Interpretable Tree-Based Anxiety Classification from Wearable Biosignals Using SHAP and Bias Diagnostics

Anne Schwerk
Department IT & Engineering, Extended Artificial Intelligence, IU International University, Bad Honnef, Germany

Ali Mehmood Khan
Department IT & Engineering, Extended Artificial Intelligence, IU International University, Bad Honnef, Germany

Stephan de la Rosa
Department IT & Engineering, Extended Artificial Intelligence, IU International University, Bad Honnef, Germany

Armin Grasnick
Department IT & Engineering, Extended Artificial Intelligence, IU International University, Bad Honnef, Germany

Thomas Bolz
Department IT & Engineering, Extended Artificial Intelligence, IU International University, Bad Honnef, Germany

Abstract

Interpretable identification of anxiety states from physiological signals is a central challenge in affective computing, particularly for wearable-based mental health monitoring. This study presents a transparent, tree-based framework for multi-class anxiety classification using multimodal biosignals acquired from an FDA-cleared wrist-worn wearable device. Physiological data were collected from 18 participants during anxiety-evoking virtual reality exposure and included electrodermal activity (EDA), blood volume pulse (BVP), and peripheral skin temperature. Subjective anxiety was repeatedly assessed using a single-item self-report on a seven-point scale and discretized into anxiety classes for supervised learning.

A decision tree classifier was trained on physiologically motivated summary features extracted from fixed-length windows and evaluated using 10-fold stratified cross-validation. TreeSHAP was applied to analyze global, class-wise, and local feature contributions. Global SHAP analysis showed that EDA-derived features dominate model decisions, followed by cardiovascular variability and temperature. Class-wise SHAP analysis revealed that different anxiety categories rely on distinct physiological feature combinations, highlighting heterogeneity not observable from aggregate performance metrics. SHAP-based feature selection further demonstrated that most predictive performance can be retained using a reduced, interpretable feature set.

Overall, the results indicate that lightweight, interpretable tree-based models can achieve competitive anxiety classification performance from wearable biosignals while enabling inspection of model reasoning.

Keywords

Affective computing, wearable biosignals, SHAP explainable AI, machine learning, algorithmic bias detection.