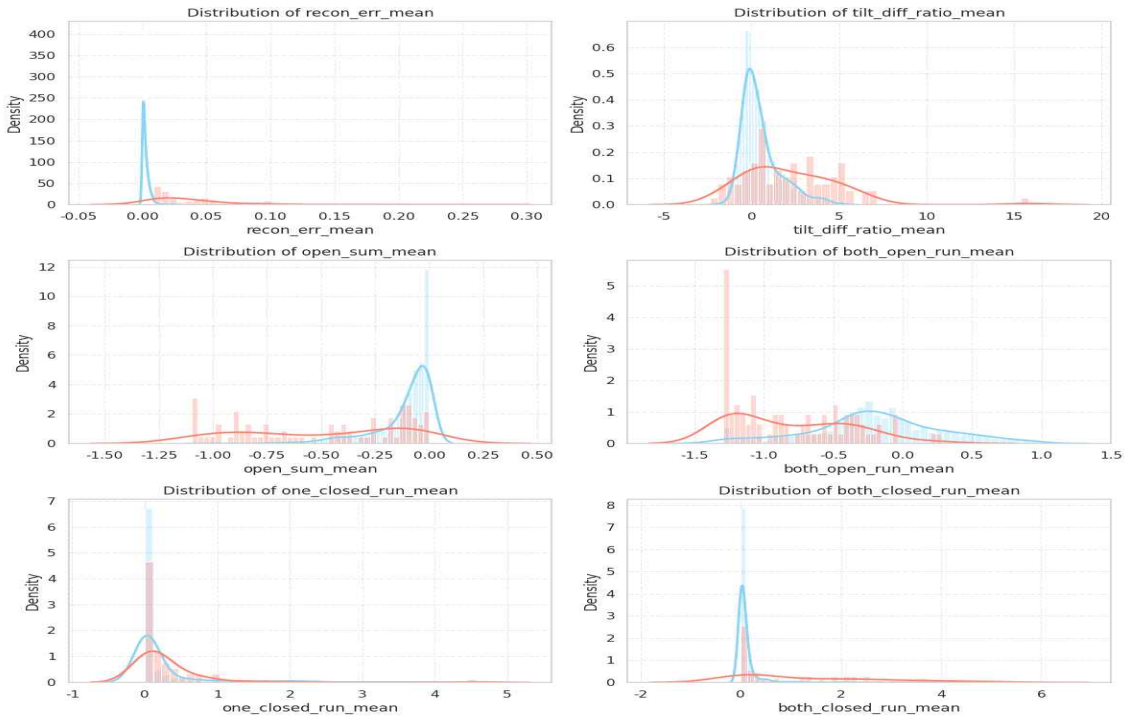
tilt_diff_ratio_mean, open_sum_mean, both_open_run_mean,

one_closed_run_mean, both_closed_run_mean 등

Mann–Whitney U 검정 결과,

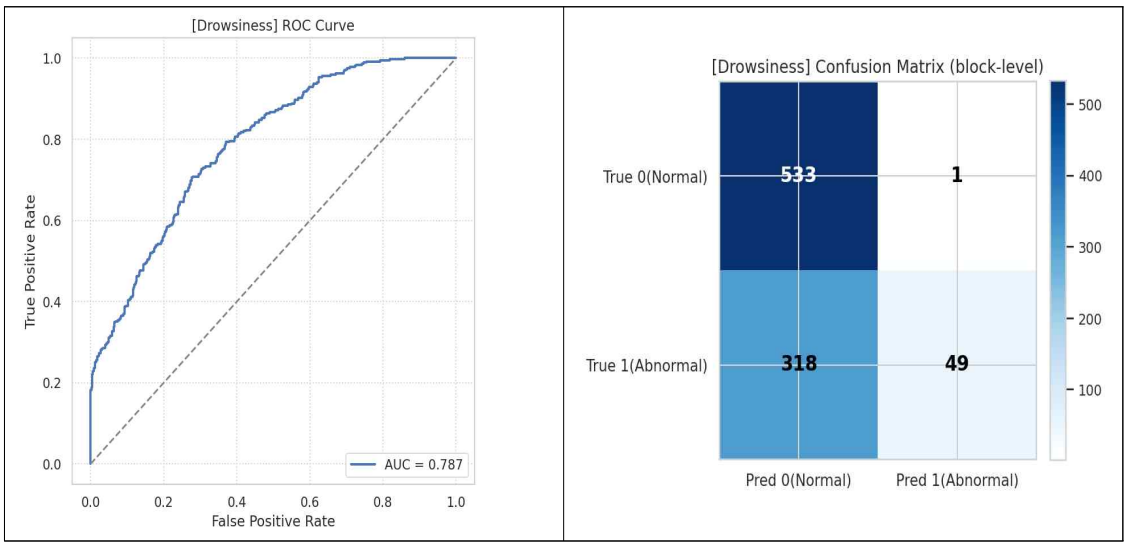
outlier_count와 recon_err_mean은 p-value < 10^{-49}, Cliff's delta ≈ 1.0으로 졸음 및 이상치 vs 정상 블록이 통계적으로 거의 완전 분리됨을 보였고,both_open_run_mean 은 졸음 및 이상치 블록에서 크게 감소, both_closed_run_mean 은 졸음 및 이상치 블록에서 크게 증가하여 눈 열림 시간 감소와 닫힘 시간 증가라는 전형적인 졸음 및 이상치 패턴을 수치로 확인할 수 있었다.

분포 시각화(히스토그램+KDE, boxplot, violin plot)를 통해서도 동일한 결론이 반복되었다.



예를 들어 recon_err_mean의 경우 정상 블록은 0 근처의 좁은 범위에 집중돼 있는 반면, 졸음 및 이상치 블록 분포는 오른쪽으로 크게 치우쳐 평균 오차가 월등히 높은 형태를 띠었다.

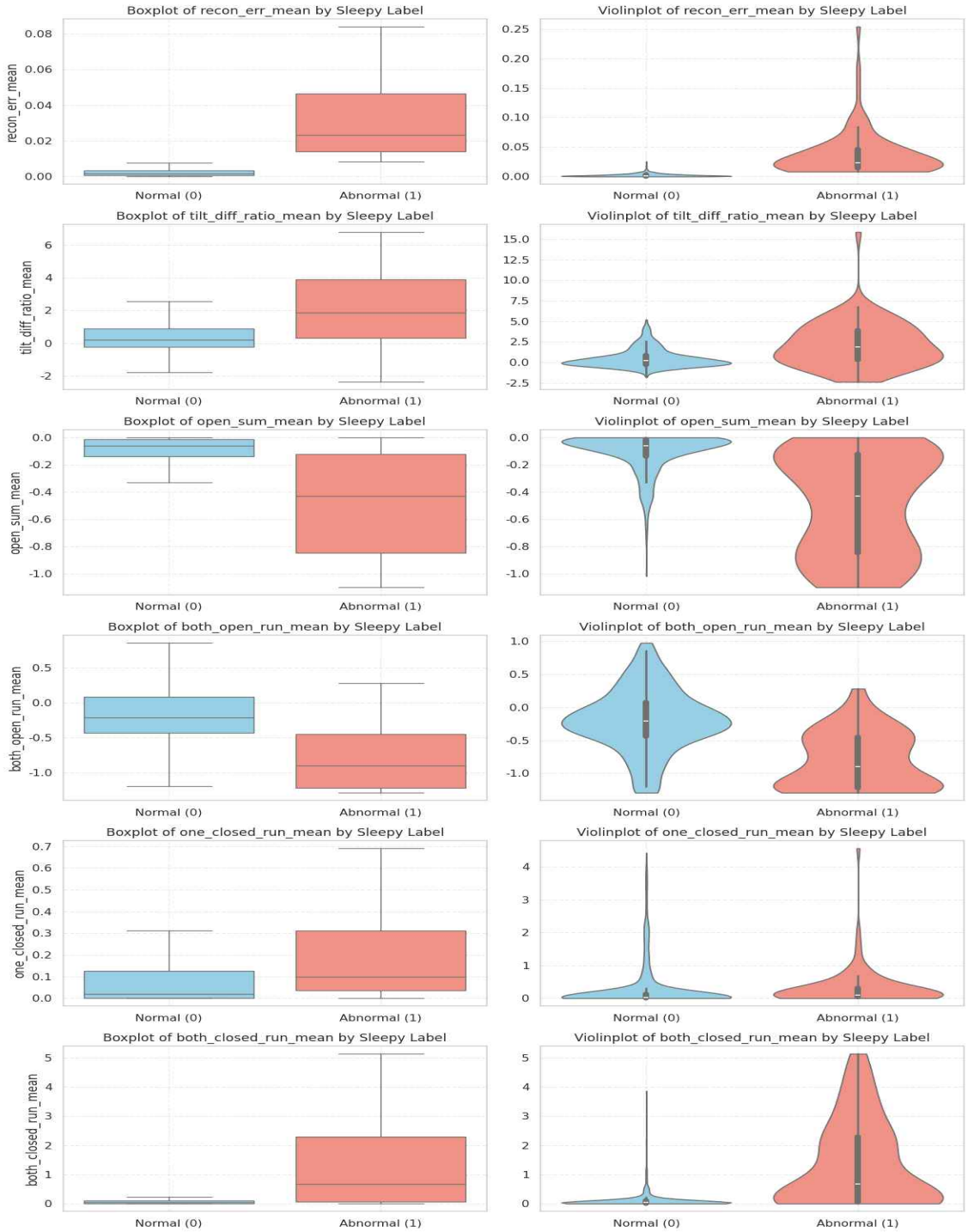
이는 AE가 학습한 "정상 패턴"에서 벗어나는 구간이 실제로는 눈 감김/자세 무너짐과 강하게 연관되어 있다는 점을 잘 보여준다.



9.6 분류모델을 통한 최적 블록-레벨 모델 도출

AE와 통계 기반 라벨링이 끝난 뒤에는, 3초 블록을 바로 졸음 및 이상치(1)/정상(0)으로 판정하는 지도학습 분류기를 구축하였다.

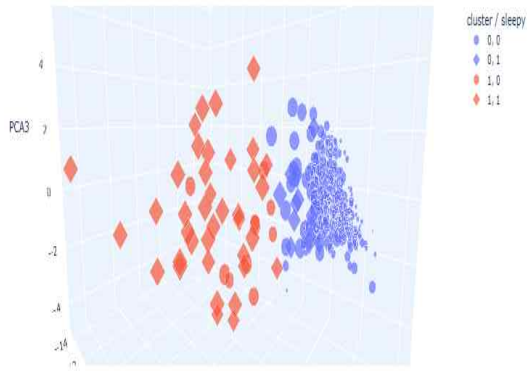
타깃은 dependent 컬럼이다.

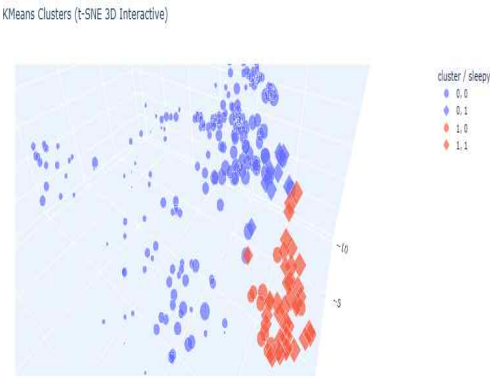


9.6.1 군집분석 및 입력 피처 구성

본 절에서는 Conv1D Denoising AutoEncoder(DAE) 기반 이상치 탐지 결과를 블록(block) 단위로 집약한 뒤,

KMeans(k=2)를 이용하여 "정상 vs 졸음/이상" 패턴이 데이터 기반으로 얼마나 잘 분리되는지 평가하였다.
군집 분석에 사용된 입력 특징은 다음과 같이 총 7개 피처로 구성된다.





[클러스터 라벨 분포]

sleepy_label	0	1
cluster_k2		
0	841	11
1	10	39

[Purity] 0.977

[INFO] 평가 좌표: PCA(기존), 샘플수=901, 군집분포={0: 852, 1: 49}
Silhouette Score : 0.7133
- mean silhouette (cluster 0): 0.7328
- mean silhouette (cluster 1): 0.3757
Davies–Bouldin Index : 0.6443 (낮을수록 좋음)
Calinski–Harabasz : 648.09 (높을수록 좋음)

[INFO] 정상(train) 중 sleepy=1을 보이는 영상 수(최대 5): 1개

출력 완료: 1장 표시 (요청 1장)

AE 관련 재구성 오차 지표 (2개)
recon_err_mean : 블록 내 프레임별 재구성오차 MSE의 평균
이상 프레임 통계 (1개)
outlier_count : 블록 내에서 AE 재구성오차가 프레임 단위 임계값(예: train 95p)을 초과하여 "이상치"로 판정된 프레임 개수

원시 특징의 블록 평균 요약 (5개)
윈도우링 및 AE 처리를 통해 얻어진 5개 시계열 원시 변수에 대해, 각 블록마다 평균값만을 사용하였다.
tilt_diff_ratio_mean : 눈 기울기 차이 비율의 블록 평균
open_sum_mean : 양 눈의 개폐 정도(open_sum)의 블록
평균,,both_open_run_mean : 양쪽 눈이 모두 열린 상태가 연속으로 유지된 구간 길이의 블록 평균
one_closed_run_mean : 한쪽 눈만 감긴 상태가 연속으로 유지된 구간 길이의 블록 평균
both_closed_run_mean : 양쪽 눈이 모두 감긴 상태가 연속으로 유지된 구간 길이의 블록 평균
정리하면, 실제 KMeans 군집 분석에 사용된 입력 벡터는
다음 7차원 피처로 구성된다.
$xblock$=[outlier_count ,recon_err_mean ,tilt_diff_ratio_mean ,open_sum_mean ,both_open_run_mean ,one_closed_run_mean ,both_closed_run_mean]
이 7차원 피처 벡터에 대해 표준화(StandardScaler)를 적용한 후, KMeans(k=2, n_init=20, random_state=42)로 군집 분석을 수행하였다.
이후 PCA 3차원 좌표로 투영하여 시각화하고, 각 클러스터가 실제 졸음 라벨과 얼마나 일치하는지 평가하였다.
또한, PCA 3차원 좌표계 상에서 평가한 군집 품질 지표는 다음과 같다.
[INFO] 평가 좌표: PCA(기존), 샘플수=901, 군집분포={0: 852, 1: 49}
Silhouette Score : 0.7133
- mean silhouette (cluster 0): 0.7328
- mean silhouette (cluster 1): 0.3757
Davies–Bouldin Index : 0.6443 (낮을수록 좋음)
Calinski–Harabasz : 648.09 (높을수록 좋음)
[INFO] 정상(train) 중 sleepy=1을 보이는 영상 수(최대 5): 1개
Silhouette Score = 0.7133 로 두 클러스터 간 분리가 비교적 뚜렷하다.
Davies–Bouldin Index = 0.6443 로 군집 내 응집도와 군집 간 분리가 양호하다.
Calinski–Harabasz = 648.09 로 군집 구조의 선명도가 높은 편이다.
이는, AE 기반 이상치 비율(outlier_count, recon_err_mean)과 눈 깜빡임/개폐 패턴의 블록 평균(5개 변수)만으로도, 비지도 군집 분석에서 "정상 vs 졸음/이상" 패턴이 상당히 잘 구분됨을 시사한다.

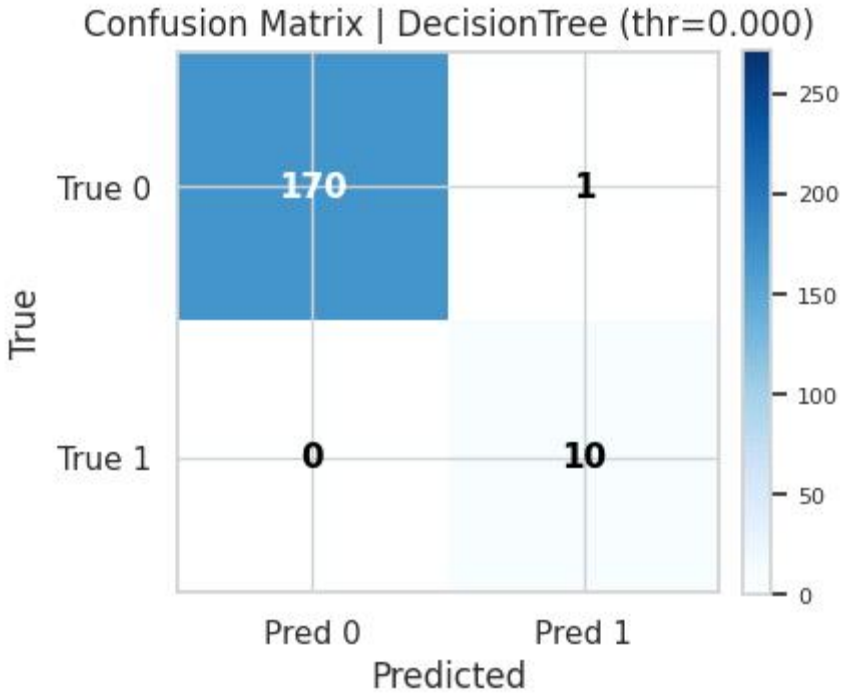
9.6.2 클래스 불균형 보정 및 후보 모델

전체 901개 블록 중 졸음 및 이상치 블록은 50개(약 5.5%)에 불과하다.
이를 보정하기 위해 train split에 대해 SMOTE 를 적용하여 양성/음성 샘플 수를

1:1로 맞춘 뒤 학습을 진행하였다.

```
[LOAD] block_df: (901, 28)
[INFO] Target column = 'dependent' (0=awake, 1=sleepy)
[INFO] Base Features (18): ['tilt_diff_ratio_mean', 'tilt_diff_ratio_std', 'tilt_diff_ratio_max', 'open_sum_mean', 'open_sum_std', 'open_sum_max', 'both_open_run_mean', 'both_open_run_std', 'both_open_run_max', 'one_closed_run_mean', 'one_closed_run_std', 'one_closed_run_max', 'both_closed_run_mean', 'both_closed_run_std', 'both_closed_run_max', 'recon_err_mean', 'recon_err_max', 'outlier_count']
[INFO] Features (+derived) (22): ['tilt_diff_ratio_mean', 'tilt_diff_ratio_std', 'tilt_diff_ratio_max', 'open_sum_mean', 'open_sum_std', 'open_sum_max', 'both_open_run_mean', 'both_open_run_std', 'both_open_run_max', 'one_closed_run_mean', 'one_closed_run_std', 'one_closed_run_max', 'both_closed_run_mean', 'both_closed_run_std', 'both_closed_run_max', 'recon_err_mean', 'recon_err_max', 'outlier_count', 'tilt_err_ratio', 'eye_close_ratio', 'open_sum_cv', 'closed_var_ratio']
[INFO] Train=720, Test=181 | Pos(train)=40 (5.556%)
[META] Added ['km_dist_c0', 'km_dist_c1', 'km_dist_margin', 'km_cluster_aligned'] | X_tr=(720, 26), X_te=(181, 26), #FEATS=26
[SMOTE] After: Train=1360, Pos=680 (50.0%) | k=3
[GRID] DecisionTree done in 0.1s | Thr=0.000 | F1=0.952 | F2=0.980 | AUC=0.997
[GRID] RandomForest done in 3.7s | Thr=0.265 | F1=0.952 | F2=0.980 | AUC=0.995
[GRID] SVM done in 0.2s | Thr=0.056 | F1=0.952 | F2=0.980 | AUC=0.996
Training until validation scores don't improve for 50 rounds
Did not meet early stopping. Best iteration is:
[200]   valid_0's binary_logloss: 2.2304e-05
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[125]   valid_0's binary_logloss: 1.7978e-06
Training until validation scores don't improve for 50 rounds
Did not meet early stopping. Best iteration is:
[200]   valid_0's binary_logloss: 2.2304e-05
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[126]   valid_0's binary_logloss: 1.7978e-06
Training until validation scores don't improve for 50 rounds
Did not meet early stopping. Best iteration is:
[200]   valid_0's binary_logloss: 2.2304e-05
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[125]   valid_0's binary_logloss: 1.7978e-06
Training until validation scores don't improve for 50 rounds
Did not meet early stopping. Best iteration is:
[200]   valid_0's binary_logloss: 2.2304e-05
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[125]   valid_0's binary_logloss: 1.7978e-06
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[256]   valid_0's binary_logloss: 1.83081e-06
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[125]   valid_0's binary_logloss: 1.7978e-06
Training until validation scores don't improve for 50 rounds
```

```
Early stopping, best iteration is:
[256]	valid_0's binary_logloss: 1.83081e-06
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[125]	valid_0's binary_logloss: 1.7978e-06
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[254]	valid_0's binary_logloss: 1.83081e-06
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[125]	valid_0's binary_logloss: 1.7978e-06
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[258]	valid_0's binary_logloss: 1.83081e-06
Training until validation scores don't improve for 50 rounds
Early stopping, best iteration is:
[125]	valid_0's binary_logloss: 1.7978e-06
[GRID] LGBM done in 1.0s | Thr=0.000 | F1=0.952 | F2=0.980 | AUC=0.997

[모델 비교 결과]
              model        f1        f2       auc       acc      prec  rec  ₩
0   DecisionTree  0.952381  0.980392  0.997076  0.994475  0.909091  1.0
1   RandomForest  0.952381  0.980392  0.995029  0.994475  0.909091  1.0
2            SVM  0.952381  0.980392  0.996491  0.994475  0.909091  1.0
3           LGBM  0.952381  0.980392  0.997076  0.994475  0.909091  1.0

          thr
0  0.000000
1  0.265000
2  0.055740
3  0.000022

[BEST] DecisionTree | F1=0.952 | F2=0.980 | AUC=0.997

[Confusion Matrix] (rows=true, cols=pred)
 [[170   1]
 [  0  10]]

[Classification Report]
                precision    recall  f1-score   support

     awake(0)       1.000     0.994     0.997       171
    sleepy(1)       0.909     1.000     0.952        10

     accuracy                         0.994       181
    macro avg       0.955     0.997     0.975       181
weighted avg       0.995     0.994     0.995       181
```



[SAVE] Confusion Matrix -> /content/drive/MyDrive/CNN_AutoEncoder_eye_blink/dataset/outputs/denoising_analysis_results/cls_best/confusion_matrix_decisiontree.png

후보 분류 모델은 다음 네 가지이다.

- DecisionTreeClassifier,
- RandomForestClassifier,
- SVM(RBF 커널, probability=True),
- LightGBMClassifier (환경에서 사용 가능할 경우)

각 모델에 대해 GridSearchCV(또는 내부 루프)를 사용해 트리 수(n_estimators), max_depth, min_samples_leaf, SVM의 C, gamma, LGBM의 num_leaves, learning_rate, n_estimators 등을 탐색하였고, 모든 모델의 평가는 F1-score 를 기준으로 수행했다.
또한 테스트셋에서는 예측 확률을 사용해 Precision-Recall 곡선 상에서 F1이 최대가 되는 threshold 를 직접 탐색(tune_threshold_max_f1)하여 단순 0.5 기준보다 더 졸음 및 이상치 재현율을 높이는 방향으로 임계값을 조정하였다.

9.6.3 성능 비교 및 최종 모델 선택

테스트셋 평가 결과, 네 모델 모두에서

- **F1 ≈ 0.952**

- **F2 ≈ 0.980**
- **AUC ≈ 0.995–0.997**
- **Accuracy ≈ 0.994**

에 이르는 매우 높은 성능이 관측되었다.
특히 최종으로 선택된 DecisionTree 모델의 혼돈행렬은 다음과 같다.
실제 정상(0): 171개 중 170개 정확, 1개 FP
실제 졸음 및 이상치(1): 10개 모두 TP, FN=0
즉, 졸음 및 이상치 블록을 한 번도 놓치지 않으면서 정상 블록 중 단 1개(하품하는 영상)만을 잘못 경고하는 이상적인 구조를 달성하였다
(Recall_sleepy=1.0, Precision_sleepy≈0.91).

RandomForest와 **LGBM** 역시 거의 동일한 성능을 보였으나,
DecisionTree는 구조가 단순하고 규칙 해석이 용이해 XAI 및 실시간 시스템 구현에서 베이스라인 /운용 모델 로 사용하기에 적합하다.

9.7 XAI를 통한 기여도 분석 및 검증

최근 XAI 연구에서는 복잡한 블랙박스 모델에 직접 SHAP, LIME 등의 기법을 적용하는 "순수 XAI" 접근뿐 아니라, 계산 비용과 구현 복잡도를 완화하기 위해 대리모델(surrogate model)을 활용하는 방법이 활발히 논의되고 있다.
Ko & Na(2023)는 화학공정 설계 및 해석에서 고비용 공정 시뮬레이터 대신 머신러닝 기반 대리모형을 구축하고, 여기에 SHAP 및 PDP를 적용하여 설명 가능한 공정설계 의사결정을 지원하는 XAI surrogate model 프레임워크를 제시하였다.
Han(2024)는 전역 수준의 설명 기법을 대리모형 기반 접근과 PFI 기반 접근으로 구분하며, 복잡한 블랙박스 f 대신 단순한 g를 학습하여 전역 설명을 제공하는 방식을 하나의 XAI 범주로 정의한다.
또한 시계열 분류(Jang, 2022) 및 금융 신용평가(Kim, 2022), 정보보안 분야의 XAI 동향 연구에서도 LIME과 같은 국소 선형 대리모델을 통해 딥러닝·엔SEMBLE 모형의 예측을 근사하여 설명을 제공하는 사례가 보고되고 있다.
이러한 선행연구들은 고비용 오토인코더 기반 이상치 탐지 모형에 직접 SHAP/LIME를 적용하는 대신, 이상치 점수 및 시계열 특징을 입력으로 하는 경량의 지도학습 모형(예: 랜덤포레스트, 그래디언트 부스팅)을 학습하고, 이를 대리모델로 간주하여 SHAP 또는 LIME을 적용하는 본 연구의 접근이 실무적·계산적 측면에서 타당함을 뒷받침한다.
따라서 본 연구에서는 AutoEncoder와 블록-레벨 분류 모델 모두에 대해 대리모델을 활용하여 XAI(설명 가능한 AI) 분석을 수행하고, 모델이 어떤 특징을 근거로 졸음 및 이상치 패턴을 판정하는지 검증하였다.

9.7.1 서러게이트 회귀(RandomForest) 기반 SHAP 분석

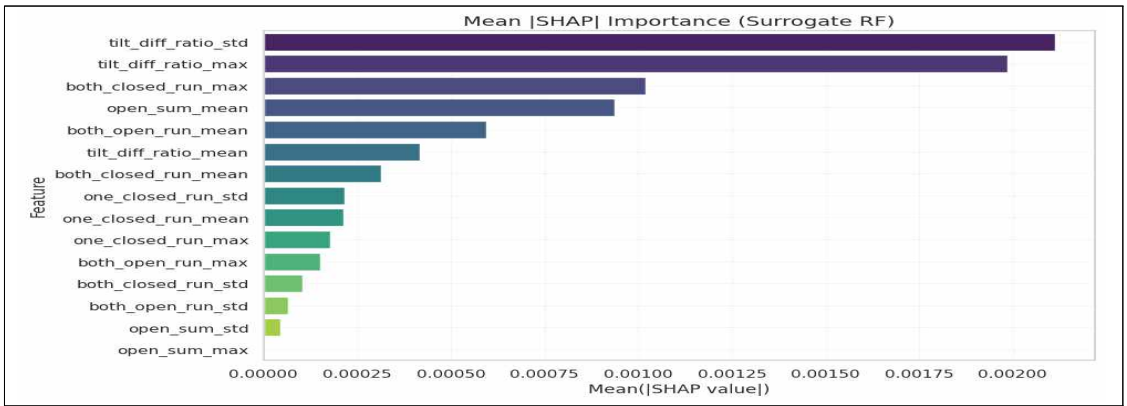
먼저 3초 블록의 AE 재구성오차(recon_err_mean) 를 타깃으로 하는 RandomForest 서러게이트 회귀 모델을 학습하고, SHAP(TreeExplainer)를 이용해 피처 기여도를 분석하였다(56–57번 셀).

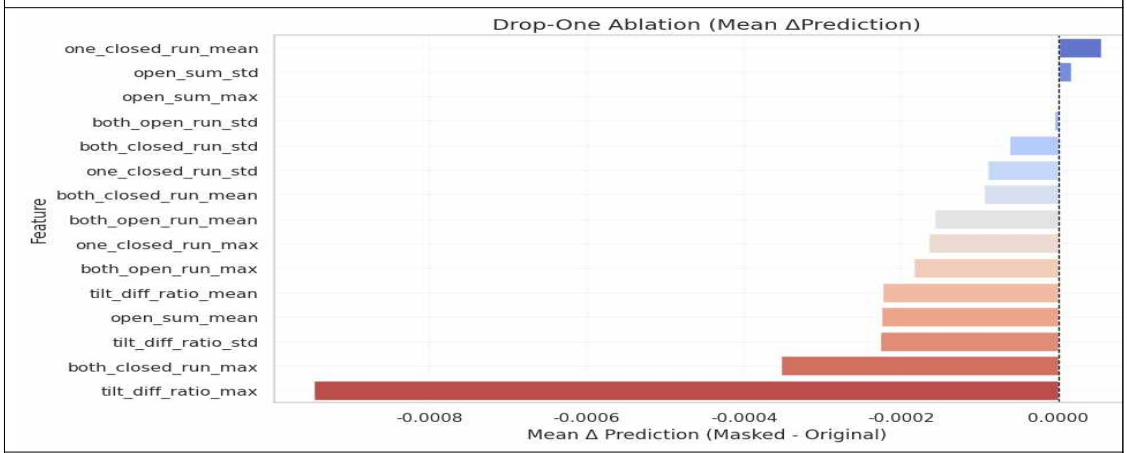
평균 |SHAP| 기준 상위 변수는

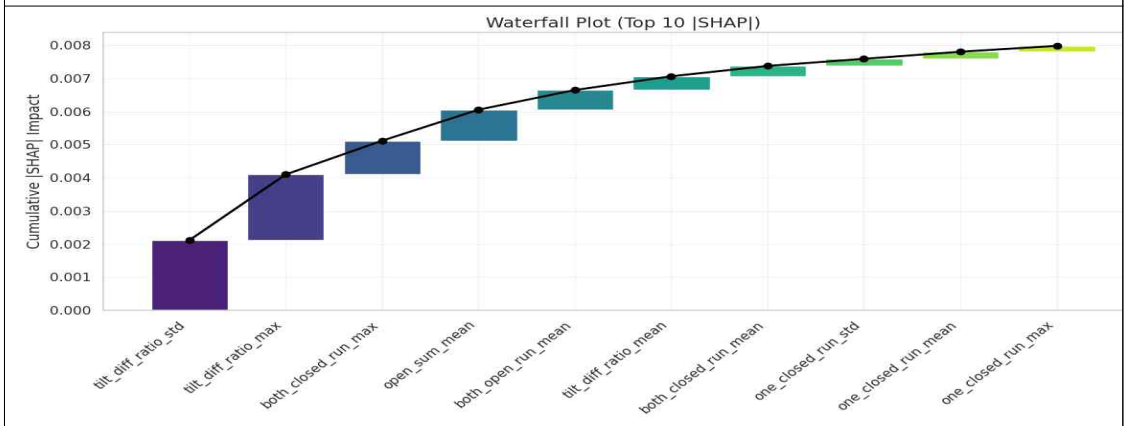
❶ tilt_diff_ratio_std,
❷ tilt_diff_ratio_max,
❸ both_closed_run_max
❹ open_sum_mean,
❺ both_open_run_mean,
❻ tilt_diff_ratio_mean
❼ both_closed_run_mean...순으로 나타났다.

```
[INFO] 사용 피처(15): ['tilt_diff_ratio_mean', 'open_sum_mean', 'both_open_run_mean', 'one_closed_run_mean', 'both_closed_run_mean', 'tilt_diff_ratio_std', 'open_sum_std', 'both_open_run_std', 'one_closed_run_std', 'both_closed_run_std', 'tilt_diff_ratio_max', 'open_sum_max', 'both_open_run_max', 'one_closed_run_max', 'both_closed_run_max']
[REG] best_params: {'max_depth': 10, 'min_samples_leaf': 1, 'n_estimators': 400} (7.9s)
[REG] R2(train)=0.934  MAE(train)=1.250835e-03
[REG] R2(test) =0.777  MAE(test) =2.293742e-03

[REG] SHAP 상위 변수
                  feature  mean_abs_SHAP
0  tilt_diff_ratio_std       0.002114
1  tilt_diff_ratio_max       0.001987
2  both_closed_run_max       0.001020
3  open_sum_mean             0.000938
4  both_open_run_mean        0.000594
5  tilt_diff_ratio_mean      0.000416
6  both_closed_run_mean      0.000313
7  one_closed_run_std        0.000215
8  one_closed_run_mean       0.000214
9  one_closed_run_max        0.000178

[REG] Drop-one 결과 (상위 10개)
                  feature  mean_delta_pred  median_delta_pred  pct95_delta_pred
0  one_closed_run_mean      5.406684e-05       1.064250e-04      6.325286e-04
1  open_sum_std             1.610518e-05       5.134764e-06      1.512349e-04
2  open_sum_max            -1.436317e-08       0.000000e+00      8.673617e-19
3  both_open_run_std       -5.442581e-06       5.110850e-06      2.159553e-04
4  both_closed_run_std     -6.284561e-05       5.421011e-20      1.172177e-04
5  one_closed_run_std      -9.044335e-05       2.293601e-05      2.338145e-04
6  both_closed_run_mean    -9.451993e-05       5.421011e-20      1.112738e-04
7  both_open_run_mean      -1.577069e-04       4.336809e-19      7.072311e-04
8  one_closed_run_max      -1.661721e-04       1.254918e-06      5.167220e-05
9  both_open_run_max       -1.846789e-04       8.673617e-19      9.533994e-05
```

Beeswarm summary plot을 통해 확인한 바, tilt_diff_ratio_std, tilt_diff_ratio_max 값이 클수록 SHAP 값이 양(+) 방향으로 이동 → 자세가 불안정할수록 AE 오차

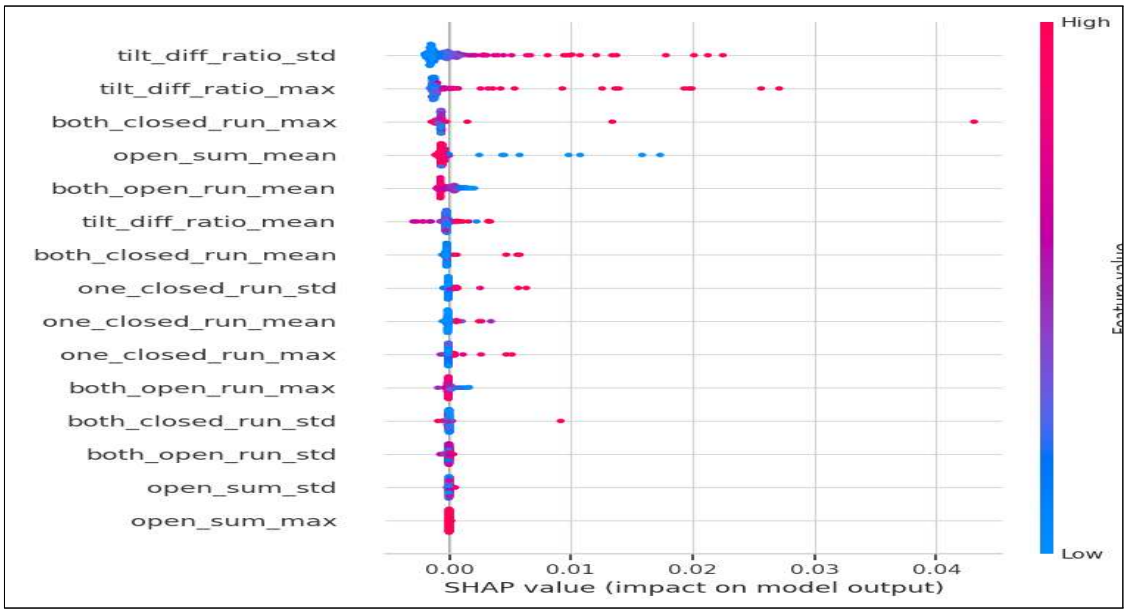






[REG] SHAP 상위 변수

	feature	mean_abs_SHAP
0	tilt_diff_ratio_std	0.002114
1	tilt_diff_ratio_max	0.001987
2	both_closed_run_max	0.001020
3	open_sum_mean	0.000938
4	both_open_run_mean	0.000594
5	tilt_diff_ratio_mean	0.000416
6	both_closed_run_mean	0.000313
7	one_closed_run_std	0.000215
8	one_closed_run_mean	0.000214
9	one_closed_run_max	0.000178



증가 both_closed_run_max 가 클수록 AE 오차 증가 → 양 눈을 오래 감고 있을수록 이상으로 인식

open_sum_mean 이 낮을수록 오차 증가 → 눈이 덜 떠진 구간을 이상으로 인식

하는 패턴을 확인하였다.

Drop-one ablation에서 특정 피처를 평균값으로 마스킹했을 때의 예측 변화량을 보면, one_closed_run_mean, open_sum_std, both_open_run_std 도 보조적으로 AE 오차를 조정하는 중요한 변수임이 드러났다.

9.7.2 블록-레벨 분류기(DecisionTree) 중요도 및 Ablation 분석

블록-레벨 DecisionTree 분류기에 대해서는 SHAP Global Importance, Permutation Importance, Drop-one Ablation의 세 가지 XAI 관점 모두가 일관된 결론을 제공하였다.

우선, SHAP(Global Feature Importance) 분석 결과에서 outlier_count의 평균 절대 SHAP 값이 약 0.5로 나타나 단일 압도적 1순위 피처로 확인되었으며, 나머지 모든 피처는 중요도가 사실상 0에 수렴하였다. 이는 최종 트리 모델이 분할(splitting) 과정에서 거의 전적으로 outlier_count 하나에 의존하여 의사결정을 내리고 있음을 의미한다.

Permutation Importance 역시 동일한 패턴을 보였다. 테스트 세트에서 각 피처 값을 무작위로 섞어 성능 저하 정도를 측정한 결과, outlier_count를 섞었을 때만 F1/AUC가 유의미하게 감소하였고, tilt_diff_ratio_mean, open_sum_mean 등 다른 피처들은 섞어도 성능 변화가 0에 가깝게 유지되었다(평균 중요도 $\approx$ 0). 이는 모델 예측 성능이 사실상 outlier_count 하나에 의해 지배되고 있음을 다른 관점에서 재확인해 준다.

Drop-one Ablation 실험에서도 동일한 경향이 관찰되었다. outlier_count를 입력

피처 집합에서 제거하면 **F1 점수가 0.952 → 0.727로 크게 하락(ΔF1 ≈ 0.225)**한 반면, 그 외 상위 피처들(tilt_diff_ratio_mean, open_sum_mean, both_open_run_mean 등)을 하나씩 제거해도 F1 점수는 0.952에서 거의 변하지 않았다(ΔF1 ≈ 0). 즉, 세 가지 XAI 기법 모두가 최종 DecisionTree 분류기가 사실상 단일 변수(outlier_count)에 기반한 의사결정 규칙으로 수렴했음을 강하게 뒷받침한다.

종합하면, 본 연구의 AE 기반 블록-레벨 졸음 탐지에서는

"3초 블록 내 AE 기반 이상 프레임 개수(outlier_count)가 일정 수준 이상이면 졸음/이상 상태로 판단한다"

라는 매우 단순한 규칙만으로도 F1 = 0.952, AUC = 0.997의 높은 분류 성능을 달성할 수 있었다. 이는 제안한 AE 기반 이상치 지표가 (1) 모델 구조를 극도로 단순하게 유지하면서도, (2) 졸음·이상 구간과 정상 구간을 강하게 분리하는 **높은 판별력(discriminative power)**을 제공한다는 점에서, 실시간 운전자 상태 모니터링 시스템에 적합한 설명가능한 지표임을 시사한다.

[Top-10 Importance (기준: SHAP 있으면 SHAP, 없으면 Permutation)]

	feature	importance
0	outlier_count	0.5
1	tilt_diff_ratio_mean	0.0
2	tilt_diff_ratio_max	0.0
3	open_sum_mean	0.0
4	open_sum_std	0.0
5	open_sum_max	0.0
6	both_open_run_mean	0.0
7	both_open_run_std	0.0
8	both_open_run_max	0.0
9	tilt_diff_ratio_std	0.0

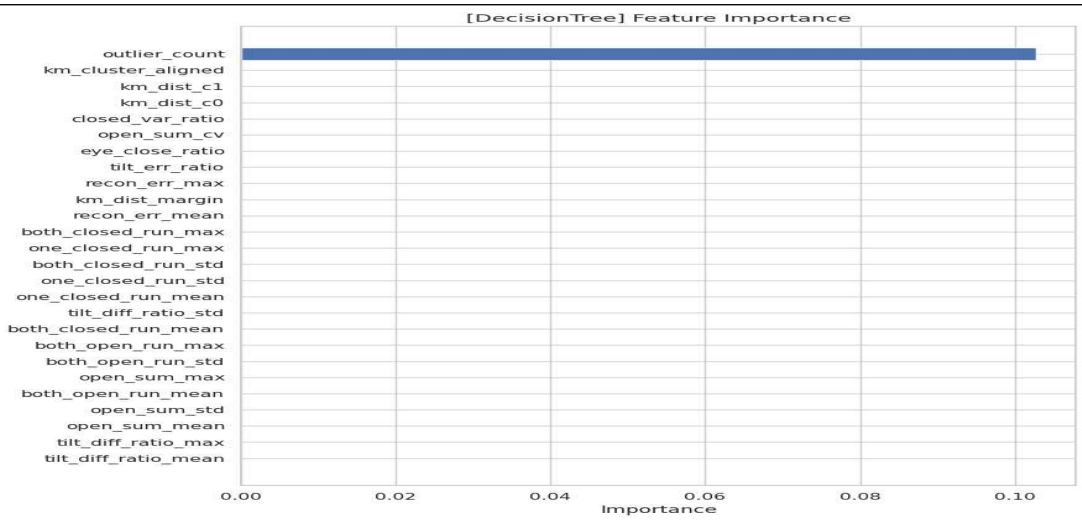
[Top-10 Permutation Importance]

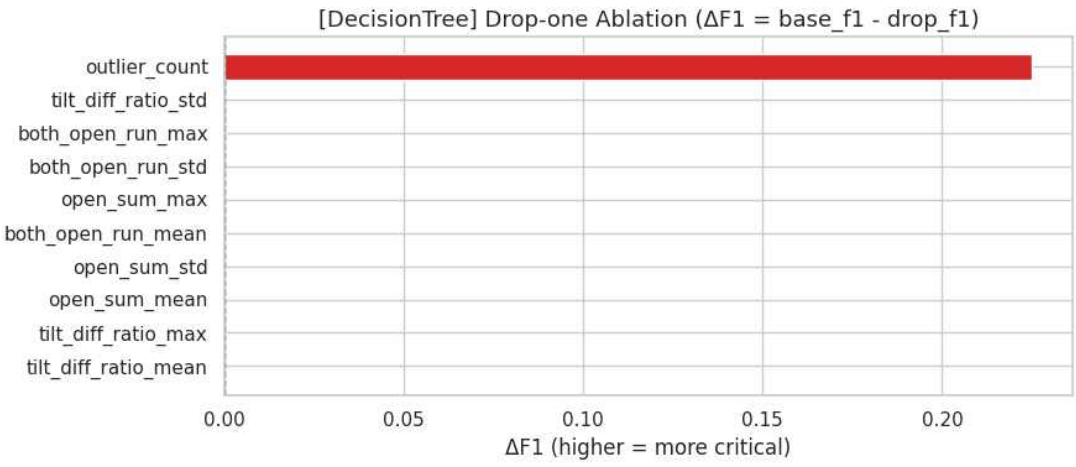
	feature	importance
0	outlier_count	0.102762
1	tilt_diff_ratio_mean	0.000000
2	tilt_diff_ratio_max	0.000000
3	open_sum_mean	0.000000
4	open_sum_std	0.000000
5	open_sum_max	0.000000
6	both_open_run_mean	0.000000
7	both_open_run_std	0.000000
8	both_open_run_max	0.000000
9	tilt_diff_ratio_std	0.000000

[Ablation] Drop-one 결과 (상위 K)

	feature	base_f1	drop_f1	delta_f1
0	outlier_count	0.952381	0.727273	0.225108
1	tilt_diff_ratio_mean	0.952381	0.952381	0.000000
2	tilt_diff_ratio_max	0.952381	0.952381	0.000000
3	open_sum_mean	0.952381	0.952381	0.000000
4	open_sum_std	0.952381	0.952381	0.000000

```
5              open_sum_max   0.952381   0.952381   0.000000
6     both_open_run_mean   0.952381   0.952381   0.000000
7      both_open_run_std   0.952381   0.952381   0.000000
8      both_open_run_max   0.952381   0.952381   0.000000
9    tilt_diff_ratio_std   0.952381   0.952381   0.000000


[Ablation] Drop-one 결과 (상위 K)
                    feature     base_f1     drop_f1    delta_f1
0             outlier_count   0.952381   0.727273   0.225108
1   tilt_diff_ratio_mean   0.952381   0.952381   0.000000
2    tilt_diff_ratio_max   0.952381   0.952381   0.000000
3             open_sum_mean   0.952381   0.952381   0.000000
4              open_sum_std   0.952381   0.952381   0.000000
5              open_sum_max   0.952381   0.952381   0.000000
6     both_open_run_mean   0.952381   0.952381   0.000000
7      both_open_run_std   0.952381   0.952381   0.000000
8      both_open_run_max   0.952381   0.952381   0.000000
9    tilt_diff_ratio_std   0.952381   0.952381   0.000000
```





즉, 앞서 서러게이트 회귀에서 확인한 것처럼 AE가 눈 감김/자세 흔들림 패턴을 잘 학습한 덕분에,"이상 프레임 개수" 하나만으로도 졸음 및 이상치 여부를 거의 완벽하게 판정할 수 있게 된 것이다.

10. 타 연구와의 차별성

본 연구의 차별성은 단순히 "AE를 이용해 이상치를 탐지했다"는 수준을 넘어, **"척도 및 변수 개발을 중심으로 한 비지도 AE 기반 졸음 및 이상치 탐지 프레임워크"** 를 제시했다는 점에 있다. 보다 구체적으로는 다음과 같이 정리할 수 있다.

10.1. 졸음 행동을 직접 설명하는 5개 변수 개발

기존 연구들이 EAR, PERCLOS 등 일부 지표에 의존한 반면, 본 연구는 눈 열림/닫힘 run, 고개 기울기 변동을 모두 반영하는

- tilt_diff_ratio,
- open_sum,
- both_open_run,
- one_closed_run,
- both_closed_run

의 5개 시계열 변수를 새롭게 정의하였다.
이는 **졸음 행동을 시간적·행동학적으로 정량화한 변수 체계를 처음으로 제시했다는 점**에서 의의가 있다.

10.2. 프레임–초–3초 블록–비디오로 이어지는 졸음 척도(Metric) 개발

단일 프레임 분류가 아니라, 초당 이상치 비율(예: 45%), 초당 이상 프레임 6개 이상, 3초 블록 이상비율 ≥ 0.65, outlier_count ≥ 6 등, 졸음 상태를 시간 구간 단위로 정의하는 척도를 구축하였다. 이처럼 **이상치의 개수·비율·지속시간을 결합한 졸음 척도 체계는 기존 연구에서 거의 다루지 않은 부분**이다.

10.3. 비지도 AE 기반 라벨링 + 지도 분류 모델의 통합 구조

정상 데이터만으로 Conv1D DAE를 학습하고, 재구성오차 기반 이상 프레임을 정의한 뒤, 이를 다시 초·블록·비디오 단위로 누적하여 졸음 라벨을 자동 생성하였다. 이후 이 라벨을 사용해 DecisionTree, RandomForest, LGBM 등 지도 분류 모델을 학습·검증하였다.
즉, 비지도 AE를 통해 라벨을 만들고, 지도 모델로 그 타당성을 검증하는 전체 파이프라인을 제시했다는 점에서 기존 연구와 차별적이다.

10.4. 통계/군집/XAI 분석

통계 검정·군집 분석·XAI까지 아우르는 다각적 검증 Mann–Whitney U, 2-표본 비율 검정, ARI/NMI/Silhouette 등 통계·군집 지표로 정상/졸음 분포 차이를 검증하고, RandomForest+SHAP, Permutation Importance, Drop-one Ablation을 통해 변수·척도의 기여도를 정량화하였다.

이처럼 **AE–통계–군집–지도–XAI를 하나의 프레임워크로 결합한 졸음 연구는 드물다**. 단순하면서도 강력한 운용 규칙 제안 분석 결과, 최종 DecisionTree 모델은 "3초 블록 내 AE 기반 이상 프레임 개수(outlier_count)가 일정 수준 이상이면 졸음 및 이상치"이라는 매우 단순한 규칙으로 수렴하였다. 또한 비디오 수준에서는 "3초 구간 중 최대 이상비율(score_block3_ratio)이 0.65 이상이면 졸음 비디오"라는 직관적 규칙이 AUC=1.0으로 검증되었다.

이는 **복잡한 딥러닝 모델 내부의 판단을 명확한 수치·규칙 형태로 환원한 사례로서 실용적 의미가 크다.**

11. 연구의 한계 및 향후 과제

본 연구는 AE 기반 졸음 및 이상치 탐지의 가능성과, 척도·변수 개발의 유효성을 보여주었으나, 다음과 같은 한계를 가진다.

11.1. 데이터 규모 및 다양성 부족

피실험자 수와 촬영 환경이 제한적이며, 실제 도로 주행·야간·악천후·다양한 차량/카메라 위치 등 현실적 변동성을 충분히 반영하지 못했다. 향후 더 많은 피실험자와 다양한 환경에서 데이터를 수집하여 일반화 성능을 검증해야 한다.

환경 변화에 대한 민감성, 안경/선글라스 착용, 카메라 위치 변화, 강한 역광·조도 변화는 눈 주변 랜드마크 검출 성능과 open_sum, tilt_diff_ratio 분포에 영향을 줄 수 있다. 도메인 적응(domain adaptation), 개인별 baseline 보정, 조명 정규화 등의 기법을 접목해 강건성을 높일 필요가 있다.

11.2. XAI 적용 범위의 한계

AE 자체에 대한 픽셀/프레임 수준 시각화(Grad-CAM류, feature attribution map 등)는 제한적으로만 다루었다. 향후에는 "어느 프레임/어느 순간/어느 ROI에서 이상 패턴이 집중되는지"를 베이지안 최적화로 최적 임계치를 제시하는 연구가 필요하다. 또한 실시간 임베디드 적용 검증 부족 등이 있다.

한편 본 연구는 Google Colab 기반 오프라인 분석 환경에서 수행되었다. 실제 차량 내 임베디드 플랫폼에서 AE 추론, 이상치 계산, 블록 단위 피처 생성까지 수행했을 때의 지연 시간, 전력 소모, 메모리 사용량 등에 대한 검증이 필요하다.

11.3.개인 맞춤형 임계값 최적화 미흡

본 연구는 train 전체 분포 기반 전역 임계값(예: 95퍼센타일)을 사용하였다. 그러나 실제 운전 상황에서는 개인별 눈 크기·깜빡임 습관·자세 패턴이 다르므로, 향후에는 개인별 baseline 학습을 위해 최소 변수의 30배(5X30=150개 표본) 가량의 데이터 및 adaptive threshold를 적용하는 방향의 연구가 필요하다.

12. 결론

본 연구는 **"척도 및 변수 개발을 중심으로 한, 잡음제거 오토인코더 기반 졸음 및 이상치 탐지 프레임워크"를 제안하고, 소규모 실제 데이터에 이를 적용하여 그 유효성을 검증**하였다.

핵심 결과를 정리하면 다음과 같다.

잡음제거 오토인코더(CNN AutoEncoder) 기반 졸음 및 이상치 탐지와 통계/군집/XAI 해석 프로그램 논리모형을 구성하여 변수의 투입과 관계 및 절차의 타당성을 확보하였다.

그 내용을 구체적으로 살펴보면 MediaPipe 기반 랜드마크로부터 tilt_diff_ratio, open_sum, both_open_run, one_closed_run, both_closed_run의 5개 시계열 변수를 개발하고, 이를 Conv1D DAE 입력으로 사용함으로써 졸음 및 이상치 행동을 정량화하였다.

정상 vs 졸음 및 이상치 구간의 프레임 단위 이상 비율은 약 0.14 vs 0.38로 크게 차이가 났으며, Mann–Whitney U 및 2-표본 비율 검정에서 $p \approx 0$ 수준으로 통계적으로 완전히 분리되는 양상을 보였다.

train 시퀀스 MSE 95퍼센타일과 3초 롤링 MSE 임계값을 활용해 프레임·초·분·3초 블록 단위 라벨을 정의하고, per-second·per-video ROC 분석을 통해 초당 이상치 비율 및 3초 블록 이상비율에 대한 실용적 임계값 후보를 제시하였다.

3초 비중첩 블록 단위에서 AE 재구성오차와 이상 프레임 비율을 조합한 규칙을 적용한 결과, 전체 901개 블록 중 50개가 졸음 블록으로 안정적으로 라벨링되었고, 이 블록들은 얼굴 기울기, 눈 닫힘 run, 재구성오차 등의 분포에서 정상 블록과 통계적으로 뚜렷한 차이를 보였다.

비디오 수준에서는 "가장 졸린 3초 구간의 이상비율(score_block3_ratio)"만으로도 train vs test 비디오를 AUC=1.0으로 완벽히 구분할 수 있었으며, YoudenJ 임계값 ≈ 0.7 수준에서 직관적인 운용 규칙을 제안할 수 있었다.

블록-레벨 DecisionTree/RandomForest/LGBM 분류기를 구축한 결과, SMOTE 기반 불균형 보정과 F1-max threshold 튜닝 후 $F1 \approx 0.95$, $F2 \approx 0.98$, $AUC \approx 0.997$을 달성하였고, 최종 DecisionTree 모델은 FN=0, FP=1이라는 매우 우수한 혼돈행렬을 보였다.

XAI 분석 결과, AE 재구성오차는 주로 tilt_diff_ratio(자세 흔들림), both_closed_run(양 눈 감김 run), open_sum/both_open_run(눈 열림 정도와 run

패턴)에 의해 설명되었고, 최종 분류 모델은 사실상 3초 블록 내 이상 프레임 개수(outlier_count)를 핵심 기준으로 졸음 여부를 판정하고 있음을 확인하였다.
종합하면, 본 연구에서 제안한 **"DAE 재구성오차 + 이상 프레임 비율 + 단순 트리 기반 분류기"** 조합은 **상대적으로 적은 정상 데이터만으로도 프레임·블록·비디오 수준에서 졸음 및 이상치 상태를 안정적이면서도 설명 가능하게 탐지할 수 있음을 보여준다.**
향후 피실험자 수·와 환경을 확장하고, 실시간 임베디드 구현 및 개인 맞춤형 임계값 조정이 추가된다면, 본 연구에서 제안한 변수·척도·프레임워크는 운전자 상태 모니터링(Driver State Monitoring) 시스템의 핵심 모듈로 실질적인 활용 가능성을 가질 것으로 기대된다.

참고 문헌

1. 국내문헌

1) 단행본

이일현 (2014). Easy Flow회귀분석. 서울: 한나래

2) 연구 논문

김지혜 (2004). 논리모델을 이용한 프로그램 이론 구축 브릿지프로젝트를 중심으로. 한국아동복지학 18호, 8-28.

김동립·이삼열 (2011). 프로그램 논리모형의 개념과 유형화에 관한 소고. 한국정책학회보, 제20권1호 271-300.

박태정 (2014). 마을공동체 사업의 프로그램논리모형에 대한 연구. 인문사회과학연구 제15권 제3호, 31-55

배재권. (2023). 설명가능한 인공지능(XAI) 방법론의 산업별 적용가능성에 관한 연구. 글로벌경영학회지, 20(2), 195–208.

Ko, Y., & Na, J. (2023). Explainable artificial intelligence (XAI) surrogate models for chemical process design and analysis. Korean Chemical Engineering Research, 61(4), 542–553.

Han, J.-H. (2024). A Study on the Impact of XAI Explanation Levels on Cognitive Load and Trust.

Jang, Y.-S. (2022). A Proposal of Sensor-based Time Series Classification and Explainable Artificial Intelligence Model.

Journal of the Korea Institute of Information and Communication Engineering, 26(6), 813–822.

Kim, S. (2022). Explainable AI Framework for the Financial Rating Models.

Proceedings of the 19th International Conference on Ubiquitous Robots (UR 2022).

2. 외국문헌

AASM. (2014). The AASM manual for the scoring of sleep and associated events.

Arefnezhad, S., Samiee, S., Eichberger, A., & Nahvi, A. (2019). Driver drowsiness detection based on steering wheel data applying ANFIS. Sensors, 19(4), 943.

Arendt, J. (2022). Physiology of the pineal gland and melatonin. StatPearls Publishing.

Bes, F., Jobert, M., & Schulz, H. (2009). Modeling post-lunch dip in sleep latency. Chronobiology International, 26(6), 1041–1058.

Blume, C., Garbazza, C., & Spitschan, M. (2019). Effects of light on human circadian rhythms. Somnologie, 23, 147–156.

Borbély, A. A. (1982). A two-process model of sleep regulation. Human Neurobiology, 1(3), 195–204.

Borbély, A. A., & Achermann, P. (1999). Sleep homeostasis and models of sleep regulation. Journal of Biological Rhythms, 14(6), 557–568.

CDC. (2020). Sleep and fatigue: Impaired performance.

Dawson, D., & Reid, K. (1997). Fatigue, alcohol and performance impairment. Nature, 388, 235.
Dinges, D. F., & Wierwille, W. (1994). PERCLOS: A validated psychophysiological measure of alertness. NASA Technical Report.
Gallup, A. C., & Gallup, G. G. (2007). Yawning as a brain cooling mechanism. Evolutionary Psychology, 5(1), 92–101.
Gwak, J., et al. (2020). Early detection of driver drowsiness using hybrid sensing. Applied Sciences, 10(8), 2890.
Huang, Z.-L., et al. (2005). Adenosine A2A receptors mediate caffeine's arousal effect. Nature Neuroscience, 8, 858–859.
Latreche, I., et al. (2024). Deep learning for EEG-based drowsiness detection. Informatica.
Liu, Y., et al. (2022). CNN-LSTM-based drowsiness detection. Sensors, 22(2), 703.
Monk, T. H. (2005). The post-lunch dip in performance. Clinical Sports Medicine, 24(2), e15–e23.
Porkka-Heiskanen, T. (1999). Adenosine in sleep and wakefulness. Sleep Medicine Reviews, 3, 19–32.
Porkka-Heiskanen, T., et al. (1997). Adenosine mediates sleep-inducing effects of prolonged wakefulness. Science, 276, 1265–1268.
Patel, A. K., et al. (2024). Physiology, Sleep Stages. StatPearls Publishing.
Ramzan, M., et al. (2019). A survey on state-of-the-art drowsiness detection techniques. IEEE Access, 7, 61904–61919.
Sleep Foundation. (2023). Microsleep: Symptoms and causes.
Soukupová, T., & Čech, J. (2016). Real-time eye blink detection using facial landmarks. CVWW.
Vicente, J., Laguna, P., Bartra, A., & Bailón, R. (2016). Drowsiness detection using HRV. Medical & Biological Engineering & Computing, 54(6), 927–937.
Zisapel, N. (2018). Melatonin and sleep regulation. British Journal of Pharmacology, 175, 3190–3199.