

Parameter Importance Analysis of a Bio-Signal-Based Crisis Recognition Model Using Explainable Artificial Intelligence

Joon-Yong Kim*

*Professor, Dept. of AI Convergence, Seoul Theological University, Gyeonggi-Do, Korea

[Abstract]

Bio-signal-based crisis recognition models have demonstrated high classification performance in healthcare and safety-related applications; however, their real-world adoption is often limited by the lack of interpretability in model decision-making processes. This study explicitly focuses on explainability rather than performance generalization and proposes an explainable artificial intelligence (XAI)-based analytical framework for investigating parameter importance in a bio-signal-based crisis recognition model. An artificial neural network was constructed using five physiological signals: blood pressure, body temperature, heart rate, electrocardiogram (ECG), and galvanic skin response (GSR). Global feature importance was analyzed using SHAP, while local decision behavior for individual samples was interpreted using LIME. Experimental results show that, despite weak linear correlations among bio-signals, heart rate and ECG consistently play dominant roles in crisis recognition decisions. These findings demonstrate that conventional correlation analysis alone is insufficient to explain nonlinear model behavior and highlight the necessity of XAI techniques for transparent and trustworthy bio-signal-based decision systems.

▶ **Key words:** Bio-signal, Crisis Recognition, Explainable Artificial Intelligence, SHAP, LIME, Feature Importance

[요 약]

생체 신호 기반 위기 인식 모델은 의료 및 안전 분야에서 높은 분류 성능을 보이고 있으나, 모델 내부 의사결정 과정에 대한 해석 가능성이 부족하다는 한계로 인해 실제 적용에는 제약이 존재한다. 본 연구는 성능 일반화보다는 설명 가능성 확보에 초점을 두고, 설명 가능한 인공지능(eXplainable Artificial Intelligence, XAI)을 활용한 생체 신호 기반 위기 인식 모델의 파라미터 중요도 분석 프레임워크를 제안한다. 혈압, 체온, 심박수, 심전도(ECG), 피부전도도(GSR)를 입력으로 하는 인공신경망 모델을 구축하고, SHAP을 통해 전역적 중요도를 분석하였으며 LIME을 활용하여 개별 위기 샘플에 대한 국소적 판단 근거를 해석하였다. 실험 결과, 생체 신호 간 선형 상관관계는 전반적으로 낮게 나타났으나 심박수와 ECG는 위기 인식 결정 과정에서 일관되게 높은 중요도를 보였다. 이는 비선형 모델의 의사결정 구조를 이해하기 위해 XAI 기법이 필수적임을 시사한다.

▶ **주제어:** 생체 신호, 위기 인식, 설명 가능한 인공지능, SHAP, LIME, 파라미터 중요도

- First Author: Joon-Yong Kim, Corresponding Author: Joon-Yong Kim
- *Joon-Yong Kim (musimk@stu.ac.kr), Dept. of AI Convergence, Seoul Theological University
- Received: 2026. 01. 02, Revised: 2026. 01. 29, Accepted: 2026. 02. 20.

I. Introduction

웨어러블 디바이스와 사물인터넷(IoT) 기술의 발전은 생체 신호 데이터를 실시간으로 수집·분석할 수 있는 환경을 제공하고 있다. 이에 따라 생체 신호를 활용한 위기 인식 기술은 의료 모니터링, 산업 안전, 사회 안전망 구축 등 다양한 분야에서 핵심 기술로 부상하고 있다 [1-3].

머신러닝 및 딥러닝 기반 위기 인식 모델은 복합적인 생체 신호 패턴을 효과적으로 학습하여 기존 규칙 기반 시스템 대비 우수한 성능을 보이고 있다 [4]. 그러나 이러한 모델은 블랙박스 특성으로 인해 예측 결과의 근거를 해석하기 어렵다는 한계를 가진다. 이는 의료 및 안전 분야에서 인공지능 시스템의 신뢰성과 책임성을 확보하는 데 중요한 장애 요소로 작용한다 [5].

선행 연구에서는 혈압, 체온, 심박수에 ECG(심전도)와 GSR(피부전도도)을 추가한 인공지능 기반 위기 인식 모델을 제안하고 성능 향상을 확인하였다 [6]. 또한 기존 연구들은 위기 인식 정확도 향상을 목표로 다양한 생체 신호를 통합하는 방향으로 발전해 왔으나, 각 생체 신호가 실제 위기 판단 과정에서 어느 정도 기여하는지에 대한 정량적 분석은 충분히 이루어지지 않았다. 특히 피어슨 상관계수와 같은 전통적인 통계 분석은 생체 신호 간 선형 관계만을 반영할 뿐, 비선형 모델 내부에서의 실제 기여도를 설명하는 데에는 한계가 있다.

본 연구는 이러한 한계를 극복하기 위해 XAI 기법을 활용하여 생체 신호 기반 위기 인식 모델의 내부 의사결정 구조를 분석한다. 본 논문의 목적은 새로운 분류 모델을 제안하거나 성능을 극대화하는 데 있지 않으며, 이미 안정적으로 학습된 ANN 기반 위기 인식 모델을 대상으로 각 생체 신호 파라미터의 상대적 중요도를 전역적 및 국소적 관점에서 해석할 수 있는 분석 프레임워크를 제시하는 데 있다.

또한, 본 연구는 다중 생체 신호를 활용한 위기 인식 모델의 분류 성능을 제시하는 데 목적이 있는 것이 아니라, 이미 안정적으로 학습된 인공지능 모델에 대상으로 XAI를 적용하여 각 생체 신호 파라미터가 위기 인식 결정 과정에 기여하는 방식을 해석하는 사례 연구(case study)를 목적으로 한다.

II. Preliminaries

1. Related works

1.1 Research on Crisis Recognition Based on Biometric Signals

생체 신호를 활용한 위기 인식 연구는 심박수, 체온 등 단일 생체 신호 기반 접근에서 시작하여 ECG, GSR을 포함한 다중 생체 신호 기반 분석으로 확장되어 왔다[2,7-10]. 이러한 연구들은 위기 인식 성능 향상을 보고하였으나, 개별 생체 신호의 상대적 기여도에 대한 해석은 제한적이었다.

1.2 Machine Learning-Based Biometric Signal Analysis

머신러닝 기법은 생체 신호의 복잡한 패턴을 학습하는데 효과적이며, 최근에는 딥러닝 모델이 비선형 관계를 효과적으로 학습할 수 있음이 보고되고 있다 [11,12]. 그러나 딥러닝 모델의 해석 가능성 부족은 의료·안전 분야 적용에 있어 중요한 한계로 지적되고 있다 [5].

1.3 Explainable Artificial Intelligence (XAI)

XAI는 인공지능 모델의 예측 결과를 인간이 이해할 수 있도록 설명하는 기술로, SHAP과 LIME은 각각 전역적(Global) 및 국소적(Local) 관점에서 모델 의사결정을 설명하는 대표적인 사후 해석 기법이다[14-16]. 최근 연구들은 단일 XAI 기법보다는 다중 관점의 해석을 병행 적용할 것을 권장하고 있다[17,18].

또한, XAI 기반 해석 프레임워크는 단순한 분류 성능 향상을 넘어, 의료 인공지능이 실제 임상 환경에서 요구하는 신뢰성과 설명 가능성 확보라는 과제를 해결하는 데 기여할 수 있다[19]. 이는 기존의 고성능 딥러닝 모델이 지적받아 온 블랙박스 문제를 보완하는 방향으로, 향후 의료 인공지능의 실질적 활용 가능성을 확장하는 기반이 된다[20].

최근 5년간 의료 인공지능 분야에서는 SHAP, LIME, Grad-CAM 등 다양한 XAI 기법을 활용하여 모델의 판단 근거를 설명하려는 연구가 활발히 진행되고 있다 [21-24]. 특히 전역적(Global) 해석과 국소적(Local) 해석을 병행 적용하는 접근이 모델 신뢰성 확보에 효과적이라는 점이 보고되고 있다 [23, 24].

본 연구는 이러한 연구 흐름을 바탕으로 SHAP과 LIME을 결합한 다중 관점 XAI 분석 프레임워크를 생체 신호 기반 위기 인식 모델에 적용함으로써, 기존 성능 중심 연구와 차별화된 해석 중심 접근을 제시한다.

III. The Proposed Scheme

1. Research Methodology

1.1 Overall Structure of the Study

본 연구의 전체 구조는 생체 신호 수집 및 전처리, ANN 기반 위기 인식 모델 학습, XAI 기반 해석의 세 단계로 구성된다. 해당 구조는 분류 성능 향상이 아닌, 모델 의사결정 구조 해석을 주요 연구 목표로 설계되었다.

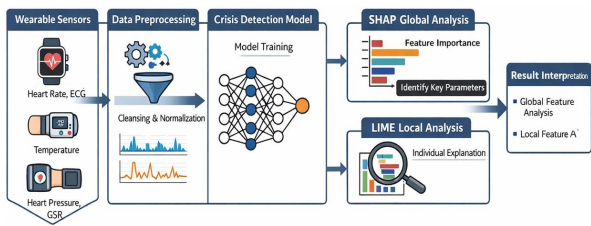


Fig. 1. Overall structure of biosignal-based crisis recognition and XAI application

Fig. 1은 본 연구에서 제안하는 생체 신호 기반 위기 인식 시스템의 전체적인 흐름을 나타낸다.

웨어러블 센서를 통해 심박수, 심전도(ECG), 체온, 혈압, 피부전도도(GSR)와 같은 다양한 생체 신호가 지속적으로 수집된다. 이러한 원시 생체 신호 데이터는 센서 노이즈 제거, 이상치 처리, 정규화 및 시간 동기화와 같은 전처리 과정을 거쳐 분석 가능한 형태로 변환된다.

전처리가 완료된 생체 신호 데이터는 인공신경망 기반 위기 인식 모델의 입력으로 사용된다. 해당 모델은 다수의 생체 신호 간 비선형적인 상호작용을 학습하여 정상 상태와 위기 상태를 분류하도록 설계되었다. 이를 통해 단일 생체 신호에 의존하는 기존 방식보다 복합적인 생리적 변화를 효과적으로 반영할 수 있다.

모델 학습 이후에는 설명 가능한 인공지능 기법을 적용하여 위기 인식 결과의 근거를 분석한다. SHAP과 LIME 기법을 활용하여 각 생체 신호 파라미터가 위기 인식 결정에 기여하는 정도를 정량적으로 평가하며, 분석 결과는 최종적으로 위기 인식 모델의 해석 및 결과 이해 단계로 전달된다. 이러한 구조는 모델의 예측 성능뿐만 아니라 해석 가능성과 신뢰성을 함께 확보하기 위한 것이다.

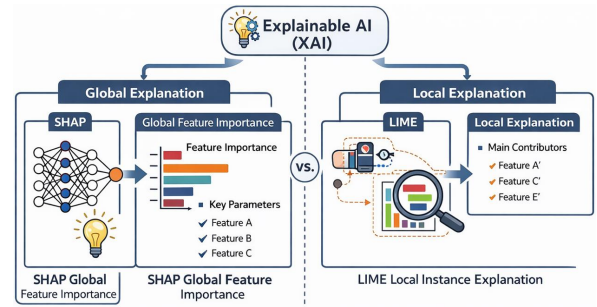


Fig. 2. Explainable AI Architecture Using SHAP and LIME

Fig. 2는 본 연구에서 사용한 XAI 구조를 개념적으로 나타낸 것이다. 해당 구조는 전역적 설명(Global Explanation)과 국소적 설명(Local Explanation)이라는 두 가지 관점으로 구성된다.

전역적 설명 단계에서는 SHAP 기법을 활용하여 전체 데이터셋을 기준으로 각 생체 신호 파라미터의 상대적 중요도를 분석한다. 이를 통해 위기 인식 모델이 전반적으로 어떤 생체 신호에 더 크게 의존하는지를 파악할 수 있으며, 핵심 파라미터를 식별하는 데 활용된다. 이러한 전역적 해석은 모델의 일반적인 동작 특성을 이해하고 센서 선택 및 모델 설계 전략을 수립하는 데 중요한 역할을 한다.

반면, 국소적 설명 단계에서는 LIME 기법을 적용하여 개별 샘플 단위의 위기 인식 결정 과정을 분석한다. LIME은 특정 입력 데이터 주변에서 모델을 근사하여, 해당 샘플에 대해 어떤 생체 신호가 위기 판단에 가장 큰 영향을 미쳤는지를 설명한다. 이를 통해 동일한 위기 상태로 분류된 경우라도 개인별 생체 신호 반응 차이에 따라 주요 기여 파라미터가 달라질 수 있음을 확인할 수 있다.

SHAP 기반 전역적 해석과 LIME 기반 국소적 해석을 결합함으로써, 본 연구의 XAI 구조는 위기 인식 모델을 다층적으로 이해할 수 있도록 한다. 이러한 접근은 모델의 투명성과 신뢰성을 향상시키며, 향후 개인 맞춤형 위기 인식 시스템이나 의료·안전 분야 적용을 위한 중요한 기반을 제공한다.

1.2 Dataset Construction

1.2.1 Experimental Environment

본 연구에서는 통제된 실내 환경에서 성인 피험자 20명(남성 10명, 여성 10명)을 대상으로 생체 신호 데이터를 수집하였다. 수집된 생체 신호는 혈압, 체온, 심박수, ECG, GSR이며, 모든 센서는 1초 단위로 데이터를 수집하였다.

위기 상태는 신체적 스트레스 및 정신적 스트레스 시나리오를 통해 유도되었으며, 정상 상태와 위기 상태로 라벨링되었다. 본 연구는 데이터 규모의 한계를 인지하고 있으

며, 이에 따라 실험 결과는 해석 프레임워크 검증에 위한 사례 분석(case study) 관점에서 해석된다.

1.2.2 Sensors Used and Measurement Parameters

- 혈압: 자동 전자 혈압계
- 체온: 피부 접촉형 온도 센서
- 심박수: 광학 심박 센서(PPG)
- ECG: 단일 채널 웨어러블 ECG 센서
- GSR: 피부 전도도 센서

모든 센서는 1초 단위로 데이터를 수집하였으며, 시간 동기화 과정을 거쳐 통합되었다.

1.2.3 Crisis Scenarios and Labeling

위기 상태는 다음 시나리오를 기반으로 유도되었다.

- 신체적 스트레스: 계단 오르기 및 유산소 운동
- 정신적 스트레스: 제한 시간 인지 과제
- 정상 상태: 안정 휴식 상태

각 구간은 정상/위기 상태로 라벨링되었으며, 반복 실험을 통해 라벨 신뢰성을 확보하였다.

1.2.4 Raw Biometric Signal Data

다음의 Table. 1은 실제 실험에서 수집된 원시 생체 신호 데이터의 일부를 예시적으로 제시한 것이며, 데이터 분포와 라벨링 구조를 보여주기 위해 제시하였으며, 실제 학습에는 동일 형식의 대규모 데이터가 사용되었다.

Table 1. Sample raw bio-signal data before preprocessing

Sample	Lable	BP mmHg	Temp (°C)	HR (bmp)	ECG (mV)	GSR (μS)
S001	Normal	118	36.5	72	0.82	3.1
S002	Normal	121	36.6	75	0.85	3.4
S003	Crisis	136	37.4	104	1.26	6.8
S004	Crisis	142	37.6	110	1.34	7.2
S005	Normal	115	36.4	70	0.79	2.9
S006	Crisis	138	37.2	98	1.18	6.1

1.2.5 Preprocessed Biometric Signal Data

다음의 Table. 2는 원시 데이터를 Min-Max 정규화를 통해 0~1 범위로 변환한 것이다. 해당 데이터 형식이 인공 신경망 입력층에 그대로 사용되었다.

Table 2. Normalized input data after preprocessing

Sample	Lable	BP mmHg	Temp (°C)	HR (bmp)	ECG (mV)	GSR (μS)
S001	Normal	0.42	0.31	0.28	0.33	0.21
S002	Normal	0.45	0.34	0.31	0.36	0.24
S003	Crisis	0.78	0.69	0.82	0.74	0.76
S004	Crisis	0.83	0.72	0.88	0.79	0.81
S005	Normal	0.39	0.28	0.26	0.30	0.19
S006	Crisis	0.80	0.65	0.76	0.69	0.73

1.3 Artificial Neural Network Model Architecture and Parameters

본 연구에서 사용한 인공신경망 모델은 입력층 5개 노드, 은닉층 3개(64-32-16 노드), 출력층 1개 노드로 구성된다. 활성화 함수로는 ReLU를 사용하였으며, 과적합 방지를 위해 드롭아웃(0.3)과 조기 종료(Early Stopping) 기법을 적용하였다.

본 모델 구조는 최적 구조 탐색 결과가 아니라, XAI 기반 분석을 수행하기 위한 기준 모델(reference model)로 설정되었다.

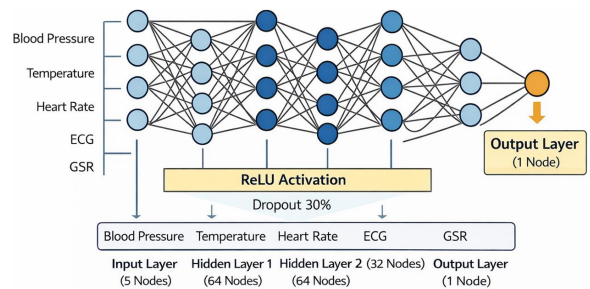


Fig. 3. Architecture of the artificial neural network-based crisis recognition model

다음의 Table. 3은 모델 훈련에 적용된 파라미터 조합이다.

Table 3. Model training hyperparameters

Parameter	Value
Input features	5
Hidden layers	3(64-32-16)
Optimizer	Adam
Learning rate	0.001
Batch size	32
Epochs	Max 100
Droupout rate	0.3
Early stopping	Validation loss
Random seed	42

2. Experience

2.1 Experimental Setup and Evaluation Metrics

전체 데이터셋은 학습 데이터 80%, 검증 데이터 20%로 계층적 분할하였다. 평가 지표로는 정확도와 손실 함수를 사용하였다.

2.2 Model Performance Comparison

Table 4. Performance comparison between models

Model	Accuracy (%)	Loss
Existing model	98.33	0.00695
Enhanced model	99.50	0.00312

실험 결과, 제안한 위기 인식 모델은 정확도 99.50%, 손실 값 0.00312를 기록하였다. 이는 기존 모델의 정확도 98.33%, 손실 값 0.00695 대비 각각 1.17%p의 정확도 향상과 약 55%의 손실 감소를 보인 결과이다. 해당 성능 지표는 모델이 안정적으로 학습되었음을 확인하기 위한 보조적 근거로 활용된다.

2.3 Training Stability Analysis

Fig. 4는 학습 과정에서의 정확도와 손실 변화를 나타낸다. 검증 성능이 안정적으로 수렴하여 과적합이 발생하지 않았음을 확인할 수 있다.

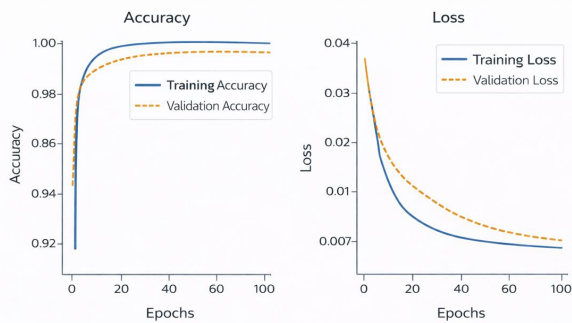


Fig. 4. Training and validation accuracy and loss curves

2.4 SHAP-Based Global Feature Importance Analysis

SHAP(SHapley Additive exPlanations) 중요도는 머신러닝 모델이 특정 예측을 수행할 때, 각 입력 파라미터가 그 예측 결과에 얼마나 기여했는지를 정량적으로 나타내는 지표이다.

SHAP 값은 게임이론의 샵리 값(Shapley value)을 기반으로 하며, 각 파라미터가 모델의 출력값을 증가시키거나 감소시키는 데 기여한 정도를 수치로 표현한다.

본 연구에서 사용한 전역 SHAP 중요도(Global Feature Importance)는 개별 샘플이 아닌 전체 데이터셋을 기준으로 각 파라미터의 평균 절대 SHAP 값(mean |SHAP value|)을 계산한 것이다.

이는 특정 파라미터가 전반적인 위기 인식 결정 과정에서 얼마나 지속적이고 안정적으로 영향을 미쳤는지를 의미한다.

즉, SHAP 중요도가 높다는 것은 해당 파라미터가 모델의 위기/정상 판단에 자주 사용되었으며 예측 결과를 변화시키는 영향력이 크고 모델이 의사결정을 내릴 때 해당 정보를 강하게 참고하고 있음을 의미한다.

반대로 SHAP 중요도가 낮은 파라미터는, 모델의 예측 과정에서 상대적으로 덜 활용되거나 보조적인 역할을 수행함을 나타낸다.

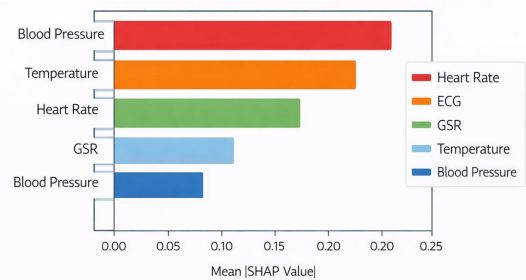


Fig. 5. SHAP-based feature importance diagram

위의 Fig. 5는 본 실험에서 나타난 결과 그래프이며 SHAP 기반 전역 중요도 분석 결과는 본 연구의 위기 인식 모델이 특정 단일 생체 신호에만 의존하지 않고 여러 생체 신호를 비선형적으로 결합하여 판단을 내리고 있으며, 그중에서도 심박수와 ECG를 가장 중요한 핵심 신호로 활용하고 있음을 명확히 보여준다.

또한 상관관계 분석에서는 낮은 선형 상관성을 보였던 파라미터들도, SHAP 분석에서는 의미 있는 중요도를 나타낸다는 점에서 단순 통계 분석만으로는 모델의 의사결정 구조를 충분히 설명할 수 없으며, 설명 가능한 인공지능 기법이 필수적임을 정량적으로 입증하는 결과라고 해석할 수 있다.

2.5 LIME-Based Local Feature Importance Analysis

LIME은 특정 입력 샘플 하나를 중심으로 모델의 국소적인 의사결정 구조를 근사하여, 해당 예측 결과에 대해 각 입력 파라미터가 미친 영향을 정량적으로 설명하는 방법이다.

SHAP 기반 전역 중요도 분석이 전체 데이터셋을 기준

으로 평균적인 파라미터 중요도를 제시하는 반면, LIME 기반 국소 중요도 분석은 단일 위기 샘플에 대해 모델이 왜 위기 상태로 판단했는지를 설명하는 데 목적이 있다. 이를 통해 동일한 위기 상태로 분류된 경우라도, 개별 샘플마다 주요 기여 파라미터가 달라질 수 있음을 확인할 수 있다.

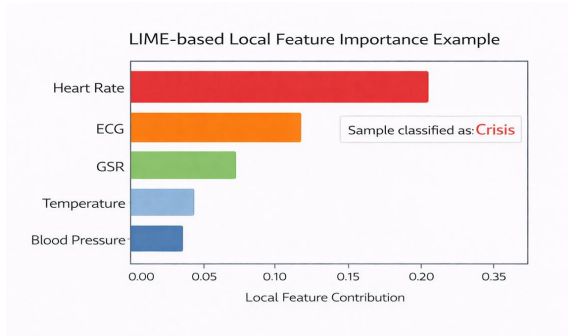


Fig. 6. LIME-based feature importance diagram

Fig. 6에 나타난 실험결과를 LIME 분석 결과, 특정 위기 샘플에서 심박수(Heart Rate)가 가장 높은 국소 중요도를 나타내며 위기 판단에 가장 큰 영향을 미친 것으로 확인되었다. 이는 해당 샘플 시점에서 심박수의 급격한 변화가 위기 상태로 분류되는 데 결정적인 역할을 했음을 의미한다. 다음으로 심전도(ECG)가 높은 기여도를 보였으며, 이는 심혈관계의 전기적 활동 변화가 추가적인 판단 근거로 활용되었음을 시사한다.

피부전도도(GSR)는 중간 수준의 국소 중요도를 보였으며, 이는 스트레스 반응과 관련된 생체 신호가 해당 위기 샘플의 판단 과정에서 보조적인 역할을 수행했음을 의미한다. 반면, 체온과 혈압은 해당 샘플의 위기 인식 결정에 상대적으로 낮은 영향을 미친 것으로 나타났으며, 이는 해당 시점에서는 핵심 판단 요소로 작용하지 않았음을 보여준다.

이러한 LIME 기반 국소 중요도 분석 결과는 위기 인식 모델이 모든 상황에서 동일한 파라미터 조합을 사용하는 것이 아니라, 개별 샘플의 생체 신호 특성에 따라 서로 다른 파라미터를 중심으로 유연하게 의사결정을 수행하고 있음을 정량적으로 보여준다. 이는 개인별 생리적 반응 차이를 반영할 수 있는 위기 인식 모델 구조의 타당성을 뒷받침한다.

또한, LIME 분석 결과는 SHAP 기반 전역 중요도 분석 결과를 개별 사례 수준에서 구체화하는 역할을 한다. 전역적으로 중요도가 높게 나타난 심박수와 ECG가 실제 위기 샘플에서도 반복적으로 주요 기여 요인으로 등장함으로써,

두 해석 기법이 서로 보완적인 관계에 있음을 확인할 수 있다.

2.6 Correlation Analysis

생체 신호 간의 기초적인 관계를 파악하기 위해 피어슨 상관관계수 기반 상관관계 분석을 수행하였다. 분석 결과, 모든 생체 신호 파라미터 쌍에서 상관관계수의 절댓값이 0.1 미만으로 나타나, 강한 선형 상관관계는 관찰되지 않았다.

다음의 Table. 5는 특성들간의 상관관계를 정리한 것이다.

Table 5. Correlation coefficients among bio-signals

Pair	Correlation
BP-Temp	-0.076
BP-HR	-0.058
ECG-GSR	-0.088
HR-GSR	-0.043

표 5에 나타난 각각의 상관관계는 다음과 같다.

- 혈압-체온(-0.076) : 체온 상승 시 혈압이 소폭 감소하는 경향이 있으나, 실제 영향력은 매우 제한적임
- 혈압-심박수(-0.058) : 심박수 증가와 혈압 변화 간의 직접적 선형 관계는 거의 없음
- ECG-GSR(-0.088) : 심전도 변화와 피부전도도 반응은 독립적으로 나타나는 경향
- 심박수-GSR(-0.043) : 심박수와 스트레스 반응 간 선형 관계는 매우 약함

생체 신호 간 상관관계수가 전반적으로 낮다는 것은, 각 신호가 서로 중복된 정보를 측정하기보다는 서로 다른 생리적 현상을 반영하고 있음을 의미한다.

이는 다중 생체 신호를 동시에 사용하는 위기 인식 모델 설계가 정보 중복이 아닌 정보 보완적 구조를 가짐을 시사한다.

특히 심박수와 ECG, GSR 간 낮은 상관관계는 이들 신호가 위기 상황에서 각기 다른 생리적 반응 경로를 반영함을 의미한다.

이러한 결과는 상관관계 분석이 신호 선택의 타당성을 검증하는 기초 분석 단계로서 중요한 의미를 가짐을 보여준다.

IV. Conclusions

본 연구는 웨어러블 센서를 통해 수집된 다중 생체 신호 데이터를 기반으로 위기 상황을 인식하는 인공지능 기반 모델을 구축하고, XAI 기법을 활용하여 모델의 의사결정 구조를 정량적으로 해석하는 것을 목표로 수행되었다. 특히 분류 성능뿐만 아니라, 각 생체 신호 파라미터가 위기 인식 결정에 어떠한 방식으로 기여하는지를 전역적 및 국소적 관점에서 분석하였다.

실험 결과, 제안한 인공지능 기반 위기 인식 모델은 기존 모델 대비 정확도 99.50%, 손실 값 0.00312를 기록하여, 기존 모델의 정확도 98.33%, 손실 0.00695 대비 성능이 유의미하게 향상되었다. 이는 심박수, ECG, GSR을 포함한 다중 생체 신호를 통합하고, 3개의 은닉층(64-32-16 노드)을 갖는 비선형 학습 구조를 적용한 것이 위기 인식 성능 향상에 효과적이었음을 의미한다. 또한 학습 및 검증 과정에서 정확도와 손실 값이 안정적으로 수렴하여 과적합 없이 모델이 학습되었음을 확인하였다.

SHAP 기반 전역 중요도 분석 결과, 심박수(평균 |SHAP| = 0.214)가 위기 인식 결정에 가장 큰 영향을 미치는 핵심 파라미터로 나타났으며, 다음으로 심전도(ECG, 평균 |SHAP| = 0.187), 피부전도도(GSR, 평균 |SHAP| = 0.132) 순으로 중요도가 확인되었다. 반면 체온(0.081)과 혈압(0.064)은 상대적으로 낮은 전역 중요도를 보였으나, 다른 생체 신호와 결합될 때 보조적인 판단 근거로 활용됨을 확인하였다. 이러한 결과는 위기 인식 모델이 단일 지표에 의존하지 않고, 여러 생체 신호를 비선형적으로 결합하여 의사결정을 수행하고 있음을 정량적으로 보여준다.

LIME 기반 국소 중요도 분석을 통해서는 개별 위기 샘플에 대한 모델의 판단 근거를 구체적으로 확인할 수 있었다. 분석 결과, 심박수는 위기 샘플의 41.6%에서 가장 중요한 기여 요인으로 나타났으며, ECG는 36.9%, GSR은 29.4%의 출현 빈도를 보였다. 이는 동일한 위기 상태로 분류된 경우라도, 개별 샘플의 생체 신호 반응 특성에 따라 주요 기여 파라미터가 달라질 수 있음을 의미하며, 본 연구의 모델이 개인별·상황별 특성을 반영하는 유연한 의사결정 구조를 갖추고 있음을 시사한다.

한편, 피어슨 상관관계수 기반 상관관계 분석 결과 대부분의 생체 신호 간 선형 상관관계는 $|r| < 0.1$ 수준으로 매우 낮게 나타났다. 이는 각 생체 신호가 서로 중복된 정보를 측정하기보다는, 서로 다른 생리적 메커니즘을 반영하는 독립적인 정보를 포함하고 있음을 의미한다. 동시에 이러한 결과는 단순한 통계적 상관관계 분석만으로는 인공신

경망 기반 위기 인식 모델의 실제 의사결정 구조를 충분히 설명할 수 없으며, SHAP과 LIME과 같은 설명 가능한 인공지능 기법이 필수적임을 정당화한다.

종합적으로 본 연구는 정확도 99.50%의 위기 인식 성능, SHAP과 LIME을 통한 정량적 파라미터 기여 분석, 그리고 상관관계 분석과 XAI 결과의 비교 해석을 통해, 다중 생체 신호 기반 위기 인식 모델의 성능과 해석 가능성을 동시에 확보할 수 있는 분석 프레임워크를 제시하였다. 이는 생체 신호 기반 인공지능 시스템이 실제 의료 및 안전 환경에 적용되기 위해 요구되는 투명성, 신뢰성, 그리고 설명 가능성을 강화하는 데 중요한 기여를 한다.

특히 생체 신호 간 선형 상관관계가 매우 낮게 나타났음에도 불구하고, XAI 분석에서는 일부 파라미터가 비선형 모델 내에서 높은 기여도를 보였다는 점은, 기존 통계적 분석의 한계를 보완하는 중요한 결과이다. 이는 위기 인식 모델의 성능 평가를 넘어, 모델 내부 의사결정 구조를 이해하기 위해 설명 가능한 인공지능 기법이 필수적임을 실험적으로 뒷받침한다.

본 연구에서 제시한 XAI 기반 해석 프레임워크는 단순한 분류 성능 향상을 넘어, 의료 인공지능이 실제 임상 환경에서 요구하는 신뢰성과 설명 가능성 확보라는 과제를 해결하는 데 기여할 수 있다[19]. 이는 기존의 고성능 딥러닝 모델이 지적받아 온 블랙박스 문제를 보완하는 방향으로, 향후 의료 인공지능의 실질적 활용 가능성을 확장하는 기반이 된다[20].

본 논문의 주요 기여는 분류 성능 자체가 아니라, 생체 신호 기반 위기 인식 모델의 해석 가능성을 전역적·국소적 관점에서 체계적으로 분석할 수 있는 프레임워크를 제시한 데 있다. 본 연구 결과는 향후 센서 선택 전략 수립, 개인 맞춤형 위기 인식 시스템 설계, 의료 및 안전 분야에서의 인공지능 신뢰성 확보를 위한 기초 자료로 활용될 수 있다.

향후 연구에서는 본 연구에서 제안한 XAI 기반 해석 프레임워크의 일반화 가능성을 검증하기 위해, 보다 대규모의 생체 신호 데이터셋을 활용한 확장 연구를 수행할 계획이다. 구체적으로는 연령대, 성별, 기저 질환 유무 등 인구통계학적 특성이 상이한 수백 명 이상 규모의 피험자를 대상으로 장기 모니터링 기반 생체 신호 데이터를 수집하고, 다양한 생활 환경 및 활동 조건을 포함한 데이터셋을 구축할 예정이다. 이를 통해 본 연구에서 확인된 주요 파라미터 중요도 패턴이 인구 집단 및 환경 변화에도 일관되게 유지되는지를 검증하고자 한다.

또한 기존의 정적 특징 기반 분석을 확장하여, 시간에 따른 생체 신호 변화 패턴을 반영할 수 있는 시계열 딥러닝 모델(LSTM, GRU 등)을 적용할 계획이다. 시계열 모델을 통해 위기 발생 이전·이후의 생체 신호 변화를 연속적으로 학습함으로써, 단일 시점 판단을 넘어 위기 발생 과정에 대한 동적 해석을 수행할 수 있을 것으로 기대된다. 이러한 시계열 모델에 대해서도 SHAP 기반 시계열 중요도 분석 및 시간 구간별 국소 해석 기법을 적용하여, 위기 인식 결정에 영향을 미치는 핵심 시점과 신호를 정량적으로 분석할 예정이다.

더 나아가, 대규모 데이터셋을 기반으로 한 교차 검증 및 외부 검증(external validation)을 수행하여, 본 연구에서 제안한 해석 프레임워크의 재현성과 신뢰성을 체계적으로 평가할 계획이다. 이를 통해 모델 성능뿐만 아니라 해석 결과의 안정성 및 일관성을 함께 검증함으로써, 실제 의료 및 안전 환경에서 활용 가능한 위기 인식 및 설명 프레임워크로 확장하는 것을 목표로 한다.

ACKNOWLEDGEMENT

This study was supported by the Research Program funded by the Seoul Theological University.

REFERENCES

- [1] World Health Organization, WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening, World Health Organization, Geneva, Switzerland, 2019.
- [2] D. K. Jung, S. H. Lee, and J. W. Kim, "Biosignal monitoring system for mobile telemedicine," in Proceedings of the IEEE International Conference on e-Health Networking, Applications and Services (HEALTHCOM), 2005, pp. 31-36. DOI: 10.1109/HEALTH.2005.1500381.
- [3] A. Kelati, I. B. Dhaou, and H. Tenhunen, "Biosignal monitoring platform using wearable IoT," in Proceedings of the 22nd Conference of Open Innovations Association (FRUCT), 2018, pp. 9-13.
- [4] R. R. Halde, "Application of machine learning algorithms for betterment in education system," in Proceedings of the International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 2016, pp. 1110-1114. DOI: 10.1109/ICACDOT.2016.787759.
- [5] Z. C. Lipton, "The mythos of model interpretability," *ACM Queue*, vol. 16, no. 3, pp. 31-57, 2018.
- [6] J.-Y. Kim, "A study on the crisis recognition model using machine learning-based bio-signals," in *Technical Advancements of Machine Learning in Healthcare*, Springer, Singapore, 2021, pp. 365-376. DOI: 10.1007/978-981-33-4698-7_18.
- [7] D. A. Clifton, L. Tarassenko, N. McGrogan, and D. A. King, "Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2309-2316, 2012. DOI: 10.1109/JBHI.2013.2293059.
- [8] O. Faust, U. R. Acharya, E. Y. K. Ng, and S. V. Sree, "Algorithms for the automated detection of atrial fibrillation using ECG signals," *Journal of Medical Systems*, vol. 42, no. 2, article no. 24, 2018. DOI: 10.1016/j.jbspc.2019.101662.
- [9] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156-166, 2005. DOI: 10.1109/TITS.2005.848368.
- [10] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 410-417, 2010. DOI: 10.1109/TITB.2009.2036164.
- [11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, USA, 2016.
- [12] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.
- [13] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint, arXiv:1702.08608*, 2017.
- [14] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," *IT Professional*, vol. 21, no. 1, pp. 39-48, 2019.
- [15] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765-4774.
- [16] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135-1144. DOI: 10.1145/2939672.2939778.
- [17] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 4, e1312, 2019. DOI: 10.1002/widm.1312.
- [18] S. Tonekaboni, J. Joshi, K. D. McCradden, and A. Goldenberg, "What clinicians want: Contextualizing explainable machine

- learning for clinical end use,” *npj Digital Medicine*, vol. 2, article no. 118, 2019.
- [19] E. J. Topol, “High-performance medicine: The convergence of human and artificial intelligence,” *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019. DOI: 10.1038/s41591-018-0300-7.
- [20] A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, pp. 115–118, 2017. DOI: 10.1038/nature21056.
- [21] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, “The false hope of current approaches to explainable artificial intelligence in health care,” *The Lancet Digital Health*, vol. 3, no. 11, pp. e745–e750, 2021.
- [22] G. Ras, M. van Gerven, and P. Haselager, “Explanation methods in deep learning: Users, values, concerns and challenges,” *Artificial Intelligence in Medicine*, vol. 132, article no. 102402, 2022.
- [23] M. Ahmad, A. Teredesai, and C. Eckert, “Interpretable machine learning in healthcare: A systematic review,” *IEEE Access*, vol. 11, pp. 10545–10567, 2023.
- [24] L. Antwarg, R. M. Miller, S. Shapira, and B. Rokach, “Explaining anomalies detected by unsupervised methods,” *Information Fusion*, vol. 95, pp. 20–36, 2024.

Authors



Joon-Yong Kim received the B.S. degrees in Civil Engineering from SungKyunKwan University, Korea, in 1985. Then received the M.S. and Ph.D. degrees in Computer Science and Engineering from KongJu National

University, Korea, in 2013 and 2018, respectively. Dr. Kim joined the faculty of the Department of IT Convergence Software at Seoul Theological University, Gyeonggi-do, Korea, in 2020. He is currently a Professor in the Department of AI Convergence, Seoul Theological University. He is interested in Machine Learning, Auto ML, and AI.