Stress-Level Management via Smartwatch Monitoring

Introduction

Stress is a pervasive health issue among students and professionals, reducing productivity and well-being. Current smartwatch-based solutions lack accuracy and real-time capabilities. This project develops a machine learning system to classify stress states (baseline, stress, recovery) using multimodal wearable sensor data: heart rate variability (HRV), blood volume pulse (BVP), electrodermal activity (EDA), skin temperature, and motion. The goal is continuous, objective stress monitoring rather than subjective questionnaire-based approaches.

Datasets

Wearable Device Dataset (PhysioNet)

Empatica E4 wristband recordings from 36 subjects (aged 18–30) undergoing structured stress induction (Stroop test, arithmetic tasks, rest periods). Raw multimodal signals (BVP at 64 Hz, EDA at 4 Hz, temperature at 4 Hz, 3-axis accelerometer at 32 Hz) plus self-reported stress scores and task labels.

Strengths:

Multimodal raw data, flexible preprocessing and feature extraction, realistic stress protocol with clear baseline–stress–recovery phases.

Limitations:

Young healthy population only, documented quality issues (duplicated signals, wristband misplacement, missing data), small sample size.

SWELL-HRV Dataset (Kaggle)

Office-work stress from 25 subjects under three conditions: neutral, time pressure, interruption. Precomputed HRV features (time- and frequency-domain) derived from ECG, plus subjective stress and task load assessments

Strengths:

Real-world context, ready-to-use HRV features, multiclass labeling, subjective–physiological correlation data.

Limitations:

HRV features only (no raw ECG or other modalities), small sample, cannot apply advanced time-series modeling directly.

Rationale for Using Both Datasets

Wearable dataset enables multimodal signal processing and temporal modeling; SWELL provides real-world validation and fast prototyping. Cross-dataset evaluation strengthens generalization claims.

Methodology

Phase 1 — Classical Feature-Based Models

Random Forest and Support Vector Machine using HRV and extracted multimodal features. Prior work reduced 67 HRV features to 8 discriminative ones with >99% accuracy. Feature selection via mutual information or mRMR reduces dimensionality and overfitting.

Phase 2 — Time-Series Models

LSTM or CNN–LSTM on raw Wearable dataset signals, learning temporal dynamics of stress onset, duration, and recovery. Preprocessing includes alignment across modalities (different sampling rates), normalization, and sliding-window segmentation (30–120 s windows).

Challenges

Small sample sizes (36 and 25 subjects) increase overfitting risk and limit generalization. Signal heterogeneity across datasets complicates model transfer. Self-reported stress labels introduce subjectivity noise. Wearable data suffers motion artefacts and missing values. Real-world stress differs from controlled lab/office settings.

Expected Outcome

Baseline stress detection model (Random Forest or SVM) with documented accuracy, precision, recall, F1-score. Advanced LSTM model capturing temporal stress dynamics. Cross-subject and cross-dataset evaluation demonstrating feasibility and limitations. Identification of practical deployment requirements (user calibration, noise handling, data diversity).

Future Work

Expand to diverse real-world data (daily life, exams, variable stressors). Develop user-adaptive models. Integrate real-time feedback layer (breathing prompts, alerts). Apply explainability methods (feature importance, attention mechanisms) for user insight. Track long-term stress trends for proactive well-being management.