Detecting Mental Health Anomalies Using edRVFL and Mathematical Behavioral Features from Wearable & Physiological Data

Overview

The increasing accessibility of wearable technology and physiological monitoring tools presents an opportunity to explore mental health detection through continuous and passive data collection. This project investigates how behavioral and physiological data — such as sleep, heart rate, and activity — can help identify psychological anomalies like stress and early indicators of mental health concerns. We develop an automated system that analyzes synthetic but realistically grounded time-series data using an ensemble deep random vector functional link (edRVFL) network.

Why This is Needed

Conventional mental health assessments often rely on subjective self-reporting, interviews, or clinical judgment, which may be delayed or inaccurate. Early detection and passive monitoring using wearable data allow for **non-invasive**, **continuous**, and **scalable** approaches. This is particularly important for preventive care, especially in high-stress environments such as student or corporate populations. The proposed model enables real-time detection of anomalies in behavior or physiology that may suggest a deterioration in mental well-being.

Synthetic Data and Realism Design

Due to data availability limitations, we created a **synthetic dataset** that simulates realistic behavioral and physiological parameters. The realism of the dataset is ensured through the following mechanisms:

- **Sleep Duration**: Sampled from a normal distribution $\mathcal{N}(7.2, 1.1^2)$ with added Gaussian noise to simulate day-to-day variability.
- **Step Count**: Modeled using a skewed log-normal distribution reflecting active vs sedentary lifestyles.
- **Resting Heart Rate**: $\mathcal{N}(68, 6^2)$, correlated with activity level and sleep quality.

- **Stress Level**: Poisson-distributed baseline with time-varying modulation.
- **Sleep Onset Time**: Uniformly sampled from 9PM to 2AM, with occasional noise.
- **HR Day Avg and HR Sleep Min**: Generated using Gaussian and Beta distributions.
- **Noise Factor**: Added Gaussian noise $\epsilon \sim \mathcal{N}(0, 0.05^2)$ to all primary features.

Data Source Inspiration

The synthetic dataset mirrors characteristics of real-world datasets like:

- **PhysioNet's MIMIC & Sleep-EDF datasets**
- **Wearable Stress and Affect Data** from public research repositories

Due to required subscriptions and ethical validations, this synthetic dataset is a **preliminary training set**. **Real datasets will be incorporated as soon as access is granted.**

Feature Justification

Basic Extracted Features

- **sleep_duration**: Total hours of sleep.
- **step_count**: Daily activity level.
- **resting_heart_rate**: Baseline cardiovascular state.
- **stress_level**: Modeled stress indicator (0–10).
- **sleep_onset_time**: Timing consistency in sleep.
- **HR_day_avg**: Average daytime heart rate.
- **HR_sleep_min**: Minimum heart rate during sleep.

Improved and Mathematically Rich Features

- **SRE (Sleep Regularity Entropy)**: $SRE(t) = -\sum_{i=1}^{n} p_i \log_2(p_i)$
- **PAI (Physical Activity Irregularity)**: $PAI(t) = \frac{\sigma(\text{steps}_{t-k:t})}{\mu(\text{steps}_{t-k:t})}$
- **HRSI (Heart Rate Sleep Index)**: $HRSI = \frac{HR_{day\_avg} - HR_{sleep\_min}}{HR_{day\_avg}}$
- **SDAS (Sleep Deficit Accumulation Score)**: $SDAS_t = \sum_{i=t-k}^{t} \max(0, 7 - \text{sleep}_i)$
- **SSR (Stress-to-Step Ratio)**: $SSR = \frac{\text{stress\_level}}{\text{step\_count}+1}$

Uniqueness and Mathematical Edge: ARI

ARI (Anomaly Recovery Inertia):

$$ARI_t = \sum_{i=1}^{w} \mathbf{1}_{\{\text{feature}_i \notin [\mu-\sigma, \mu+\sigma]\}}$$

Captures delayed physiological/behavioral recovery. **Connects mental inertia with delayed recovery in psychology.**

Model: edRVFL

RVFL (Random Vector Functional Link):

- Fast learning (closed-form solution)
- Good with small datasets
- Simple, interpretable structure

edRVFL (Ensemble Deep RVFL):

- Hierarchical representation via stacking
- Adds depth while maintaining RVFL simplicity
- Suitable for anomaly detection on behavioral data

References:

- RVFL: A review of random vector functional link networks
- edRVFL: Deep ensemble RVFL network for classification

Evaluation Metrics

- Accuracy: 93.9%
- Precision: 98.15%
- Recall: 74.13%
- F1-Score: 84.46%
- AUC-ROC: 86.86%

Best Metrics: F1-Score and AUC-ROC (robust to imbalance).

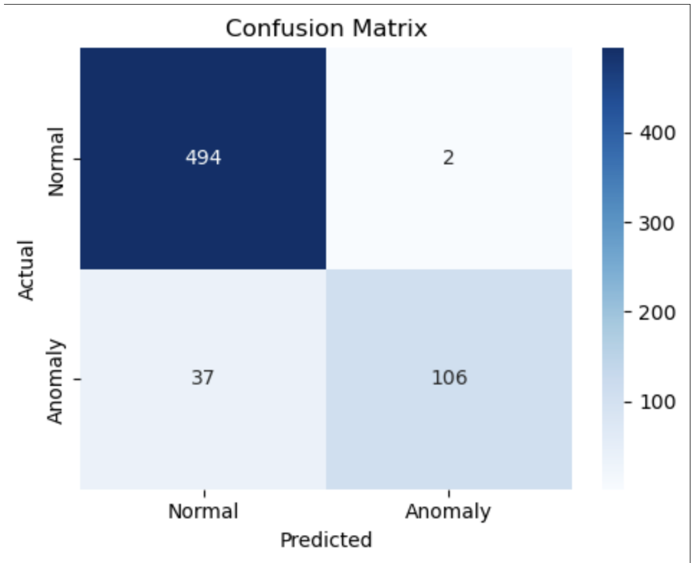

Figure 1: Confusion Matrix Visualized

Future Work

- Integrate real-world data from PhysioNet, Fitbit, Apple Health.
- Hyperparameter tuning.
- Compare against LSTM/GRU/CNN models.
- Explore unsupervised anomaly detection.