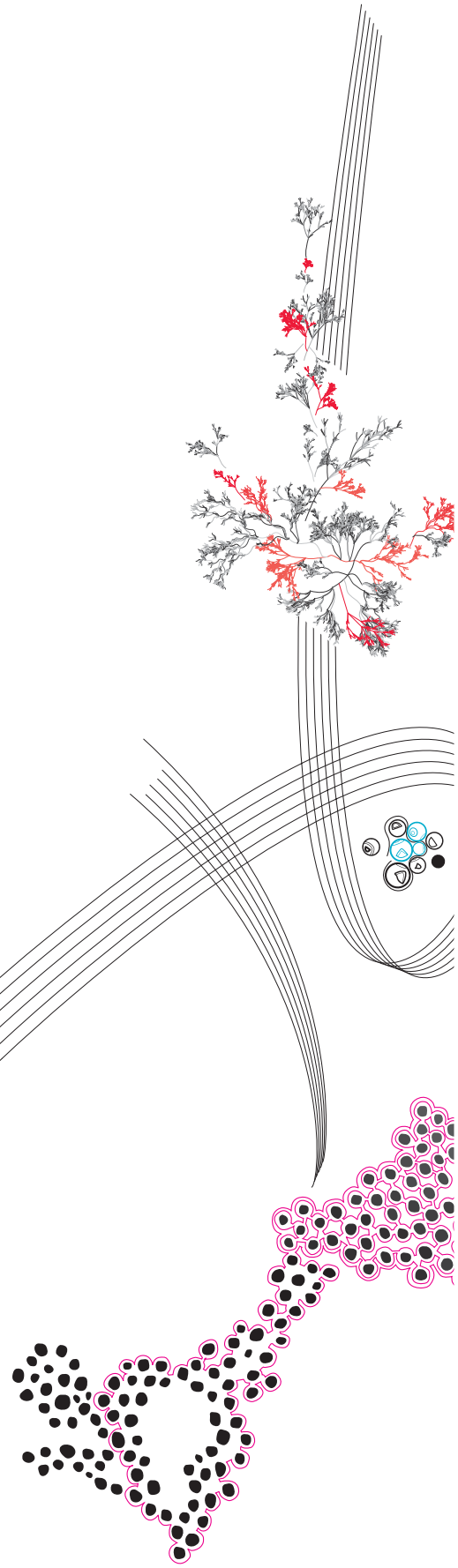

MSc Embedded Systems
Final Project

Predicting Positive Psychological States using Machine Learning and Digital Biomarkers from wearable data

Lingxi Wu

Supervisor:

Dr. Arlene John
Dr. Jorge Piano Simoes
Dr.ir. Beijnum, Bert-Jan van
Dr. E. Talavera Martínez

March, 2025

Department of Biomedical Signals and Systems
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

Contents

Abstract

Wearable devices provide continuous physiological data, offering opportunities for real-world mental health monitoring. While prior research has primarily focused on detecting stress and negative psychological states, the prediction of positive psychological states remains underexplored. This study investigates whether digital biomarkers derived from smartwatch data can predict daily psychological states based on the PERMA framework (Positive Emotion, Engagement, Relationships, Meaning, and Accomplishment).

Using data from 34 participants wearing research-grade smartwatches for eight days, physiological signals—including heart rate variability (HRV), electrodermal activity (EDA), and accelerometer-derived movement features—were processed into machine-learning-ready features. Machine learning models, including Random Forest, Long Short-Term Memory networks (LSTM), Convolutional Neural Networks (CNNs), and Transformers, were trained to predict daily self-reported psychological states. Explainable AI techniques, such as SHAP and LIME, were employed to identify key physiological markers, while conformal prediction assessed model uncertainty.

Results indicate that while absolute prediction accuracy remains a challenge, CNN models achieved a threshold accuracy of up to 62%, with accelerometer-based features showing the highest predictive importance. Self-esteem-related constructs demonstrated the most distinguishable physiological patterns, while anger was more detectable than anxiety or sadness. Uncertainty analysis further highlighted the potential for improving confidence in predictions.

These findings suggest that wearable-derived biomarkers hold promise for tracking positive psychological states, albeit with limitations in predictive accuracy. Future work should focus on refining feature extraction, improving model generalization, and integrating multimodal data sources to enhance real-world applicability in mental health monitoring.

Predicting Positive Psychological States using Machine Learning and Digital Biomarkers from wearable data

Lingxi Wu

March 2025

Chapter 1

Introduction

In recent years, wearable devices have revolutionized the collection and analysis of physiological and behavioral data. Devices such as Fitbit[1], Garmin[2], and Apple Watch[3] offer continuous tracking of metrics like heart rate, skin conductance, physical activity, and sleep patterns, enabling insights into human psychological states in real-world contexts. This shift represents a transformative opportunity to advance mental health monitoring and research outside controlled laboratory settings.

Mental health challenges, including stress, anxiety, and depression, are recognized as global priorities. Traditional detection methods, relying on self-reported questionnaires or structured interviews, are often limited by recall bias and lack real-time applicability. Wearable devices address these limitations by providing continuous, objective data, capturing the nuanced, moment-to-moment fluctuations of psychological states[4][5]. While much of the existing research focuses on stress or other negative mental states[6], understanding positive psychological states—such as happiness and resilience—is equally critical to enhancing well-being, productivity, and social connection.

Despite the ubiquity of wearable data, leveraging it to predict positive psychological states remains under explored[7]. Addressing this gap, this study investigates the feasibility of predicting and understanding positive psychological states using wearable device data, combining machine learning and explainable AI techniques. It also combines ecological momentary assessment to track psychological states of participants over time, acknowledging that these fluctuate over time and are context-dependent.

The research methodology encompasses key stages: preprocessing raw wearable signals into interpretable features using frameworks like FLIRT[8], applying machine learning algorithms to predict psychological outcomes, and employing explainable AI techniques to identify the most impactful features. Additionally, uncertainty quantification methods, such as conformal prediction, are integrated to ensure robust and reliable predictions. Together, these steps provide a comprehensive framework for processing the physiological signals and predicting positive psychological states.

The contributions and research questions of this work are:

- **Main Research Question: Can positive psychological states be predicted using machine learning and digital biomarkers derived from wearable data?**
- **Sub Research Question 1:** How can a data processing pipeline be designed and

implemented to extract meaningful features, handle missing data, and analyze temporal patterns from wearable data for predicting psychological states?

- **Sub Research Question 2:** How can interpretable machine learning techniques be employed to identify and validate the key physiological markers associated with psychological states, ensuring the predictions are transparent and actionable?

Chapter 2

Related Works

The use of wearable technology and machine learning for predicting psychological states has become an active research area, encompassing diverse aspects such as wearable data preprocessing, multi-modal feature engineering, psychological state prediction, and theoretical frameworks like the PERMA model. This section reviews prior work in these areas and highlights the gaps addressed by this study.

2.1 Mental Health Monitoring and Digital Phenotype

Wearable Devices for Mental Health Monitoring

The variability in assessment tools for diagnosing and screening mental disorders is significant, as highlighted by an analysis of 126 different questionnaires and interviews [9]. Such heterogeneity can result in inconsistencies in diagnosis and treatment outcomes. Wearable devices offer a promising alternative by enabling continuous, non-invasive monitoring of physiological signals, providing a foundation for detecting and predicting mental states. In [10], digital phenotypes for stress detection were introduced, leveraging large-scale wearable data to identify stress biomarkers. Depression severity monitoring through wearable and mobile sensors was proposed in [11], showcasing the potential of digital phenotyping in mental health assessment. Recent advancements in mental health monitoring emphasize generalizability and personalization in stress prediction. The integration of passive sensing with actionable insights for clinical mental healthcare has been explored, demonstrating how wearable data can guide behavioral interventions [12]. Similarly, ensemble learning models for stress prediction have been developed, showing robust generalization across synthetic datasets [13]. In addition, PredictEYE, a personalized time-series model that predicts mental states using eye-tracking data, has highlighted the potential of personalization in mental health applications [14].

Beyond research, digital phenotyping is gradually achieving regulatory recognition, paving the way for its adoption in medical diagnostics and treatment evaluation. A milestone in this field has been reached with the first-ever regulatory qualification of a digital biomarker—the stride velocity 95th centile (SV95C)—as a primary endpoint for monitoring disease progression in Duchenne muscular dystrophy (DMD) [15]. This development marks a significant shift, demonstrating that digital health technologies can meet the rigorous standards of regulatory agencies. As digital phenotyping continues to gain validation, its application in mental health and other medical domains is expected to expand,

reinforcing its role in clinical decision-making and personalized healthcare.

Multimodal and Advanced Modeling Techniques

The integration of multiple sensors and modalities has proven to be highly effective in enhancing psychological state prediction. A review of machine learning techniques for mental health diagnoses highlighted deep learning's potential to address challenges in student populations [16]. The development of multivariate time-series models for mental state prediction incorporated temporally observable physical and behavioral data, demonstrating the value of comprehensive modeling approaches [17]. A multimodal framework, M3Sense, leveraged representation learning to detect mental health conditions using limited labeled data [18]. Expanding this approach, the integration of environmental sensor data revealed its value for assessing cognitive function [19]. Research on behavioral patterns collected via smartphones established links to personality traits, further underscoring the role of multimodal data in psychological research [20].

Improvements in combining multiple models and matching data from different sources have made multimodal models more powerful. A notable example is Husformer, a multimodal transformer for human state recognition, which demonstrated the effectiveness of deep learning architectures in multimodal analysis [21].

Digital Biomarkers

The integration of digital biomarkers into machine learning workflows has significantly advanced mental health research. Studies on illness activity in depressive episodes have presented automated pipelines that leverage multivariate time-series data for symptom monitoring [22, 23]. These findings emphasize the importance of combining domain knowledge with machine learning techniques. The feasibility of designing models for predicting emotional states from passively sensed data was demonstrated through hierarchical models that accounted for individual differences [24]. Despite these advancements, digital phenotyping remains underexplored in monitoring and predicting positive emotional states that are not associated with mood disorders, highlighting a key research gap.

Applications of Stress Detection and Cognitive Monitoring

Wearable-based stress detection has yielded promising results in both real-world and controlled settings. The feasibility of monitoring stress using heart rate variability (HRV) and photoplethysmography in everyday environments was demonstrated despite challenges posed by environmental factors [25]. Similarly, Bayesian change point detection techniques have been applied to track acute stress responses, enhancing the granularity of stress assessment [26].

In the field of cognitive monitoring, the synergy between digital cognitive tests and wearable devices has been explored for mild cognitive impairment detection. Improved classification performance was observed when physiological and behavioral features were combined, reinforcing the importance of multimodal data integration in cognitive assessments [27, 28].

2.2 Data Preprocessing, Feature Engineering, and Imputation for wearable data

The preprocessing and feature engineering of wearable data play a critical role in building accurate predictive models. The FLIRT toolkit automates key tasks such as artifact de-

tection, signal filtering, and feature extraction, significantly simplifying the preparation of wearable data for machine learning [8]. This toolkit primarily focuses on three specific types of data—Heart Rate Variability (HRV), Electrodermal Activity (EDA), and Accelerometer (ACC)—to generate features. The issue of missing data has been addressed with a multivariate imputation framework, improving the handling of large gaps in time-series sensor data [29].

The importance of optimized feature selection and extraction has been further demonstrated through research that optimized wearable biosensor data for stress classification using explainable AI [30]. Frequency spectrum analysis for electrodermal activity (EDA) signals has been proposed, showing superior stress detection accuracy with fewer data points [31]. The significance of HRV and ACC data in enhancing predictive models using wearable devices has also been highlighted [32]. Studies have shown that combining HRV and ACC data improves the accuracy of health monitoring systems, while machine learning techniques help enhance HRV measurement accuracy by mitigating errors caused by user movement.

By leveraging these data sources alongside advanced processing techniques, researchers can develop more accurate and reliable models for various health-related outcomes.

Chapter 3

Technical Background

3.1 Interpretable Machine Learning and Uncertainty Quantification

Interpretable machine learning has become essential in wearable data research, enhancing the transparency and applicability of predictive models. Stress detection has benefited from interpretable models that integrate physiological markers with behavioral features like keystroke dynamics [33]. Additionally, explainable deep learning methods, such as those leveraging integrated gradients, have provided valuable insights into the relationships between physiological signals and psychological states [34].

The reliability of machine learning models has been further strengthened by the ability to estimate prediction uncertainty. Conformal prediction, a distribution-free uncertainty quantification method, has been introduced to ensure confidence in model predictions and improve their applicability in real-world scenarios [35].

3.2 Theoretical Models for Mental Health

The PERMA model provides a theoretical framework for understanding well-being through five core elements: Positive Emotion, Engagement, Relationships, Meaning, and Accomplishment, as shown in Figure 3.1 [36]. Its application in mental health interventions has been recommended, particularly for addressing anxiety [37]. The model's multidimensional approach has also been validated in student populations, demonstrating links between its elements and factors such as life satisfaction and physical vitality [38]. Furthermore, its utility in fostering mental well-being has been highlighted in studies focusing on undergraduate students [39].

This study integrates elements of the PERMA model into predictive frameworks for positive psychological states, finding connections between theoretical constructs and data-driven methods. While previous research has predominantly focused on symptom monitoring and mood disorders, this work extends the scope of digital phenotyping by evaluating its potential for predicting positive emotional states, an area that remains largely unexplored.

3.3 Ecological Momentary Assessment

Ecological Momentary Assessment (EMA), also known as the Experience Sampling Method (ESM), is a research methodology that involves collecting data from participants in real-time and within their natural environments. This approach requires individuals to report on their thoughts, feelings, behaviors, and contexts as they occur, thereby minimizing recall bias and enhancing ecological validity[40]. By capturing data in real-time, EMA reduces the reliance on participants' memory, thereby minimizing recall bias. This leads to more accurate and reliable data that reflect participants' genuine experiences in their natural settings. EMA allows researchers to gather comprehensive information about the situational factors influencing participants' psychological states. This includes understanding how specific contexts or events impact mood and behavior, providing a nuanced view of the dynamics of psychological experiences.By collecting data at multiple points over time, EMA enables the examination of fluctuations in psychological states, capturing both short-term variations and long-term trends. This temporal sensitivity is crucial for understanding the dynamics of mood and behavior. EMA can be tailored to specific research needs, allowing for various data collection methods such as self-reports, physiological measurements, or passive data collection through mobile devices. This flexibility makes it applicable across diverse psychological research areas[41].

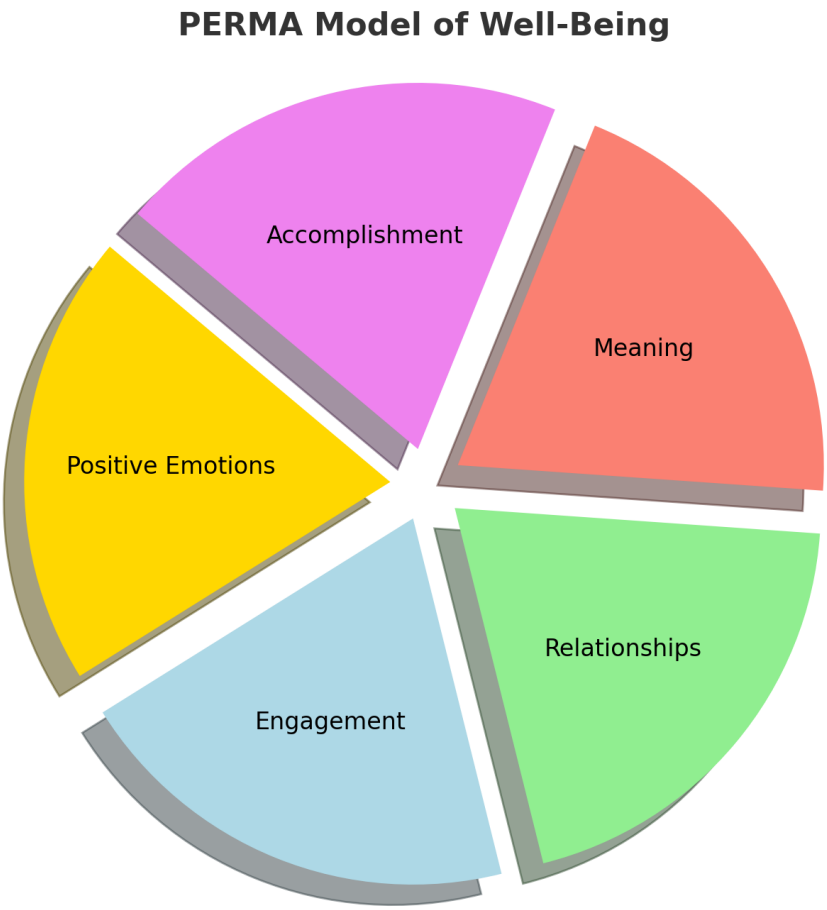

FIGURE 3.1: Example of Seligman's PERMA model[36]

Chapter 4

Methodology

4.1 Dataset Overview

The study was conducted under the approval of the University of Twente Ethics Committee. The Ethics Committee of the University of Twente's Faculty of Behavioural Sciences approved this study (project 240133, obtained on 23/02/2024). This dataset offers a rich combination of physiological and psychological data, uniquely suited for studying the relationship between wearable signals and mental states in young adults. While the dataset is relatively small compared to those used by large technology companies, it provides a controlled and ecologically valid foundation for academic research.

4.1.1 Participant Demographics

The dataset was collected by the PHT (Psychology, Health, and Technology) department in the BMS (Behavioural Management and Social Sciences) faculty at the University of Twente. Data collection occurred between March 19, 2024, and May 13, 2024, involving 34 participants. The participants were university students aged 18–24, with a gender distribution of 37% men and 63% women. Recruitment was conducted via university-wide announcements, and all participants provided informed consent prior to their involvement in the study. The data collection, participant recruitment, and study design were managed by the original research team from this department, and the author of this study was not involved in the data collection process.

4.1.2 Data Collection Procedure

Participants were equipped with Empatica Embrace Plus wearable devices[42], which continuously monitored their physiological signals and activity patterns throughout the study period of 8 days. The devices recorded multiple modalities, including: **HRV:** Capturing cardiovascular activity, **EDA:** Measuring skin conductance as an indicator of emotional arousal, **ACC:** Detecting movement and activity levels. Participants were instructed to wear the smartwatch on their non-dominant hand to minimize interference with daily activities while ensuring consistent data collection. However, individual differences in variations in movement patterns could introduce inconsistencies in accelerometer (ACC) measurements. Given that ACC-derived features are highly sensitive to motion artifacts, variations in wrist positioning may affect feature accuracy, particularly for activity intensity and entropy-based movement metrics.

All signals were recorded at a sampling frequency of 64 Hz.

4.1.3 Psychological Assessments

In addition to physiological data, participants completed daily self-report surveys assessing their psychological states. These surveys measured constructs such as stress levels, positive and negative emotions, and overall well-being, using validated psychological instruments. Data collection was conducted using the M-Path app [43], with participants completing the surveys at the end of each day.

During the intake survey, participants selected time slots that best fit their schedules, ranging between 8 and 11 hours. They were given a 1-hour window to complete the survey. If they had not filled in the ecological momentary assessment (EMA) within the first 30 minutes, they received a reminder to ensure timely completion. Table 4.1 provides an overview of the survey questions and corresponding scales. Each question is answered using a discrete scale from 0 to 10, where 0 represents 'strongly disagree' and 10 represents 'totally agree'. These questions are designed to capture the different domains of the PERMA model.

TABLE 4.1: Survey Questions with Codes and Label

Label	**Code**	**Question**
0	P1	I felt joyful today.
1	N1	I felt anxious today.
2	P2	I felt positive today.
3	N2	I felt angry today.
4	N3	I felt sad today.
5	P3	I felt content today.
6	SE1	On the whole, I was satisfied with myself today.
7	SE2	I was inclined to feel like a failure today.
8	SE3	I certainly felt useless today.
9	SE4	I feel that I had a number of good qualities today.
10	MIL1	Today I understood my life's meaning.
11	MIL2	Today I was searching for meaning in life.
12	MIL3	Today my life was full of value.
13	MIL4	Today I was highly committed to certain goals in my life.
14	MIL5	Today I could comprehend what my life is all about.
15	PRO1	Today I received help and support from others when I needed it.
16	PRO2	Today I felt loved.
17	PRO3	Today I was satisfied with my personal relationships.

The dataset includes 272 days of data, amounting to 6528 hours of recordings, with each participant contributing 192 on average. In addition to physiological data, the dataset contains metadata such as timestamps, gender, nationality.etc. To ensure privacy and ethical

compliance, all data was anonymized, with participant IDs replacing personal identifiers. Figure4.1 shows the distribution of answers as labels.

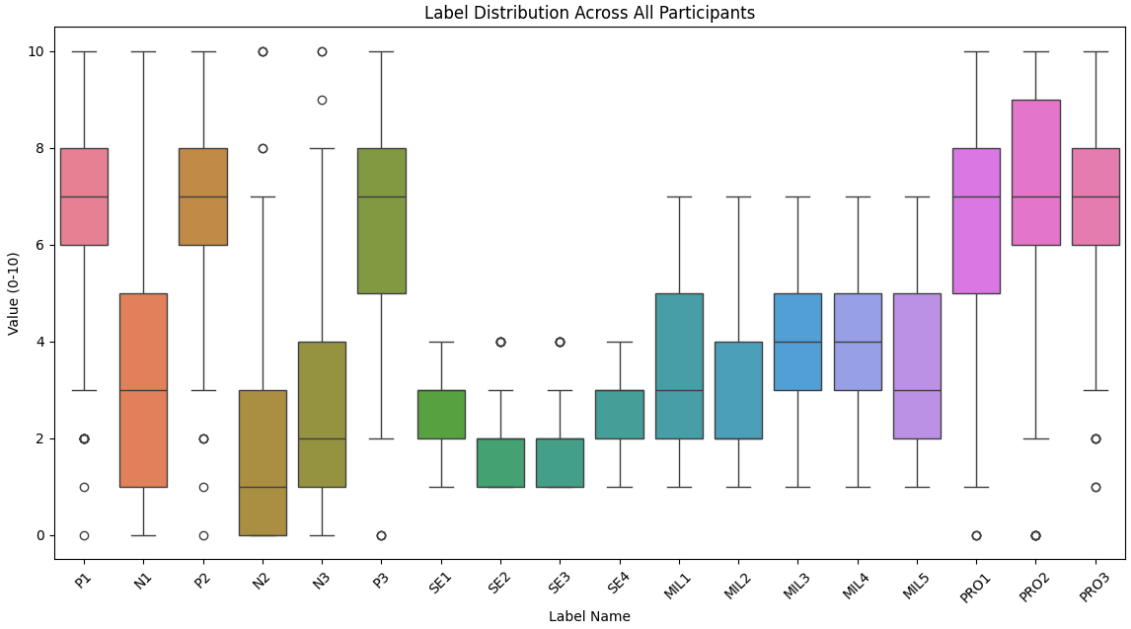

FIGURE 4.1: Label Distribution of answers to questions in the survey, range from 0 to 10

4.2 Data Processing

To transform the raw wearable data into machine-learning-ready features, a structured pipeline was implemented, as illustrated in Figure 4.2. The steps involved in this process are described below.

4.2.1 Data preprocessing

This section describes the signal processing workflow implemented in Python for extracting physiological data from `.avro` files. The data collected by the embraceplus can be downloaded from the cloud as .avro files [44]. The data consists of multiple physiological modalities such as ACC, gyroscope, EDA, temperature, and more. To illustrate the nature of these signals, figure 4.3 shows a 1-day segment of raw data for each modalityThe following subsections explain the methods used for data extraction and processing for each modality. The *Empatica Documentation* [45]was consulted to ensure the data adhered to the expected standards and formats for physiological signal processing.

Directory Structure and File Handling A script was developed to traverse the directory structure containing participant data files. To prevent redundant computations, the script maintains a log file to track previously processed files. The processed data for each participant is systematically stored in a structured output directory. For each `.avro` file, the participant ID and date are extracted from the directory structure. Additionally, the script dynamically generates an output directory to organize the results.

Accelerometer Signal Processing The accelerometer data consists of raw digital values that are converted into physical units (g-force). The timestamps for each data point are

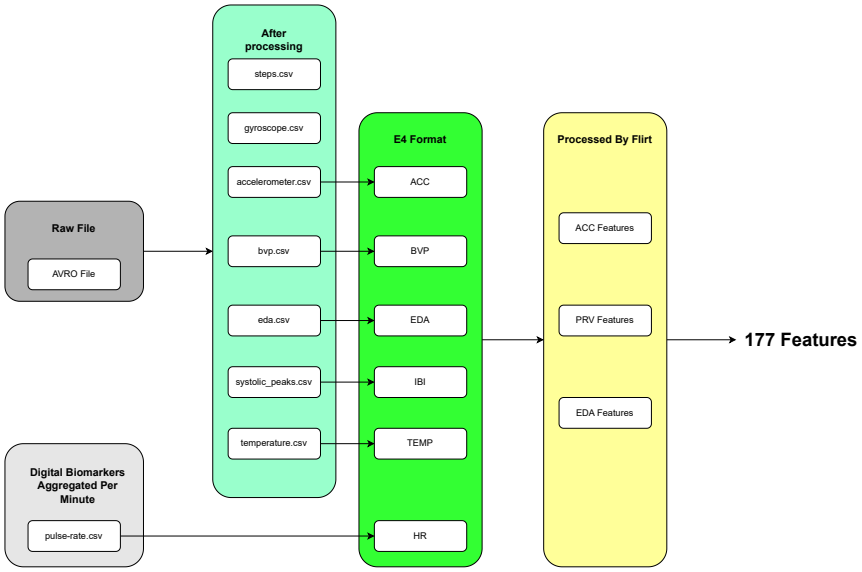

FIGURE 4.2: Workflow for Data Preprocessing and Feature Engineering. Raw data was processed, transformed to E4 format, and feature-engineered using the FLIRT toolkit.

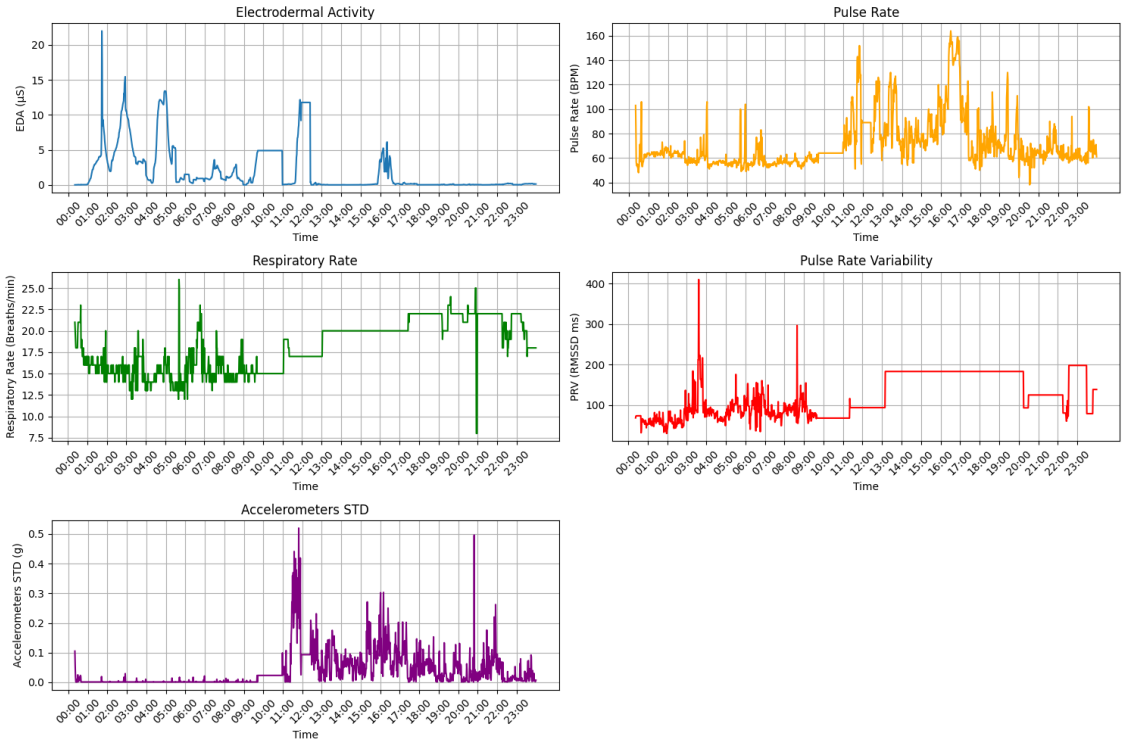

FIGURE 4.3: Example raw data for each modality

computed as:

$$t_i = t_{\text{start}} + i \cdot \frac{10^6}{f_s}, \tag{4.1}$$

where t_i is the timestamp of the i-th data point, t_{start} is the initial timestamp, f_s is the sampling frequency in Hz. The raw digital values $(x_{\text{digital}}, y_{\text{digital}}, z_{\text{digital}})$ are converted to physical units using the following transformation:

$$x_g = x_{\text{digital}} \cdot \frac{\Delta_{\text{physical}}}{\Delta_{\text{digital}}}, \quad y_g = y_{\text{digital}} \cdot \frac{\Delta_{\text{physical}}}{\Delta_{\text{digital}}}, \quad z_g = z_{\text{digital}} \cdot \frac{\Delta_{\text{physical}}}{\Delta_{\text{digital}}}, \tag{4.2}$$

where $\Delta_{\text{physical}} = \text{physicalMax} - \text{physicalMin}$, represents the sensor's measurement range in g-force, and $\Delta_{\text{digital}} = \text{digitalMax} - \text{digitalMin}$, corresponds to the range of digital values determined by the sensor's resolution.

Gyroscope Signal Processing The gyroscope data is processed in a similar manner to the accelerometer data, converting digital values into physical units measured in degrees per second (dps):

$$x_{\text{dps}} = x_{\text{digital}} \cdot \frac{\Delta_{\text{physical}}}{\Delta_{\text{digital}}}, \quad y_{\text{dps}} = y_{\text{digital}} \cdot \frac{\Delta_{\text{physical}}}{\Delta_{\text{digital}}}, \quad z_{\text{dps}} = z_{\text{digital}} \cdot \frac{\Delta_{\text{physical}}}{\Delta_{\text{digital}}}. \tag{4.3}$$

The timestamps are calculated using:

$$t_i = t_{\text{start}} + i \cdot \frac{10^6}{f_s}. \tag{4.4}$$

EDA Processing The EDA values are extracted as-is, paired with timestamps computed as:

$$t_i = t_{\text{start}} + i \cdot \frac{10^6}{f_s}. \tag{4.5}$$

Temperature Signal Processing Temperature data is processed similarly to EDA data, where raw values are directly mapped to timestamps using the formula:

$$t_i = t_{\text{start}} + i \cdot \frac{10^6}{f_s}. \tag{4.6}$$

Blood Volume Pulse (BVP) and Systolic Peaks Processing The BVP values are extracted and paired with timestamps using the same timestamp formula. Systolic peak timestamps are stored directly from the raw data without further transformations. Step counts are processed similarly to other modalities. The raw tag timestamps are stored without additional modifications. The processed data is saved separately for each modality, such as: ACC data, gyroscope data, and EDA activity. Processed files are logged to ensure that they are not reprocessed. To ensure robust execution, exceptions during file processing are caught and logged, allowing the processing workflow to continue with other files.

Figure 4.4 shows the workflow of processing raw data before making adjustments for FLIRT toolkit.

FIGURE 4.4: Workflow for Data Preprocessing from Raw data before further processing for FLIRT [8] toolkit.

The overall workflow can be summarized as first, convert raw digital values into meaningful physical units, then compute precise timestamps based on sampling frequency and start time. The processed data is saved in a structured format for subsequent analysis. Finally individual sensor streams were extracted from the AVRO files and saved as regular files, including Step counts, Gyroscopic motion data, ACC data, BVP data, EDA data, detected systolic peaks for calculating inter-beat intervals (IBI), Skin temperature. The inter-beat intervals (IBI) were computed by taking the difference between consecutive systolic peaks. PRV filtering was optionally applied for further refinement and pulse-rate aggregated digital biomarkers (e.g., heart rate) per minute.

4.2.2 Adjustments for FLIRT Toolkit Compatibility

The FLIRT feature extraction toolkit is optimized for Empatica E4 data formats, which is the previous generation of Empatica product family and does not natively support the latest Empatica Embrace Plus model. Therefore, adjustments were necessary to ensure the processed files adhered to the FLIRT-compatible input format. The processed data files were iteratively traversed from the base directory with each file's contents reformatted and saved into a structured directory. Each data modality was processed using the following pipeline: Conversion of timestamps to the required format. Transformation of raw values to adhere to FLIRT's input specifications. Logging and error handling for tracking progress.

EDA Processing EDA data required timestamps in seconds and values rounded to six decimal places. The processing formula is:

$$t_{\text{initial}} = \left\lfloor \frac{t_{\text{unix}}^{(0)}}{10^6} \right\rfloor, \tag{4.7}$$

where $t_{\text{unix}}^{(0)}$ is the first timestamp. The formatted data for FLIRT is structured as:

$$\text{EDA\_formatted} = \{t_{\text{initial}}, 4, \text{EDA values rounded to 6 decimal places}\}. \tag{4.8}$$

Blood Volume Pulse (BVP) Processing BVP data were processed with timestamps converted to seconds and formatted with the header:

$$\text{BVP\_formatted} = \{t_{\text{initial}}, 64, \text{BVP values}\}. \tag{4.9}$$

Temperature Data Processing Temperature data required a similar format as EDA:

$$\text{TEMP\_formatted} = \{t_{\text{initial}}, 4, \text{Temperature values}\}. \tag{4.10}$$

ACC Data Processing The accelerometer values were scaled to fit within the range $[-128, 127]$ using min-max scaling:

$$\text{ACC\_scaled} = \left\lfloor 127 \cdot \frac{\text{ACC} - \min(\text{ACC})}{\max(\text{ACC}) - \min(\text{ACC})} \right\rfloor, \tag{4.11}$$

and formatted as:

$$\text{ACC\_formatted} = \{\text{Header: } t_{\text{initial}}, \text{Sampling rate: } 64, \text{Scaled ACC values}\}. \tag{4.12}$$

Inter-Beat Interval (IBI) Data Processing The systolic peaks were used to compute inter-beat intervals (IBI) as:

$$\text{IBI} = \frac{t_{\text{current}} - t_{\text{previous}}}{10^3}, \tag{4.13}$$

where t_{current} and t_{previous} are consecutive systolic peak timestamps in milliseconds. The data were structured with relative timestamps:

$$\text{IBI\_formatted} = \{\text{Relative time}, \text{IBI values}\}. \tag{4.14}$$

Pulse Rate (PR) Processing Pulse rate values were scaled and formatted as:

$$\text{PR\_formatted} = \{t_{\text{initial}}, 1, \text{Pulse rate values}\}. \tag{4.15}$$

Output and Compression The processed data for each participant and modality were stored in respective subdirectories. These directories were then compressed into archives for efficient handling and transfer.

Error Handling and Logging Any processing errors were logged to facilitate debugging and ensure uninterrupted workflow execution.

The modifications ensure that the processed data can match the FLIRT input requirements, also retain data fidelity during transformations and allow for seamless feature extraction using FLIRT.

4.3 Feature Extraction Using FLIRT

The *Feature generation toolkit for wearable data* (FLIRT) was utilized to extract a comprehensive set of features from the processed data. FLIRT uses an overlapping sliding-window approach to generate machine-learning-ready features. Key feature categories include:

ACC Features: Statistical and time-series features derived from accelerometer data, capturing movement intensity and patterns.

PRV Features: Features derived from IBI data, providing metrics for pulse rate variability, which serves as a proxy for heart rate variability.

EDA Features: Statistical features calculated from electrodermal activity, which reflect physiological responses such as stress and arousal.

Table 2 presents the features generated by FLIRT based on various physiological signals. In total, 178 features are extracted, aggregated over a one-hour timestep and time window.

4.4 Dataset preparation

The generated features was further processed to ensure consistency and readiness for further analysis. The preprocessing steps included the removal of corrupted entries, interpolation of missing values, and standardization where necessary. The detailed process is outlined below.

Missing values were addressed in a multi-step approach:

Removing Columns with Excessive Missing Values Columns with more than 90% missing values were removed:

$$C_{\text{drop}} = \{c \in C \mid \text{missing}(c) > 0.9 \times N\}, \tag{4.16}$$

where C is the set of columns, N is the number of rows, and missing(c) represents the count of missing values in column c.

Imputing Remaining Missing Values Missing values were replaced with the column-wise median:

$$x_{i,j} = \begin{cases} x_{i,j}, & \text{if } x_{i,j} \neq \text{NaN}, \\ \text{median}(x_{\cdot,j}), & \text{if } x_{i,j} = \text{NaN}, \end{cases} \tag{4.17}$$

where $x_{i,j}$ denotes the i-th row and j-th column of the dataset.

Handling Infinite Values Infinite values ($+\infty$ and $-\infty$) were replaced with NaN for uniform handling.

Ensuring Consistent Data Types To maintain uniformity, columns representing timestamps were converted to a standardized datetime format:

$$t_{\text{converted}} = \text{to\_datetime}(t_{\text{original}}), \tag{4.18}$$

where to_datetime($\cdot$) ensures compatibility across different formats.

Combining Data Across Participants Each participant's data file was augmented with an identifier derived from the file name:

$$\text{participant\_id} = \text{file\_name}_{\text{base}}. \tag{4.19}$$

The cleaned data from all participants was then concatenated into a unified dataset:

$$D_{\text{combined}} = \bigcup_{i=1}^{M} D_i, \tag{4.20}$$

where D_i is the cleaned dataset for participant i and M is the total number of participants.

Data Integrity Checks To ensure data quality, the combined dataset was examined for:

Residual Missing Values:

$$\text{missing\_summary} = \{c \in C_{\text{combined}} \mid \text{missing}(c) > 0\}. \tag{4.21}$$

Residual Infinite Values:

$$\text{inf\_check} = \sum_{i=1}^{N} \sum_{j=1}^{C} 1(x_{i,j} = \infty \vee x_{i,j} = -\infty), \tag{4.22}$$

where $1(\cdot)$ is the indicator function.

The preprocessing steps can be summarized as follows: Removed corrupted columns with excessive missing values. Imputed remaining missing values using the median. Combined all participant data into a unified dataset. Verified data integrity and addressed residual issues. The preprocessed dataset was saved to ensure it was ready for subsequent analysis steps. This workflow improved data quality while facilitating reliable and reproducible analysis.

After preprocessing the raw dataset, the data was split into three versions for analysis:

Original Dataset: Retained the original features without further transformations.

Scaled Dataset: Standardized the features to have zero mean and unit variance:

$$z_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j}, \tag{4.23}$$

where $x_{i,j}$ is the original value, μ_j is the mean, and σ_j is the standard deviation of the j-th feature.

PCA Dataset: Applied Principal Component Analysis (PCA) to the already scaled dataset to reduce dimensionality, retaining 95% of the explained variance:

$$\mathbf{X}_{\text{PCA}} = \mathbf{X} \cdot \mathbf{W}, \tag{4.24}$$

where $\mathbf{W}$ is the matrix of principal components.

The dataset was divided into training and testing subsets for each label. To ensure no data leakage between participants, a grouped split was applied based on participant IDs. This guarantees that data from the same individual does not appear in both training and testing sets

4.5 Feature Selection

The extracted features were examined for relationships and potential redundancies. Correlation analysis and exploratory data analysis (EDA) were conducted to understand the interrelationships among the features, ensuring the dataset's quality for downstream machine learning tasks.

4.6 Modelling

This section describes the experimental methodology for training and evaluating machine learning models to predict each label in the dataset. The experiments were conducted in three stages: dataset preparation, model training, and result comparison.

Models and Training Procedure Various kinds of machine learning models were trained separately in the dataset:

Random Forest (RF): A tree-based ensemble model used for classification tasks [46].

Convolutional Neural Network (CNN): A deep learning model designed to learn spatial hierarchies in the feature space:

$$f(\mathbf{x}) = \text{softmax}(\mathbf{W}_3 \cdot \text{ReLU}(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3), \tag{4.25}$$

where $\mathbf{W}_i$ and $\mathbf{b}_i$ are the weights and biases of the i-th layer [47].

Long Short-Term Memory (LSTM): A recurrent neural network capable of capturing temporal dependencies in sequential data:

$$h_t = \sigma(W_x \cdot x_t + W_h \cdot h_{t-1} + b), \tag{4.26}$$

where h_t is the hidden state at time t, and σ is the activation function [48].

Transformer: A deep learning architecture based on self-attention mechanisms, enabling efficient modeling of long-range dependencies in sequential data:

$$\mathbf{Z} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}, \tag{4.27}$$

where $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ represent the query, key, and value matrices, respectively, and d_k is the dimensionality of the keys [49].

XGBoost (XGB): A gradient boosting framework designed for high-performance tree-based learning tasks:

$$F(x) = \sum_{m=1}^{M} f_m(x), \quad f_m \in \mathcal{F}, \tag{4.28}$$

where each $f_m(x)$ is a regression tree and $\mathcal{F}$ is the space of all possible trees [50].

LightGBM (LGBM): A gradient boosting algorithm optimized for efficiency and scalability using a histogram-based learning approach:

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{4.29}$$

where $l(y_i, \hat{y}_i)$ is the loss function, and $\Omega(f_k)$ represents the regularization term for tree complexity [51].

Table 4.3 and table 4.2 displays the hyperparameter and structure of the models.

Model/Algorithm	Component	Hyperparameter/Setting	Value
Transformer Model	Input	Input shape	`input_shape`
	Self-Attention	num_heads	4
	Self-Attention	key_dim	64
	Attention Dropout	dropout rate	0.3
	Layer Normalization	epsilon	1×10^{-6}
	Feed–Forward Dense 1	units	128
	Feed–Forward Dense 1	activation	ReLU
	Feed–Forward Dense 1	kernel_regularizer (l2)	0.01
	FFN Dropout	dropout rate	0.3
	Feed–Forward Dense 2	units	64
	Feed–Forward Dense 2	activation	ReLU
	Feed–Forward Dense 2	kernel_regularizer (l2)	0.01
	Output Layer	units	5
	Output Layer	activation	Linear
	Compilation	Optimizer (Adam, lr)	0.0003
	Compilation	Loss	MAE
GRU–Based Model	Masking Layer	mask_value	0.0
	Masking Layer	input_shape	`input_shape`
	GRU Layer	units	64
	GRU Layer	return_sequences	False
	GRU Layer	kernel_regularizer (l2)	0.005
	Layer Normalization	–	Default
	Dropout 1	dropout rate	0.3
	Dense Layer	units	32
	Dense Layer	activation	ReLU
	Dense Layer	kernel_regularizer (l2)	0.005
	Dropout 2	dropout rate	0.4
	Output Dense	units	`len(LABEL_COLUMNS)`
	Output Dense	activation	Linear
	Compilation	Optimizer (Adam, lr)	0.0001
	Compilation	Loss	MAE
CNN	Input	input_shape	(`X_train.shape[1]`, 1)
	Conv1D Layer 1	filters	64
	Conv1D Layer 1	kernel_size	3
	Conv1D Layer 1	activation	ReLU
	Conv1D Layer 2	filters	32
	Conv1D Layer 2	kernel_size	3
	Conv1D Layer 2	activation	ReLU
	Flatten	–	–
	Dense Layer	units	128
	Dense Layer	activation	ReLU
	Dropout	dropout rate	0.3
	Output Dense	units	11
	Output Dense	activation	Softmax
	Compilation	Optimizer (Adam, lr)	0.001
	Compilation	Loss	sc_crossentropy
	Training	epochs	10
	Training	batch_size	32

TABLE 4.2: Hyperparameters for Neural Network Models

Model/Algorithm	Component	Hyperparameter/Setting	Value
LSTM	Input	input_shape	(`X_train.shape[1]`, 1)
	LSTM Layer 1	units	64
	LSTM Layer 1	return_sequences	True
	LSTM Layer 2	units	32
	Dense Layer	units	128
	Dense Layer	activation	ReLU
	Dropout	dropout rate	0.3
	Output Dense	units	11
	Output Dense	activation	Softmax
	Compilation	Optimizer (Adam, lr)	0.001
	Compilation	Loss	sc_crossentropy
	Training	epochs	10
	Training	batch_size	32
XGBoost Classifier	General	objective	multi:softmax
	General	num_class	`num_classes`
	General	eval_metric	mlogloss
	General	use_label_encoder	False
	Class Weighting	scale_pos_weight	`class_weights`
	Tree Method	tree_method	gpu_hist
LightGBM Classifier	General	objective	multiclass
	General	num_class	`num_classes`
	Data Imbalance	is_unbalance	True
	Tree Complexity	num_leaves	15
	Tree Complexity	min_data_in_leaf	5
	Tree Depth	max_depth	6
	Learning Rate	learning_rate	0.05
	Feature Fraction	feature_fraction	0.8
	Bagging	bagging_fraction	0.8
	Bagging	bagging_freq	5
	Regularization	lambda_l1	0.5
	Regularization	lambda_l2	0.5
	Parallelization	n_jobs	-1
RandomForest	General	random_state	42
	General	n_jobs	-1

Table 4.3: Hyperparameters for Classical ML Models

4.7 Metrics

For each model, the training set was used to fit the model, while the testing set evaluated performance. Metrics such as accuracy, top-3 accuracy, and threshold accuracy were calculated as follows:

Top-3 Accuracy: The *Top-3 Accuracy* measures the proportion of predictions where the true label y_i is among the top three labels $\{\hat{y}_{i,1}, \hat{y}_{i,2}, \hat{y}_{i,3}\}$ predicted by the model. It is defined as:

$$\text{Top-3 Accuracy} = \frac{1}{N} \sum_{i=1}^{N} 1\left(y_i \in \{\hat{y}_{i,1}, \hat{y}_{i,2}, \hat{y}_{i,3}\}\right), \tag{4.30}$$

where N is the total number of samples. $\hat{y}_{i,1}, \hat{y}_{i,2}, \hat{y}_{i,3}$ are the top three predicted labels for sample i, ranked by confidence and 1(condition) is an indicator function that equals 1 if the condition is true and 0 otherwise.

This metric reflects the model's ability to make correct predictions within its top three guesses, which is particularly useful when the labels represent approximate states or when slight variations in ranking are acceptable.

Threshold Accuracy: The *Threshold Accuracy* evaluates the proportion of predictions $\hat{y}_i$ that fall within a specific tolerance range around the true label y_i. For a label range of $[0, 10]$, with a threshold tolerance of ± 1, it is defined as:

$$\text{Threshold Accuracy} = \frac{1}{N} \sum_{i=1}^{N} 1\left(|\hat{y}_i - y_i| \leq 1\right), \tag{4.31}$$

where $|\hat{y}_i - y_i| \leq 1$ ensures the prediction $\hat{y}_i$ is within one unit of the true label y_i. All other terms are as previously defined. This metric is meaningful in scenarios where predictions close to the true label are practically as valuable as exact matches. For example, predicting a "stress level" of 5 instead of the true value of 6 might still provide actionable insights.

Each model's weights were saved for reproducibility. By evaluating both *Top-3 Accuracy* and *Threshold Accuracy*, we assess the model's ability to make accurate ranked predictions (*Top-3 Accuracy*) and predict labels within a reasonable tolerance range (*Threshold Accuracy*).

4.8 Interpretability

To enhance the interpretability of the machine learning models, advanced methods from Explainable AI (XAI) were employed. These methods enable a deeper understanding of model behavior and the contributions of individual features:

MAPIE (Mean and Prediction Interval Estimation)[52]: This technique was utilized to quantify the uncertainty of predictions by providing confidence intervals, thereby enhancing the reliability of the model's outputs.

LIME (Local Interpretable Model-Agnostic Explanations)[53]: LIME was applied to interpret individual predictions by approximating the complex model with a locally

interpretable surrogate linear model. This allowed for instance-specific explanations of predictions.

SHAP (SHapley Additive exPlanations)[54]: SHAP values were computed to assess the contribution of each feature to the model's predictions. This method provided both global insights into feature importance and local interpretability for individual predictions.

4.9 Machine Learning Method

After data cleaning, standardization was applied to normalize the feature values across the dataset. Principal Component Analysis (PCA) was then performed to reduce dimensionality while retaining the most informative components. To ensure the reliability and robustness of the models, a 5-fold cross-validation strategy was employed.

Subsequently, five machine learning models—1D Convolutional Neural Network (1D-CNN), 1D Long Short-Term Memory (1D-LSTM), LightGBM (LGBM), Random Forest, and XGBoost—were implemented and evaluated across three datasets. The input data consisted of one-dimensional arrays representing all extracted features within an hourly window. The task was formulated as a classification problem, where each response was categorized into one of 11 classes, corresponding to integer values ranging from 0 to 10. Each of the 18 survey questions was treated as a separate prediction target, with models trained and evaluated independently for each label. To align the question responses with the extracted hourly features, answers from the same day were assigned to the final timestamp within the corresponding sequence. However, due to the nature of the problem, treating each one-dimensional feature array within an hourly window as an independent input to predict the same daily label led to the loss of crucial temporal information. To address this issue, the features were subsequently transformed into two-dimensional arrays, preserving their original temporal structure. Padding was applied to ensure uniform sequence lengths across all samples.

To further explore the problem and enhance prediction accuracy and reliability, we refined our approach by restructuring the label space and modifying the prediction task. The original 18 labels were clustered into five broader categories based on their underlying characteristics: Positive Emotions, Negative Emotions, Self-Esteem, Meaning in Life, and Social Support. This transformation shifted the task from multi-class classification to regression on the clustered values. Additionally, the 1D feature representations were reformulated into 2D sequences with temporal dependencies. To capture these sequential patterns effectively, we implemented and evaluated advanced deep learning architectures, including Transformers and LSTMs. The following sections present the evaluation metrics and confusion matrices for these models.

Chapter 5

Results

5.1 Feature Analysis

5.1.1 t-SNE Projection of Features by Sensor Modality

Figure 5.1 presents a t-distributed Stochastic Neighbor Embedding (t-SNE) visualization of the extracted features, where each point represents a feature vector from the dataset. The points are color-coded based on the dominant sensor modality:

- **Red (ACC - Accelerometer)**
- **Blue (HRV - Heart Rate Variability)**
- **Green (EDA - Electrodermal Activity)**

The distribution of feature points provides insights into the clustering behavior of different sensor modalities:

- **ACC features (red)** dominate the feature space, suggesting that the majority of extracted features are derived from accelerometer data.
- **EDA features (green)** form a distinct cluster in the upper-right region, indicating that they capture unique information different from ACC and HRV.
- **HRV features (blue)** are scattered throughout the space, with some overlap with both ACC and EDA, suggesting that HRV-based features share similarities with the other modalities but do not form a strongly independent group.

This visualization helps to understand the relationship between different sensor modalities and how distinct or overlapping their feature representations are.

5.1.2 Correlation Between Sensor Modalities

Figure 5.2 illustrates the mean correlation coefficient between features derived from different sensor modalities. The following modality pairs are analyzed:

- **HRV vs. ACC (blue bar)**: Displays a relatively low correlation, suggesting that HRV and accelerometer features capture largely independent information.
- **EDA vs. ACC (gray bar)**: Shows the highest correlation, indicating that electrodermal activity and accelerometer data share some common patterns, possibly due

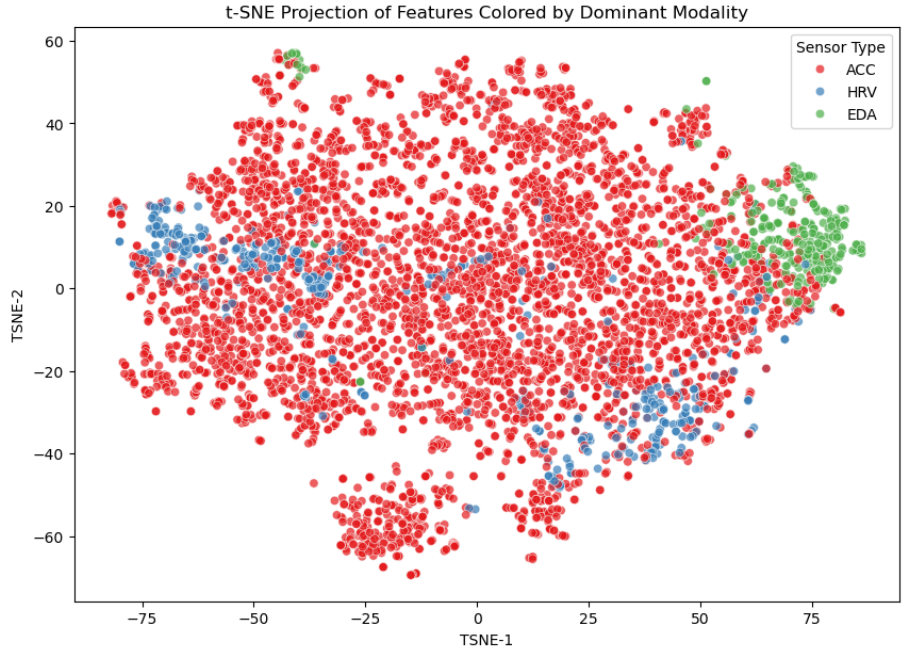

FIGURE 5.1: t-SNE Projection of Features Colored by Dominant Modality

to movement-related artifacts in EDA signals.

- **HRV vs. EDA (red bar)**: Exhibits the lowest correlation, confirming that HRV and EDA features are the most independent, as expected given their distinct physiological processes.

This correlation analysis helps determine the level of redundancy or complementarity among sensor modalities. The low correlation between HRV and EDA suggests that combining them in predictive models may provide richer and more diverse information, improving classification or regression tasks. Conversely, the higher correlation between EDA and ACC suggests that these modalities might contribute redundant information if not handled properly.

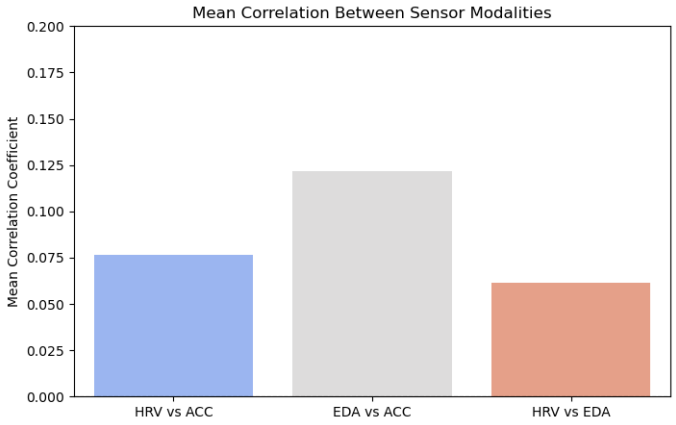

FIGURE 5.2: Mean Correlation Between Sensor Modalities

Key Observations

- The t-SNE visualization shows that ACC features dominate the dataset, while EDA forms a distinct cluster, and HRV features are more dispersed.
- The correlation analysis indicates that EDA and ACC features have the strongest relationship, while HRV and EDA features are the most independent.
- These findings suggest that combining features from different sensor modalities, particularly HRV and EDA, may provide complementary information and improve model performance.

To identify the most relevant features for predicting the target variable, multiple feature selection techniques were employed, including Recursive Feature Elimination (RFE), ANOVA, Random Forest feature importance, LASSO regression, and mutual information. These methods were combined to select the top 20 features based on their consistent importance across the techniques.

Figure 5.4 5.3 5.5 5.6 5.7 presents the top 20 features identified through this process, highlighting the features deemed most informative by the majority of the methods used.

Table 5.1 presents the feature importance values of different physiological modalities (HRV, ACC, and EDA) across five psychological factors: Positive Emotions, Negative Emotions, Self-Esteem, Meaning in Life, and Social Support.The importance of each feature is computed using three methods:

- **Pearson Correlation:** Computes the absolute correlation between each feature and the target variable to measure linear relationships.
- **Mutual Information (MI):** Estimates nonlinear dependency using mutual information regression.
- **Random Forest Importance:** A Random Forest Regressor (20 trees, max depth 5) is trained, and feature importance is derived from the average decrease in prediction error when a feature is used in decision trees.

Modality-level importance is obtained by summing the feature importances within each sensor modality (HRV, ACC, EDA).

Key Observations:

- **ACC (Accelerometer)** consistently has the highest feature importance across all psychological factors, indicating that movement-related features play a crucial role in predicting these states. The highest importance value is observed for Self-Esteem (0.801596).
- **HRV (Heart Rate Variability)** shows moderate importance across all factors, with its highest contribution to Social Support (0.164899). This suggests that heart rate variability is somewhat relevant for psychological well-being but less dominant compared to ACC.
- **EDA (Electrodermal Activity)** has relatively lower importance values, except for Meaning in Life (0.171652), where it surpasses HRV. This may indicate that skin conductance responses are more related to deeper cognitive and emotional processing.
- Overall, ACC dominates in all psychological aspects, while HRV and EDA contribute more selectively depending on the specific factor.

These results highlight the significance of movement-related data in psychological prediction while also suggesting that physiological signals such as HRV and EDA may be useful in complementary ways.

Modality	Pos. Emotions	Neg. Emotions	Self-Esteem	Meaning	Social Supp.
HRV	0.1313	0.1312	0.1166	0.1052	0.1649
ACC	0.7565	0.7486	0.8016	0.7229	0.6952
EDA	0.1061	0.1183	0.0790	0.1717	0.1374

TABLE 5.1: Modality-Level Feature Importance

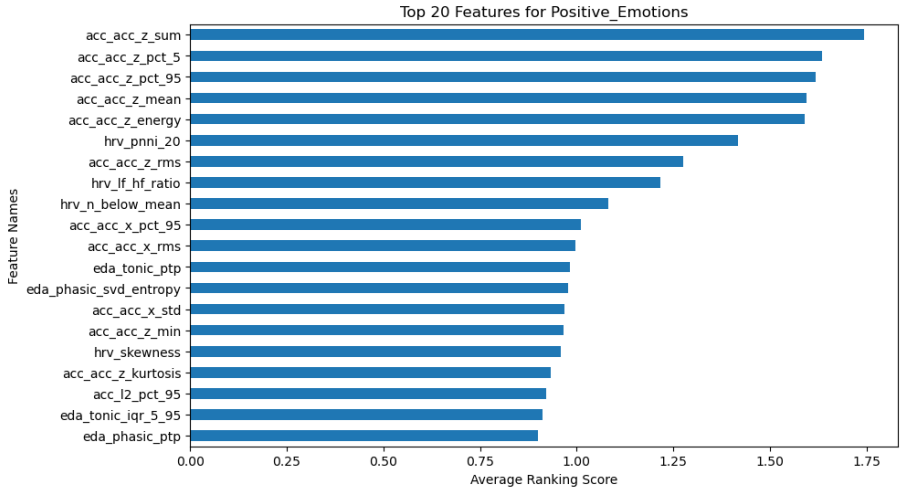

FIGURE 5.3: Top 20 features selected for Positive_Emotions

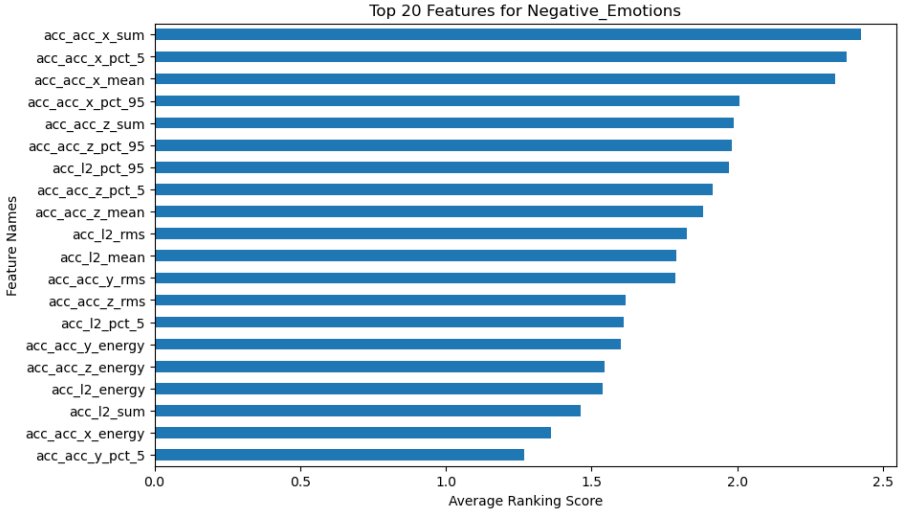

FIGURE 5.4: Top 20 features selected for Negative_Emotions

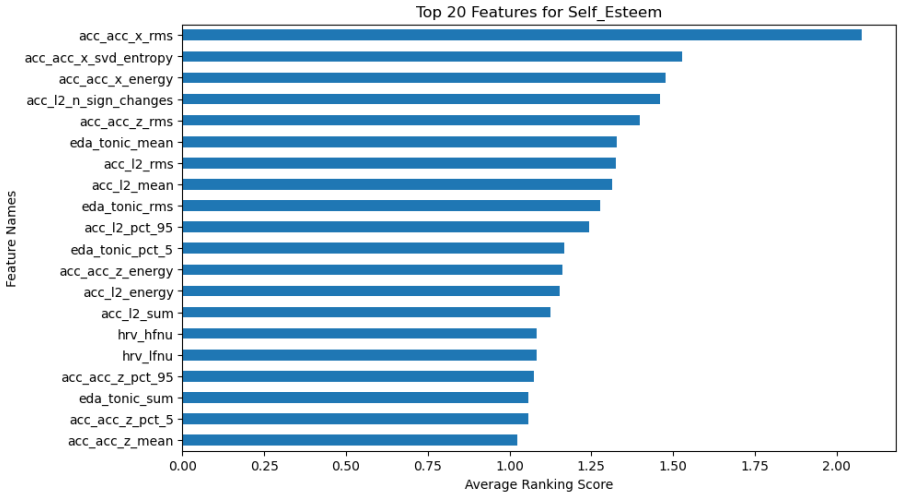

FIGURE 5.5: Top 20 features selected for Self_Esteem

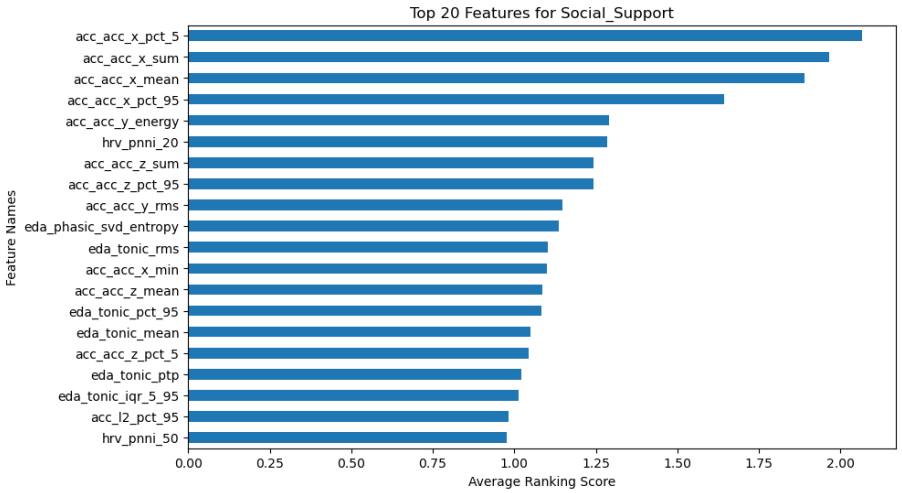

FIGURE 5.6: Top 20 features selected for Social_Support

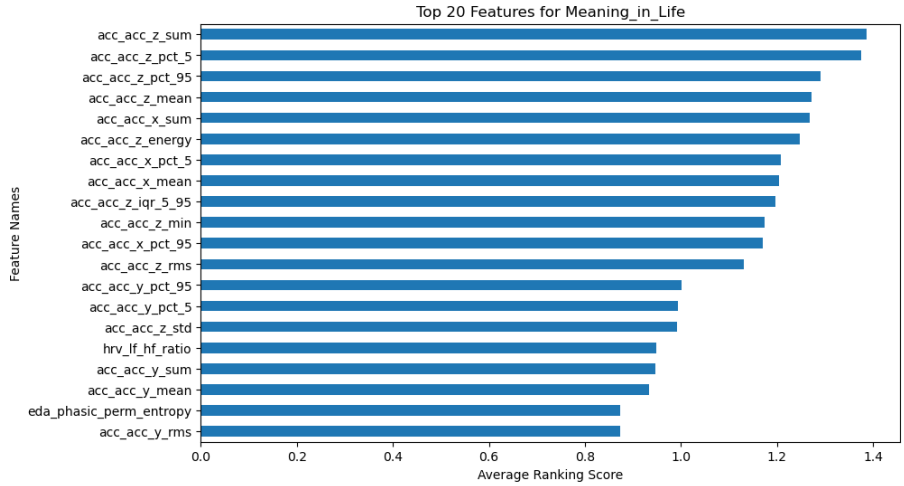

FIGURE 5.7: Top 20 features selected for Meaning_In_Life

5.2 Model Performance

Model and Dataset Comparison

Table 5.2 and 5.3 shows the performance comparison between models and datasets.

TABLE 5.2: Performance Metrics (Accuracy, Precision, Recall)

Model	Dataset	Accuracy	Precision	Recall
CNN	original	0.2795 ± 0.0359	0.1154 ± 0.0319	0.2795 ± 0.0359
CNN	pca	0.2200 ± 0.0194	0.2416 ± 0.0319	0.2200 ± 0.0194
CNN	scaled	0.2169 ± 0.0221	0.2442 ± 0.0358	0.2169 ± 0.0221
LGBM	original	0.2305 ± 0.0259	0.2427 ± 0.0308	0.2305 ± 0.0259
LGBM	pca	0.2325 ± 0.0257	0.2401 ± 0.0295	0.2325 ± 0.0257
LGBM	scaled	0.2304 ± 0.0250	0.2437 ± 0.0329	0.2304 ± 0.0250
LSTM	original	0.2598 ± 0.0334	0.2013 ± 0.0263	0.2598 ± 0.0334
LSTM	pca	0.2674 ± 0.0305	0.1321 ± 0.0219	0.2674 ± 0.0305
LSTM	scaled	0.2643 ± 0.0321	0.1581 ± 0.0346	0.2643 ± 0.0321
Random Forest	original	0.2429 ± 0.0299	0.2399 ± 0.0350	0.2429 ± 0.0299
Random Forest	pca	0.2566 ± 0.0286	0.2316 ± 0.0305	0.2566 ± 0.0286
Random Forest	scaled	0.2442 ± 0.0310	0.2442 ± 0.0359	0.2442 ± 0.0310
XGB	original	0.2295 ± 0.0244	0.2414 ± 0.0310	0.2295 ± 0.0244
XGB	pca	0.2302 ± 0.0224	0.2384 ± 0.0261	0.2302 ± 0.0224
XGB	scaled	0.2299 ± 0.0246	0.2419 ± 0.0335	0.2299 ± 0.0246

TABLE 5.3: Performance Metrics (F1 Score, Threshold Accuracy, Top 3 Accuracy)

Model	Dataset	F1 Score	Threshold Accuracy	Top 3 Accuracy
CNN	original	0.1426 ± 0.0282	0.6112 ± 0.0447	0.6443 ± 0.0465
CNN	pca	0.2147 ± 0.0283	0.5446 ± 0.0224	0.5776 ± 0.0245
CNN	scaled	0.2122 ± 0.0303	0.5396 ± 0.0304	0.5755 ± 0.0313
LGBM	original	0.2137 ± 0.0322	0.5524 ± 0.0356	0.6006 ± 0.0348
LGBM	pca	0.2118 ± 0.0324	0.5529 ± 0.0347	0.5970 ± 0.0310

Model	Dataset	F1 Score	Threshold Accuracy	Top 3 Accuracy
LGBM	scaled	0.2143 ± 0.0322	0.5528 ± 0.0380	0.5992 ± 0.0336
LSTM	original	0.1963 ± 0.0354	0.5910 ± 0.0408	0.6222 ± 0.0369
LSTM	pca	0.1573 ± 0.0241	0.5909 ± 0.0384	0.6352 ± 0.0423
LSTM	scaled	0.1787 ± 0.0368	0.5960 ± 0.0419	0.6318 ± 0.0352
Random Forest	original	0.2124 ± 0.0370	0.5753 ± 0.0404	0.5338 ± 0.0285
Random Forest	pca	0.2099 ± 0.0328	0.5883 ± 0.0377	0.5491 ± 0.0224
Random Forest	scaled	0.2133 ± 0.0379	0.5762 ± 0.0414	0.5338 ± 0.0275
XGB	original	0.2142 ± 0.0308	0.5545 ± 0.0341	0.5934 ± 0.0303
XGB	pca	0.2124 ± 0.0293	0.5565 ± 0.0330	0.5932 ± 0.0345
XGB	scaled	0.2147 ± 0.0307	0.5538 ± 0.0340	0.5925 ± 0.0309

Figures 5.8, 5.9,5.12 5.13, 5.11 and 5.10 illustrate the performance of the Random Forest, LSTM, and CNN models evaluated on three datasets: the original dataset, the dataset with PCA-applied features, and the scaled dataset. The performance was assessed using three metrics: accuracy, Top-3 accuracy, and threshold accuracy.

Accuracy Comparison Figure 5.8 shows the overall accuracy of each model across the datasets. Key observations include that LSTM almost consistently achieved the highest accuracy across all datasets, indicating its ability to model sequential dependencies in the data. Random forest performed closely behind LSTM, suggesting that tree-based methods might be as effective on these datasets or under these preprocessing conditions. CNN displayed slightly lower accuracy compared to the other models.

Top-3 Accuracy Comparison Figure 5.9 presents the Top-3 accuracy for the models. Notable findings are that LSTM maintained its advantage, achieving almost the highest Top-3 accuracy across all datasets, emphasizing its robustness in providing accurate predictions among its top three outputs. CNN also performed strongly, demonstrating its ability to rank predictions effectively. Random Forest exhibited a smaller margin between its accuracy and Top-3 accuracy, suggesting a less diverse set of high-confidence predictions compared to the other models.

Threshold Accuracy Comparison Figure 5.10 highlights the threshold accuracy, where predictions within a ± 1 range of the true label are considered correct. Observations include that random forest and LSTM showed similar performance, with both models achieving high threshold accuracy across all datasets. This indicates their capability to make predictions close to the true labels even when exact matches are not achieved. CNN lagged slightly behind, but its performance improved noticeably under this metric compared to standard accuracy, suggesting its predictions are often near the true values.

General Trends Across Datasets The PCA-applied dataset generally resulted in better performance for random forest compared to the original and scaled datasets. This indicates that dimensionality reduction using PCA likely enhanced the quality of the features by removing noise and redundancies. The scaled dataset exhibited comparable performance to the original dataset, suggesting that scaling alone did not significantly alter the predictive performance.

5.2.1 Performance per label

Performance Comparison for Individual Labels

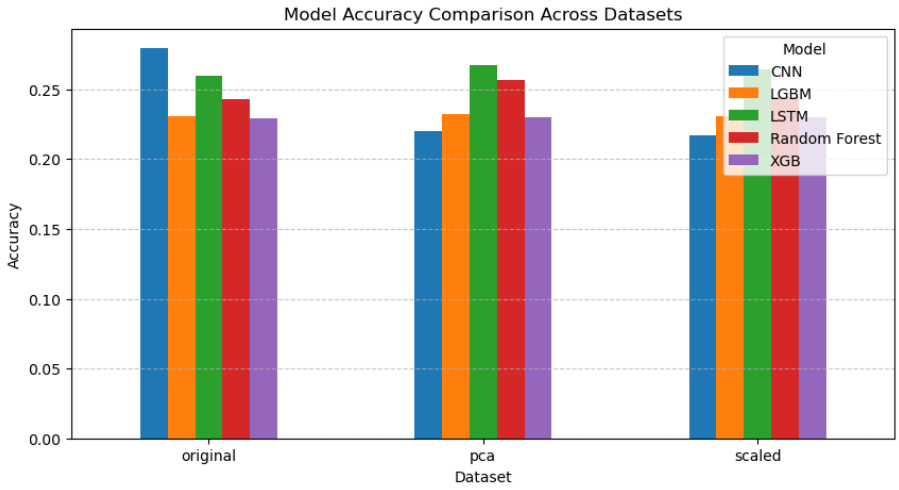

FIGURE 5.8: Accuracy Comparison of Random Forest, LSTM and CNN in 3 datasets

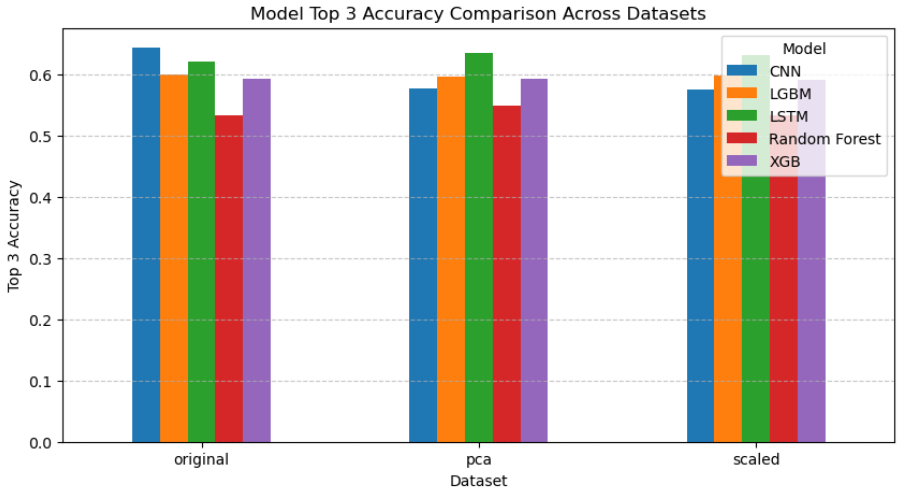

FIGURE 5.9: Top-3 Accuracy Comparison of Models in 3 datasets

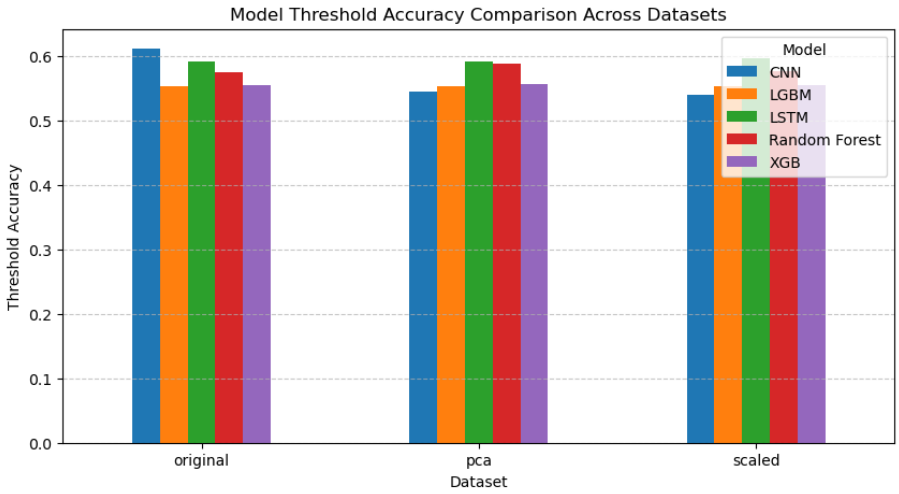

FIGURE 5.10: Threshhold Accuracy Comparison of Models in 3 datasets

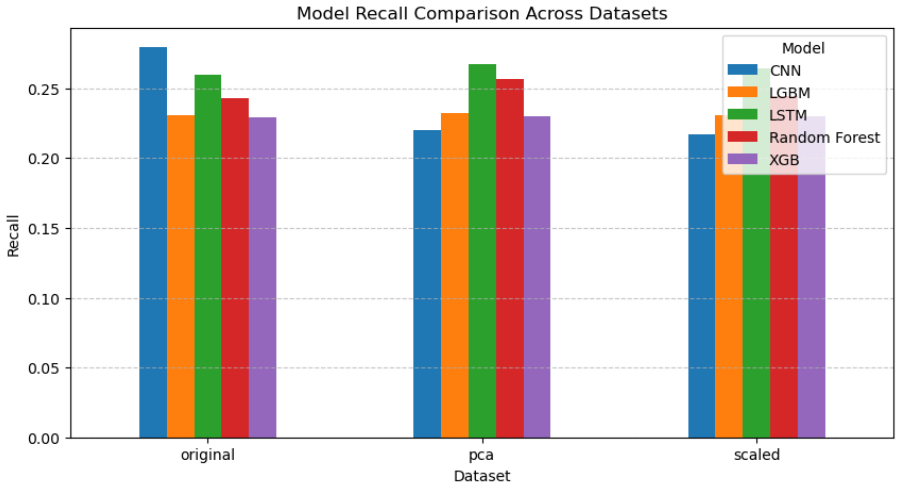

FIGURE 5.11: Recall Comparison of Models in 3 datasets

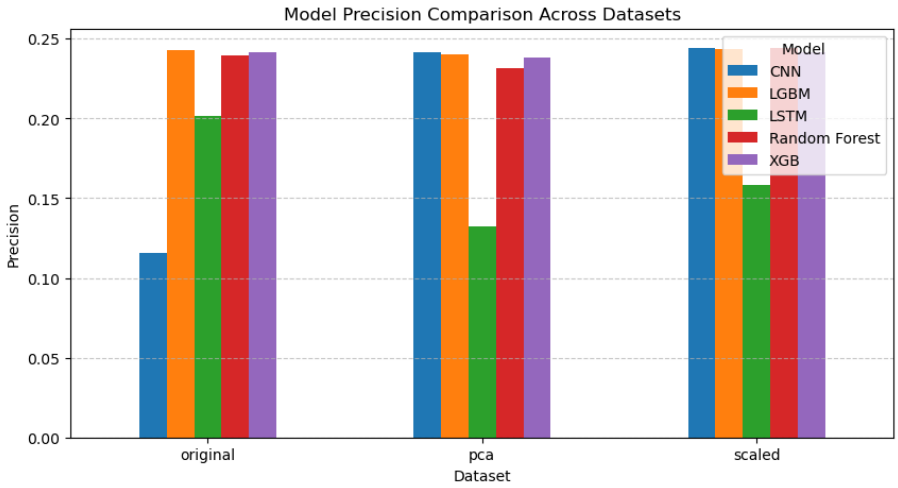

FIGURE 5.12: Precisio Comparison of Models in 3 datasets

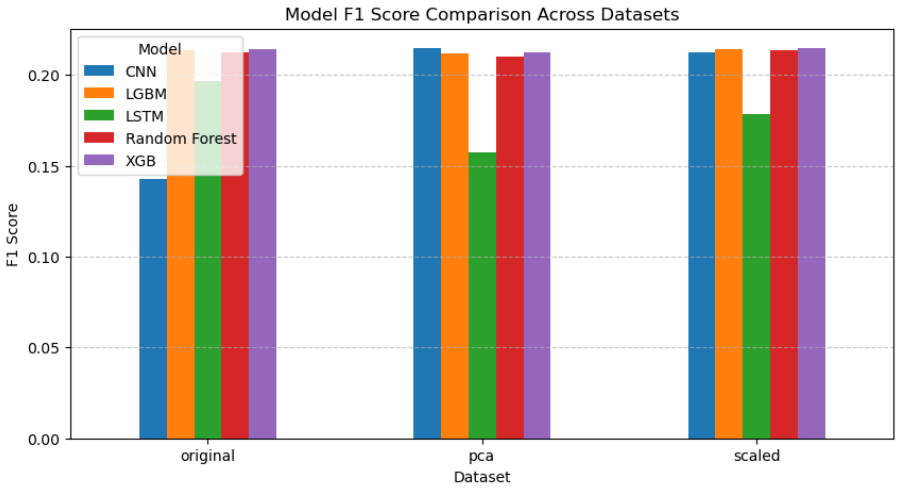

FIGURE 5.13: F1Score Comparison of Models in 3 datasets

Figures 5.19, 5.20, and 5.21 provide a detailed comparison of the performance of Random Forest, LSTM, and CNN models across all labels using three evaluation metrics: threshold accuracy, Top-3 accuracy, and overall accuracy.

Threshold Accuracy Across Labels Figure 5.19 illustrates the threshold accuracy, where predictions are considered correct if they fall within a ± 1 range of the true label. Key insights include that LSTM consistently achieved the highest threshold accuracy for most labels, leveraging its sequential modeling capabilities to generate accurate predictions near the true value. Random forest also performed strongly, achieving similar threshold accuracy to LSTM for several labels, particularly in categories where spatial patterns are critical. CNN exhibited slightly lower threshold accuracy across most labels, suggesting that its feature-splitting mechanism might struggle with certain label-specific complexities.

Top-3 Accuracy Across Labels Figure 5.20 presents the Top-3 accuracy, reflecting the proportion of predictions where the true label is within the top three predictions. Observations include that LSTM achieved high Top-3 accuracy across all labels, with CNN slightly lower, indicating its robustness in ranking predictions effectively. Random Forest showed competitive Top-3 accuracy but lagged behind the other models, especially for labels associated with more nuanced patterns.

Overall Accuracy Across Labels Figure 5.21 highlights the overall accuracy for each label. Key findings are that LSTM demonstrated the highest accuracy for almost all labels. Random Forest showed strong performance overall, with its accuracy closely matching LSTM's for most labels. CNN displayed lower accuracy compared to LSTM and random forest but not too much, but its performance not too bad, suggesting it may excel in similar scenarios.

Trends Across Labels The LSTM model performed particularly well for labels reflecting its strength in sequence modeling. The Random Forest model, while slightly less effective overall, demonstrated robust performance for certain labels. The CNN also showed comparable performances.

5.2.2 General Observations

Each confusion matrix provides insight into how well the model differentiates between different psychological states. A strong diagonal presence suggests accurate classification, while significant off-diagonal misclassifications indicate areas where the model struggles. To illustrate representative trends, we present confusion matrices for a subset of labels where notable patterns emerged.

5.2.3 Case Study: "Satisfaction with Personal Relationships" (PRO3)

The confusion matrices for PRO3 (Figures 5.14, 5.15, and 5.16) reveal a distinct trend among the models. The LSTM model (Figure 5.14) shows a strong concentration of predictions around classes 5 and 6, indicating a bias towards mid-range classifications. The Random Forest model (Figure 5.15) displays a more dispersed classification pattern, suggesting it captures variations better but at the cost of increased misclassifications. In contrast, the CNN model (Figure 5.16) primarily predicts a few dominant classes, which may indicate overfitting or inadequate feature extraction.

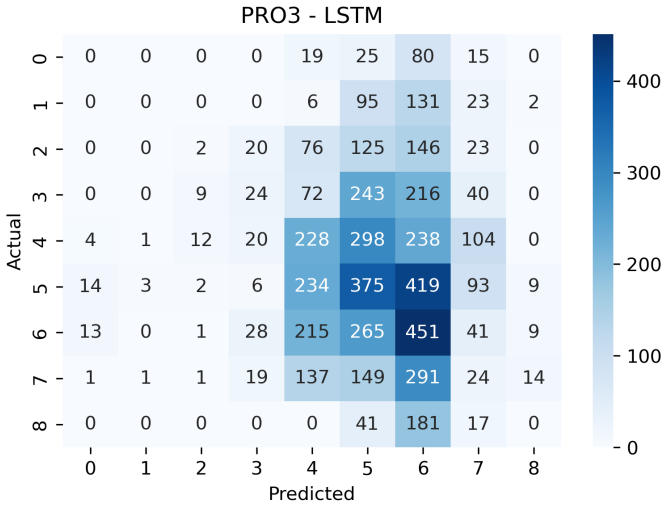

FIGURE 5.14: Confusion matrix for PRO3 using LSTM, illustrating a mid-range classification bias.

5.2.4 Case Study: "Feeling Angry Today" (N2)

A striking observation can be made from the LSTM model's predictions for anger levels (Figure 5.17). The model overwhelmingly predicts class 0, indicating a tendency to classify most responses as neutral or low anger. This suggests a difficulty in capturing variations in anger levels, which may impact its applicability in real-world monitoring of psychological states.

5.2.5 Case Study: "Feeling Positive Today" (P2)

Similarly, for predicting positive feelings (Figure 5.18), the LSTM model tends to classify responses within a narrow range, favoring mid-scale values while underestimating extreme positive or negative states. This highlights a potential challenge in capturing nuanced variations in psychological well-being.

The results indicate that different models exhibit distinct classification behaviors. The LSTM model frequently favors mid-range values, suggesting that while it captures general trends, it lacks precision in extreme cases. The Random Forest model produces more varied classifications but at the cost of increased misclassification. The CNN model, in contrast, appears to overly simplify predictions, potentially due to inadequate feature extraction. These findings suggest that further refinements in modeling strategies and feature representation may be necessary to enhance classification performance across all psychological states.

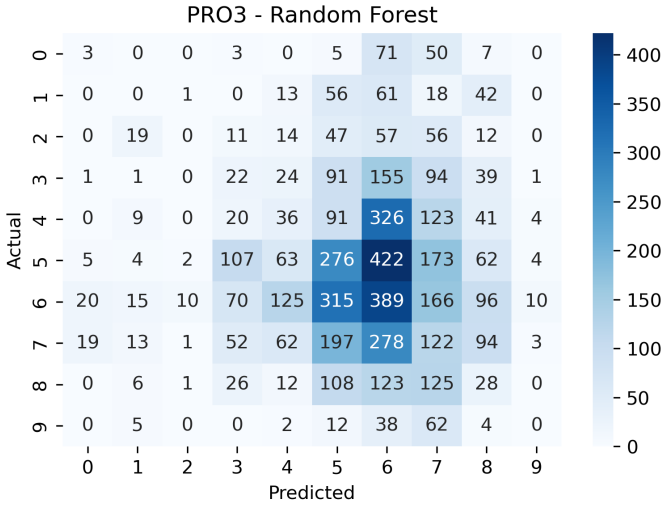

FIGURE 5.15: Confusion matrix for PRO3 using Random Forest, demonstrating greater class dispersion.

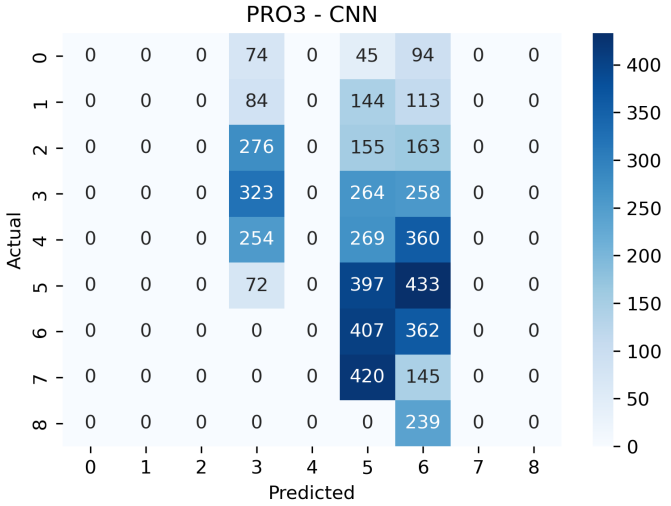

FIGURE 5.16: Confusion matrix for PRO3 using CNN, showing a concentration in a few classes.

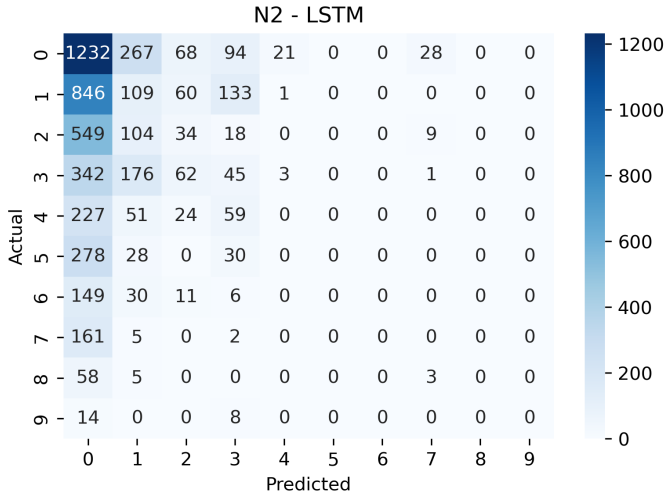

FIGURE 5.17: Confusion matrix for N2 using LSTM, indicating limited sensitivity to extreme values.

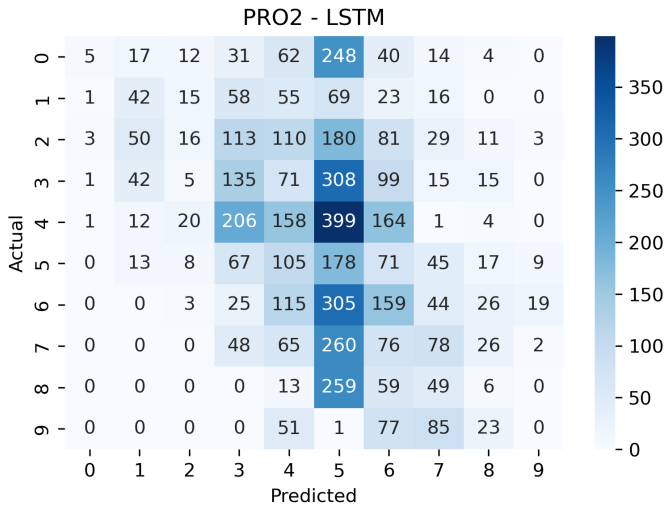

FIGURE 5.18: Confusion matrix for P2 using LSTM, showing a strong bias towards neutral predictions.

5.3 Construct-Specific Observations

Self-Esteem Constructs (SE Labels) All four SE constructs achieve the highest accuracy, suggesting that self-esteem-related emotions are physiologically distinct. Constructs like SE1 ("satisfied with self") and SE4 ("good qualities") likely align with positive emotional states, while SE2 ("failure") and SE3 ("useless") align with negative self-assessment. The ability of models to perform well across both positive and negative self-esteem constructs indicates robustness in identifying a broad spectrum of physiological signals. **MIL Constructs** MIL2 ("searching for meaning in life")'s moderate accuracy suggests that searching for meaning involves a mix of cognitive and emotional processes, potentially reflected in subtle physiological signals. This construct may also overlap with general stress markers, as searching for meaning could imply uncertainty or introspection. MIL5 ("comprehension of life's meaning") has better accuracy than other MIL constructs suggests that comprehending meaning may evoke calm, reflective physiological states that are easier to detect.

N2 (Anger) Anger tends to evoke clear physiological responses, such as increased heart rate, respiration, and skin conductance. The better performance on this label compared to other negative emotions (e.g., N1: "anxiety" or N3: "sadness") supports this.

Among them, self-Esteem constructs are most distinct and could be a indication that models excel at detecting self-esteem-related emotions, possibly because these constructs involve strong and identifiable physiological changes. Positive self-esteem constructs (e.g., SE1, SE4) likely reflect relaxed, stable states, while negative self-esteem constructs (e.g., SE2, SE3) may trigger stress-related responses. Anger is more detectable than anxiety or sadness, like N2 (anger) is better identified than other negative emotions, likely due to the pronounced physiological markers of anger compared to the subtler signals associated with anxiety or sadness. MIL constructs can represent mixed states like MIL2 (searching for meaning) and MIL5 (comprehension of meaning) are inherently ambiguous constructs that may overlap with both positive and negative emotional states, contributing to their moderate performance.

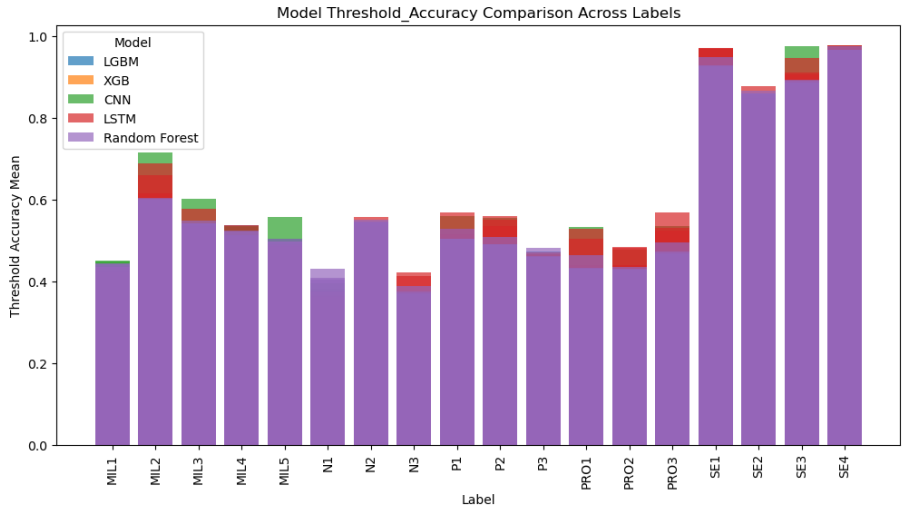

FIGURE 5.19: Threshhold Accuracy Comparison of Models in all labels

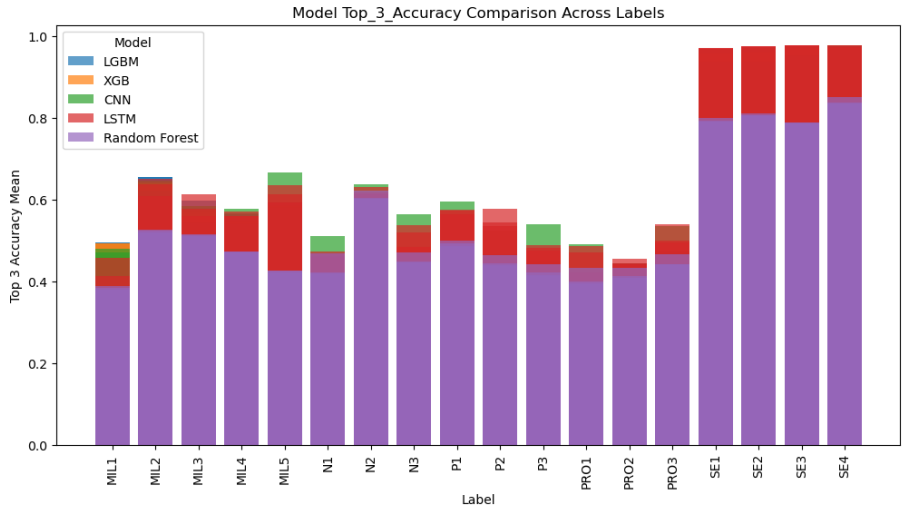

FIGURE 5.20: Top-3 Accuracy Comparison of Models in all labels

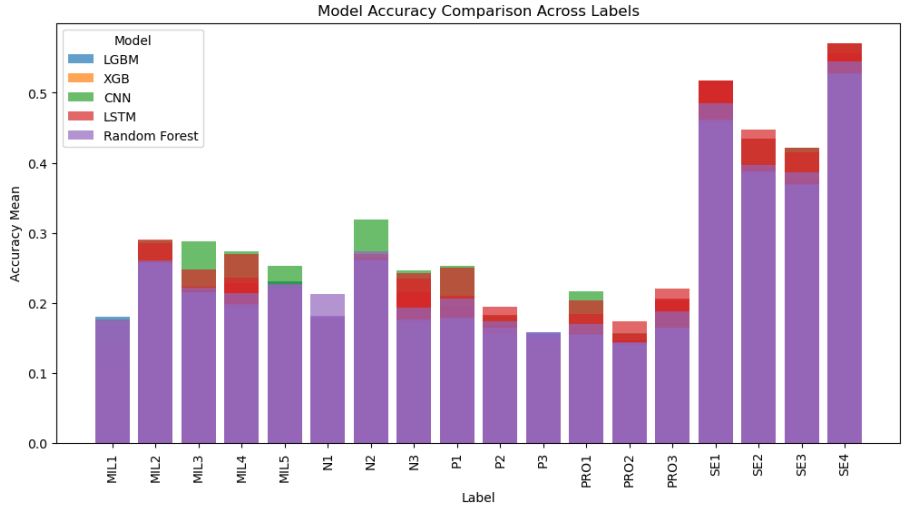

FIGURE 5.21: Accuracy Comparison of Models in all labels

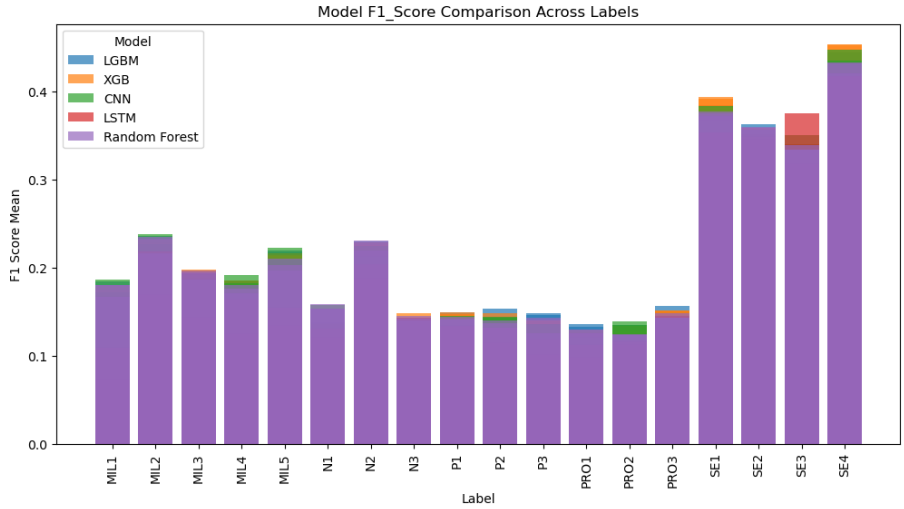

FIGURE 5.22: F1-Score Comparison of Models in all labels

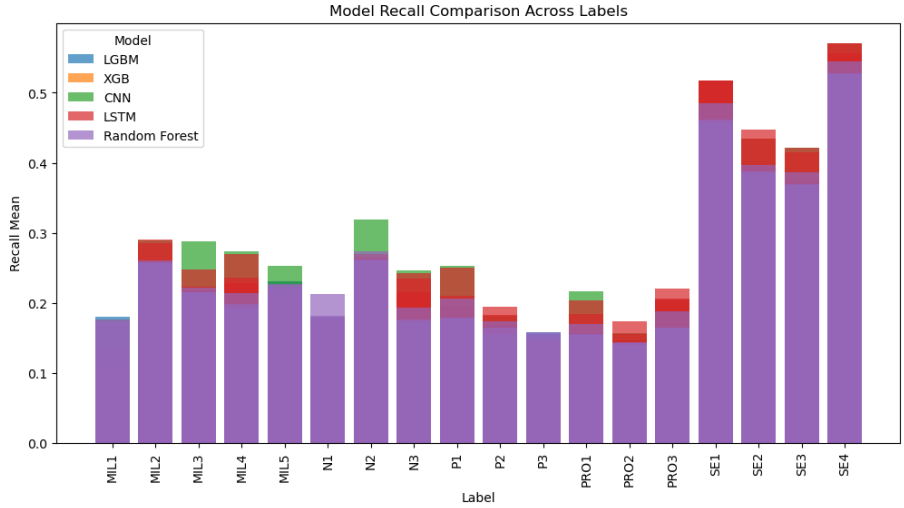

FIGURE 5.23: Recall Comparison of Models in all labels

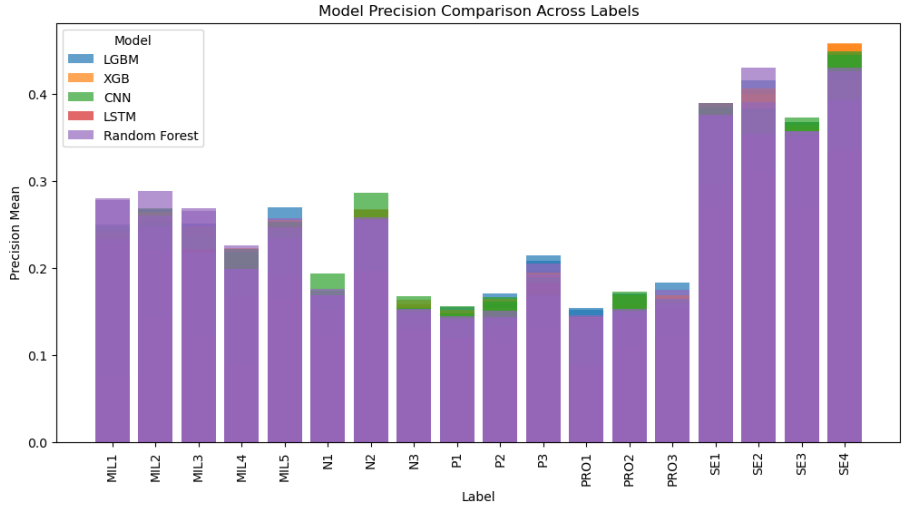

FIGURE 5.24: Precision Comparison of Models in all labels

5.4 Uncertainty Analysis

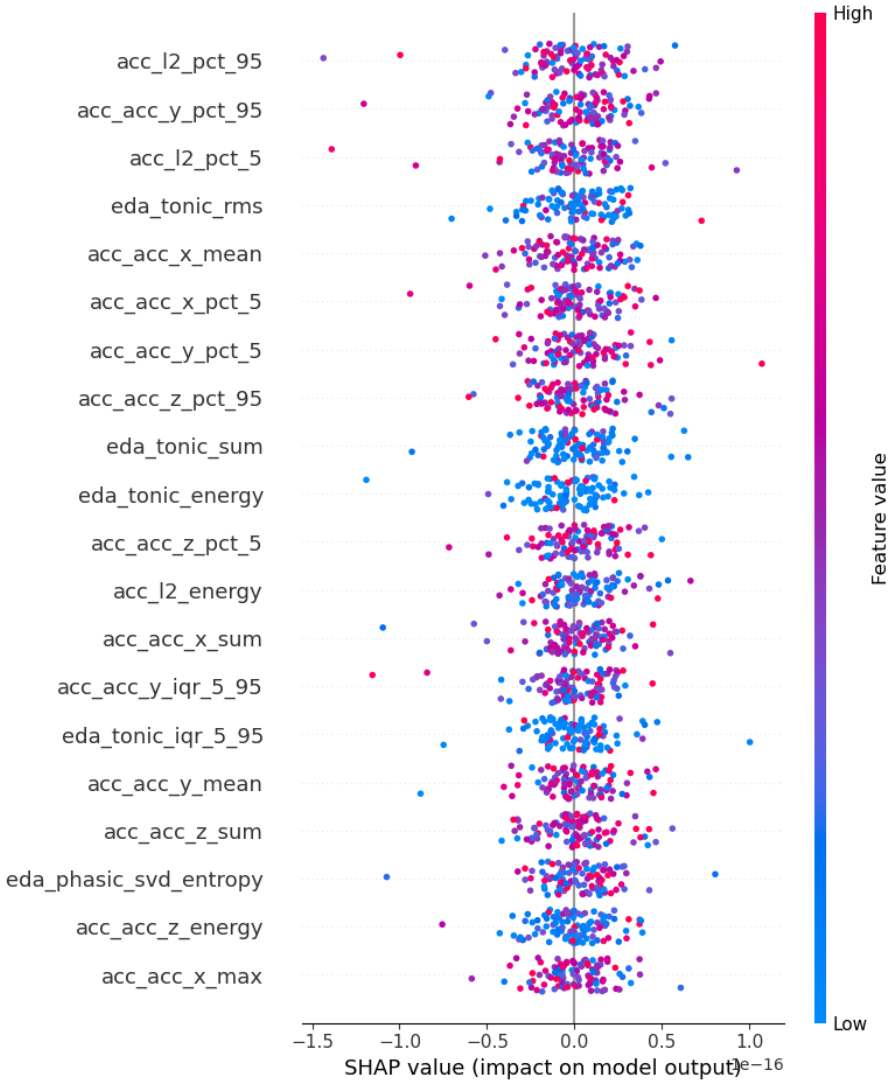

FIGURE 5.25: Shap analysis of MIL4 for random forest on scaled dataset

Figure 5.25 presents the SHAP analysis of label prediction for MIL4 in the scaled dataset. This figure illustrates the overall feature importance across all predicted classes. The classification model organizes the response variables into multiple categories, corresponding to different psychological states. In this visualization, `acc_l2_pct_95`, `eda_tonic_rms`, and `acc_acc_x_mean` emerge as the most influential features, with consistently high SHAP values.

Notably, `eda_tonic_rms` plays a significant role in higher-class predictions, while `acc_l2_pct_95` and `acc_acc_y_pct_95` exhibit strong contributions across the entire label range. Features such as `acc_acc_x_pct_5` and `eda_tonic_energy` exhibit moderate but consistent importance, suggesting their role as secondary predictors. This figure highlights the dominance of movement-related features (`acc_*`) in classification decisions, along with the notable impact of electrodermal activity features (`eda_*`). The results shown in this plot apply to the scaled dataset, with additional results for other datasets provided in the appendix.

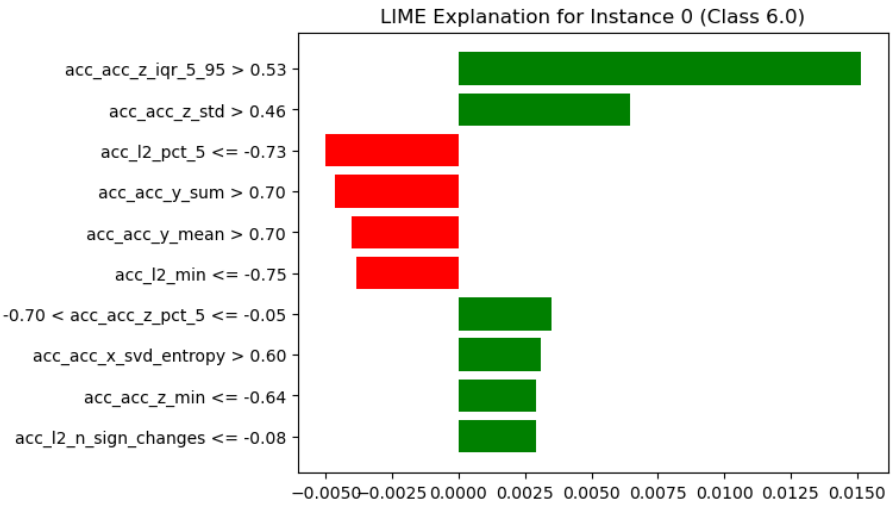

FIGURE 5.26: Lime analysis of MIL4 for random forest on Scaled dataset

Figure 5.26 shows an example of LIME analysis for MIL4 in the scaled dataset, demonstrating the local feature importance for a specific instance classified as Class 6. The full LIME results for all other classes are provided in the appendix. In this specific prediction, `acc_acc_z_iqr_5_95` and `acc_acc_z_std` are identified as the dominant contributors, positively influencing the model's decision. These features suggest that variations in acceleration, particularly along the Z-axis, are key factors in determining classification outcomes for this instance.

Conversely, `acc_l2_pct_5` and `acc_acc_y_sum` negatively impact the classification, indicating that lower values of these features decrease the likelihood of predicting Class 6. The LIME explanation highlights how specific motion patterns influence the model's decision-making for individual cases, revealing unique local variations that might not be fully captured by global SHAP analysis. This underscores the complementary role of both explainability techniques: SHAP provides a broad understanding of feature influence across all predictions, while LIME offers granular insight into individual classification outcomes. The visual breakdown in Figure 5.26 emphasizes the need to ensure balanced feature representation and interpretability in the final model.

5.5 Clustered regression

To evaluate the performance of the LSTM and Transformer models, we compare key metrics across five cross-validation folds. The results are summarized in Table 5.4.

5.5.1 Performance Analysis

The Transformer model consistently outperforms the LSTM across all key metrics:

- **Lower training and validation loss:** The Transformer achieves a significantly lower training loss (1.2714 vs. 2.6058) and validation loss (2.0798 vs. 2.2974), indicating superior optimization and generalization.
- **Improved mean absolute error (MAE):** The Transformer achieves lower MAE in both training (0.6906 vs. 1.8630) and validation (1.5026 vs. 1.5560), suggesting

TABLE 5.4: Comparison of LSTM and Transformer Models

Metric	LSTM	Transformer
Train Loss	2.61 ± 0.11	**1.27** ± 0.04
Validation Loss	2.30 ± 0.08	**2.08** ± 0.10
Train MAE	1.86 ± 0.13	**0.69** ± 0.02
Validation MAE	1.56 ± 0.09	**1.50** ± 0.08
Residual Error	0.77 ± 0.09	**-0.48** ± 0.09

better predictive accuracy.

- **Reduced residual error:** The mean residual error for the Transformer is much closer to zero (-0.4820) compared to LSTM (0.7734), indicating that Transformer predictions are more centered around the true values.

Given the limitations observed in previous models, we explore LSTM and Transformer models trained on 2D timestep data, where a regression approach is used but mapped into bins corresponding to a scale of 0 to 10. These models demonstrate improved performance compared to earlier approaches and allow for a more structured analysis of general trends due to the aggregation of labels into broader psychological aspects such as positive and negative emotions.

Figures 5.27-5.31 present confusion matrices for the LSTM model, while Figures 5.32-5.36 show results for the Transformer model. Both models exhibit strong diagonal dominance, indicating that the classification aligns well with the expected psychological states. The Transformer model, in particular, demonstrates improved differentiation between mid-range and extreme values, suggesting a more nuanced representation of psychological trends.

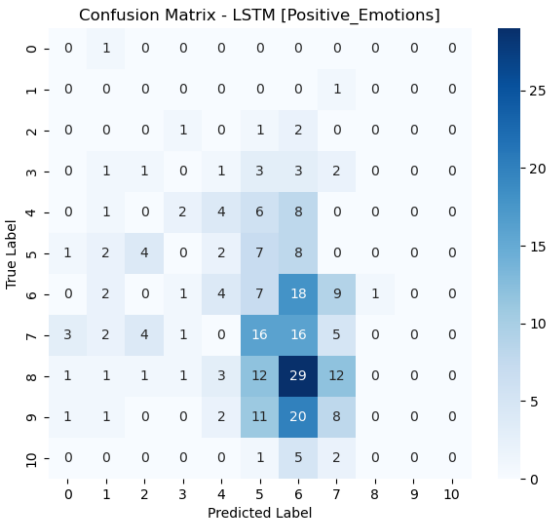

FIGURE 5.27: Confusion matrix for LSTM - Positive Emotions.

Consistency in Mid-Range Predictions: Both models show strong consistency in

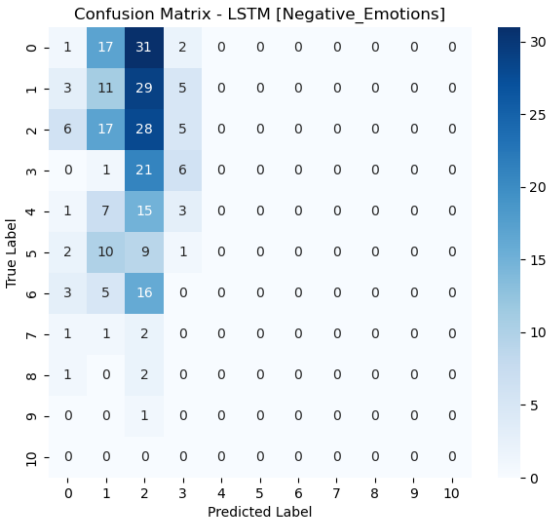

FIGURE 5.28: Confusion matrix for LSTM - Negative Emotions.

predicting mid-range values across multiple psychological states. This suggests that they have successfully learned underlying patterns in the data, particularly in distinguishing between moderate levels of emotional and psychological states. The concentration of predictions in these ranges aligns with real-world distributions, where extreme values are less frequent.

Transformer Model Captures Greater Variance: The Transformer model, in particular, exhibits a more distributed set of predictions compared to the LSTM. This indicates that it generalizes better across different classes rather than overfitting to a few dominant ones. This effect is most noticeable in *Positive Emotions* and *Negative Emotions*, where the predictions are spread across multiple labels rather than being overly concentrated in a single class.

Strong Predictive Performance in Self-Esteem: Both models demonstrate particularly high recall for *Self-Esteem*, successfully predicting the most frequent class with a high level of confidence. This suggests that the models effectively identify underlying features associated with self-esteem, capturing meaningful relationships between physiological signals and self-reported psychological states.

Reduced Ambiguity in Social Support Predictions: In the case of *Social Support*, both models show a relatively well-defined prediction pattern, with fewer extreme misclassifications. The predictions remain centered around a clearly identifiable range, suggesting that the models have successfully learned relevant physiological markers associated with social well-being.

Transformer Model Differentiates Negative Emotions More Effectively: For *Negative Emotions*, the Transformer model outperforms the LSTM in terms of prediction spread. Unlike the LSTM, which exhibits some degree of overfitting to a few key classes, the Transformer demonstrates a more balanced classification across different emotional levels. This suggests that the Transformer architecture better captures subtle variations in negative emotional states, which could be crucial for real-world applications.

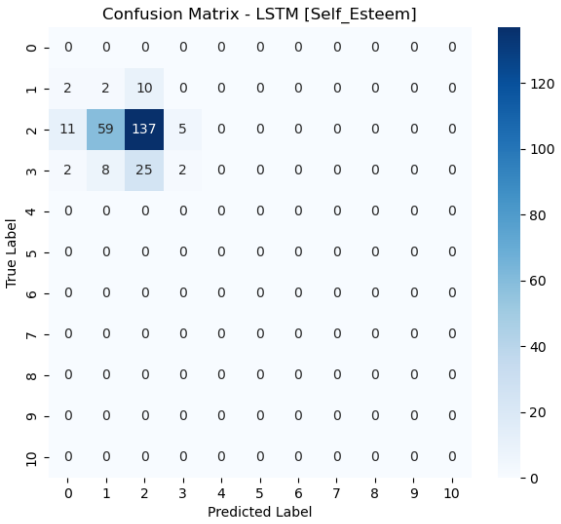

FIGURE 5.29: Confusion matrix for LSTM - Self-Esteem.

No Extreme Overconfidence in Misclassifications: One important observation is the lack of extreme misclassifications, where the models would consistently predict highly incorrect labels. Instead, errors tend to remain within adjacent or similar categories, indicating that the models maintain a reasonable level of robustness. This behavior suggests that even when misclassifications occur, the predictions are not drastically incorrect, reflecting a controlled decision-making process rather than random guessing.

Alignment with Expected Psychological Trends: Given that psychological states often follow a normal distribution—where mid-range values are more common than extremes—the observed prediction patterns align with real-world expectations. This suggests that the models are not arbitrarily assigning labels but rather capturing meaningful structure from the data.

LSTM Model Shows Strength in Specific Cases: While the Transformer model demonstrates better generalization, the LSTM model still performs well in certain cases, particularly in *Self-Esteem* and *Social Support*. The LSTM's ability to capture temporal dependencies may contribute to its consistent predictions in structured psychological constructs, where short-term fluctuations are less impactful.

Limitations and Future Considerations: While the models perform well in many aspects, some challenges remain. Both models show a tendency to overpredict mid-range values, which could suggest difficulty in differentiating between extreme psychological states. Additionally, class imbalance in certain categories (such as *Self-Esteem*) may lead to biases towards dominant labels, reducing sensitivity to less frequent classes. Future improvements could explore data balancing techniques or loss adjustments to enhance prediction performance across the full range of psychological states.

5.5.2 MAPIE on Clustered labels

To assess the reliability of predictions from the LSTM and Transformer models, we employed the MAPIE technique. The results are visualized in Figure 5.37 and 5.38, where each subplot represents uncertainty estimates for the five psychological labels: *Positive*

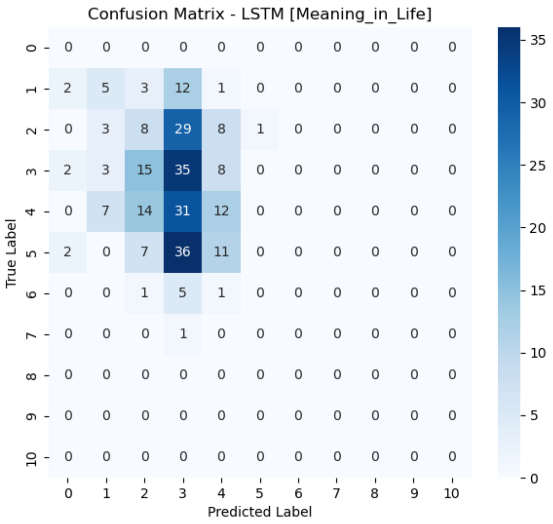

FIGURE 5.30: Confusion matrix for LSTM - Meaning in Life.

Emotions, Negative Emotions, Self-Esteem, Meaning in Life, and *Social Support.*

- **Positive Emotions & Social Support:** The Transformer model demonstrates slightly lower uncertainty compared to the LSTM, as evident from the shorter error bars. Both models, however, exhibit considerable variance in certain high-uncertainty samples.
- **Negative Emotions:** The Transformer model shows a wider spread of predictions, but its confidence intervals are more structured. The LSTM produces more uniform but larger intervals, indicating higher uncertainty.
- **Self-Esteem:** The Transformer exhibits consistently smaller error bars, suggesting a better-calibrated model with lower uncertainty. The LSTM shows larger confidence intervals, indicating reduced predictive reliability.
- **Meaning in Life:** The Transformer produces slightly narrower confidence intervals, suggesting more stable predictions. The LSTM shows greater variability with more outliers.
- **Overall Trends:** The Transformer model generally shows lower uncertainty across labels, especially in *Self-Esteem* and *Negative Emotions*. In contrast, the LSTM model struggles with high variance, suggesting it is less confident in its predictions.

The results indicate that the Transformer model is more reliable in predicting psychological states, as reflected by consistently lower uncertainty estimates. This further supports the earlier performance comparisons, where the Transformer outperformed the LSTM in mean absolute error (MAE) and residual error analysis.

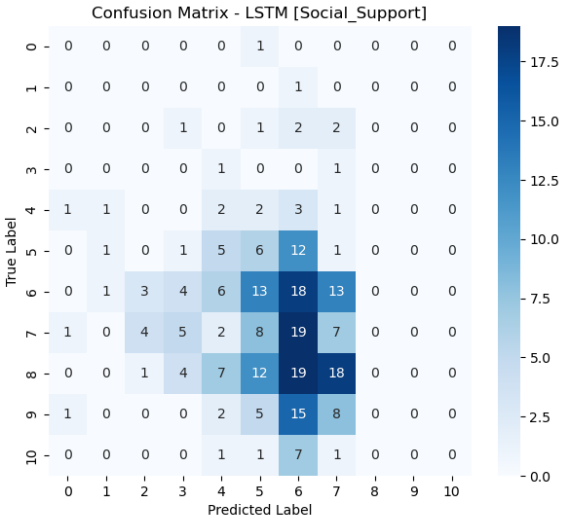

FIGURE 5.31: Confusion matrix for LSTM - Social Support.

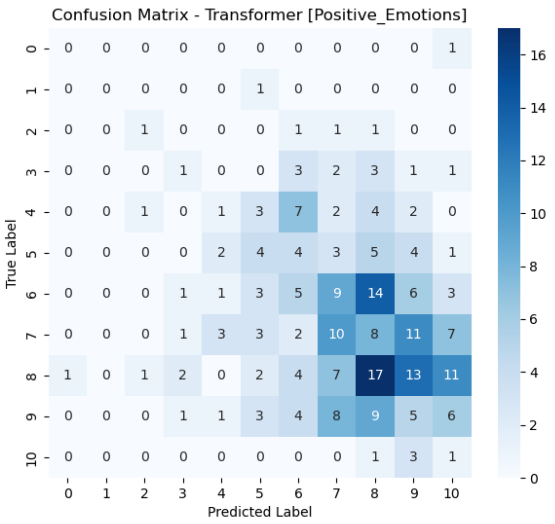

FIGURE 5.32: Confusion matrix for Transformer - Positive Emotions.

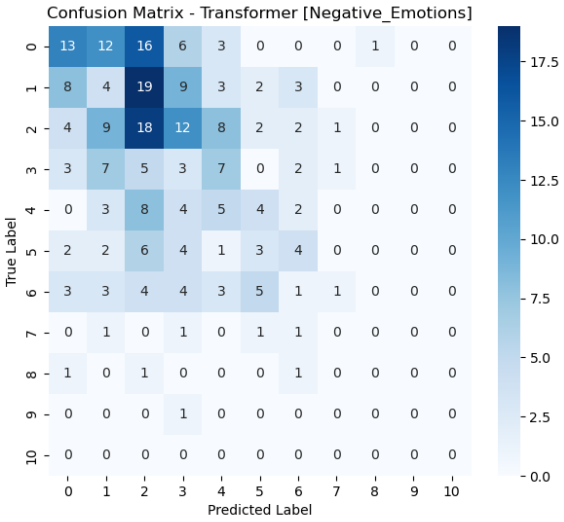

FIGURE 5.33: Confusion matrix for Transformer - Negative Emotions.

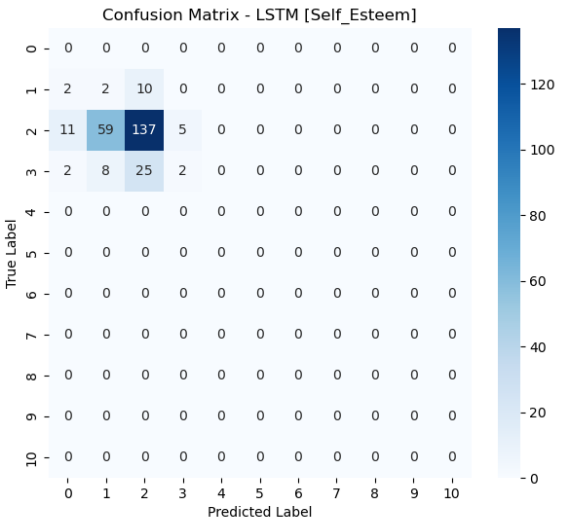

FIGURE 5.34: Confusion matrix for Transformer - Self-Esteem.

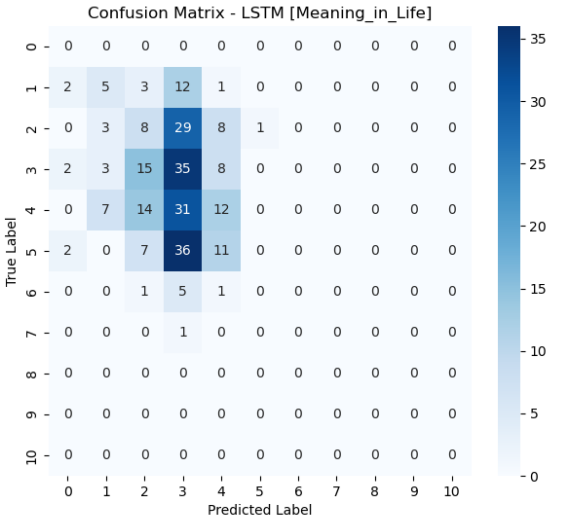

FIGURE 5.35: Confusion matrix for Transformer - Meaning in Life.

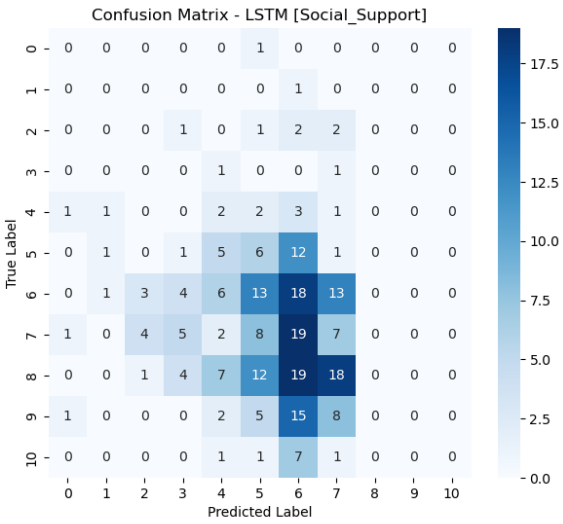

FIGURE 5.36: Confusion matrix for Transformer - Social Support.

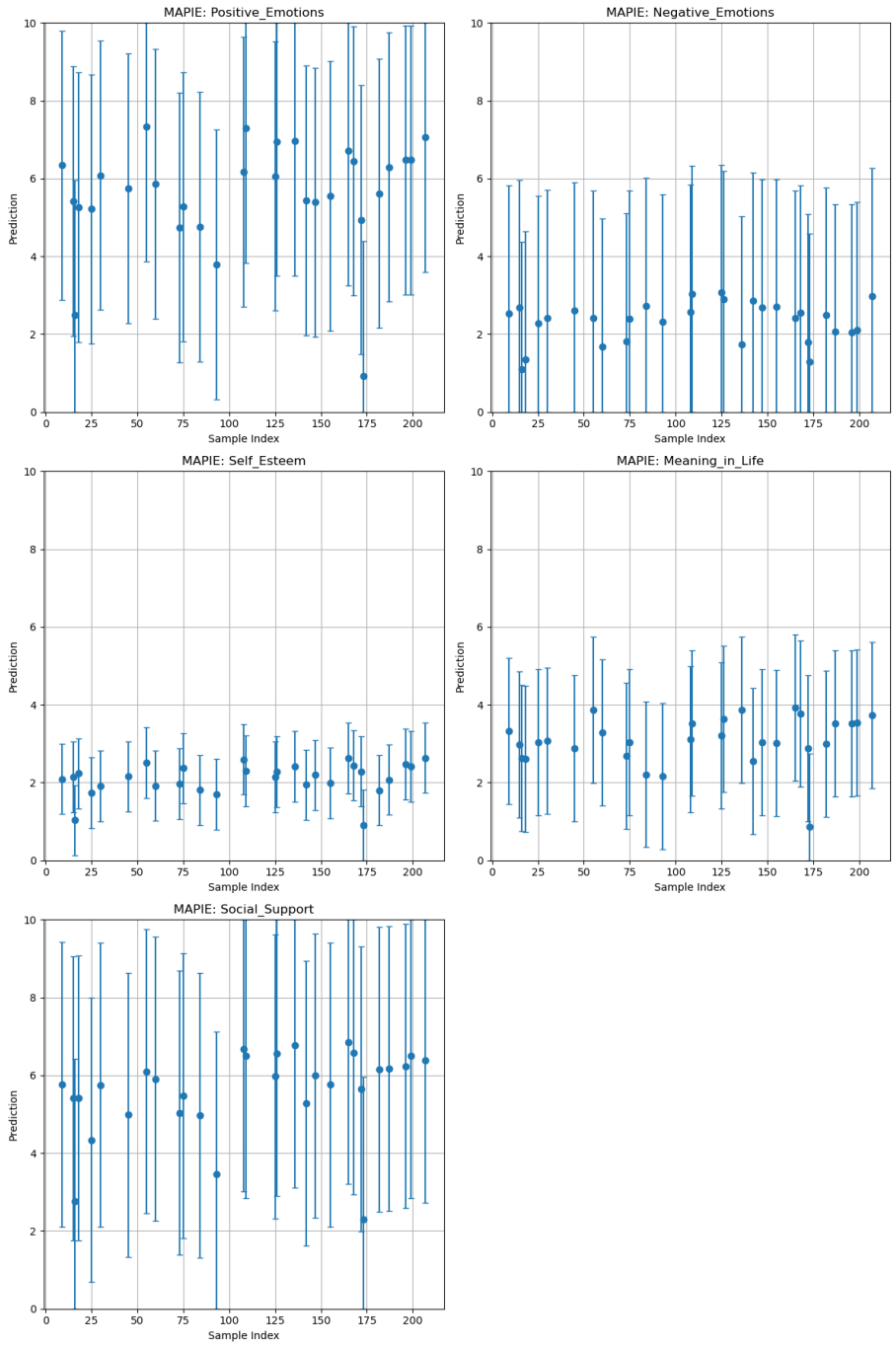

FIGURE 5.37: MAPIE analysis for all labels for LSTM model

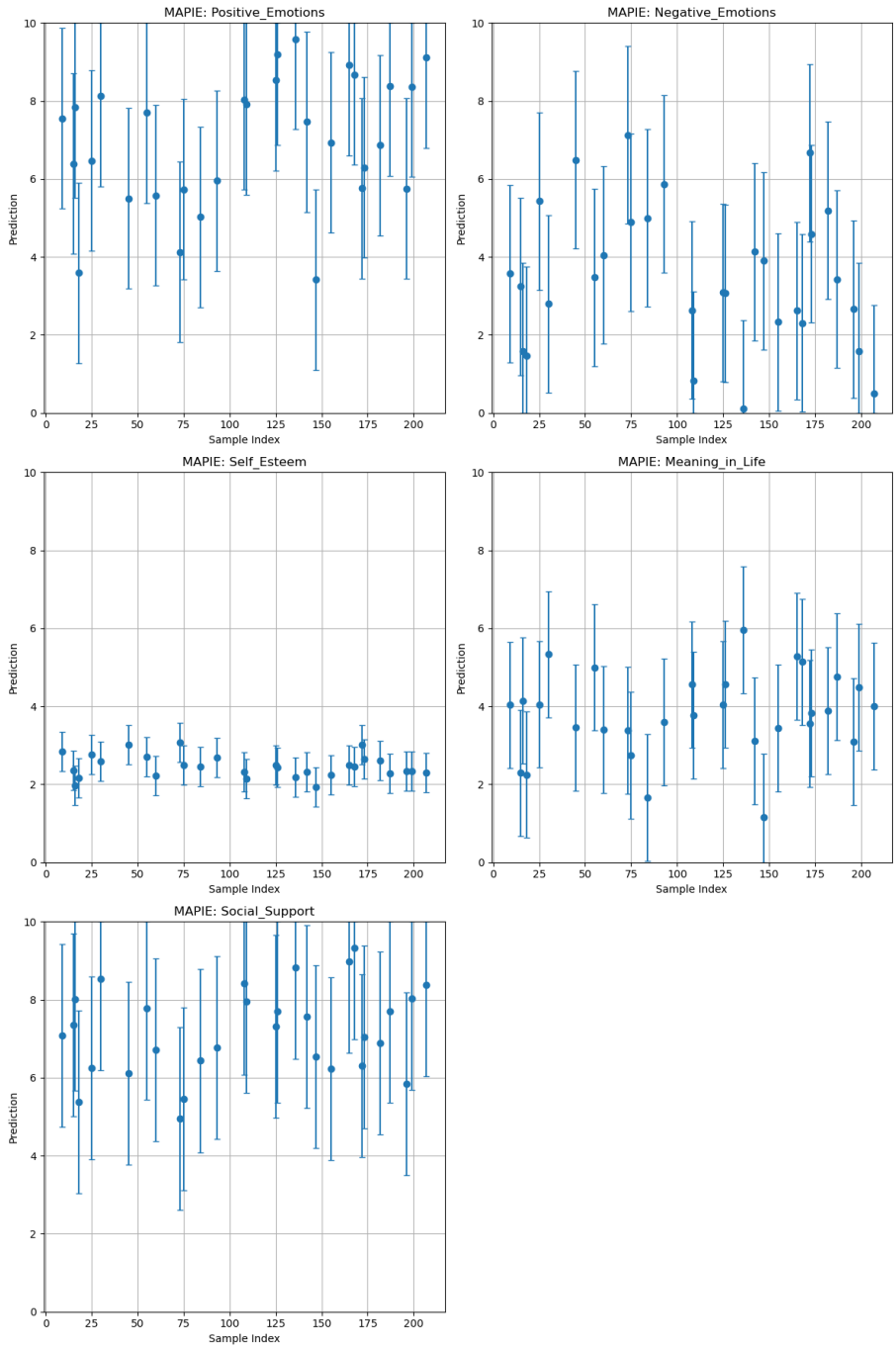

FIGURE 5.38: MAPIE analysis for all labels for Transformer model

Chapter 6

Discussion

6.1 Key Insights

Recent work has demonstrated the potential of wearables in predicting resilience, a positive psychological trait[7]. Their study applied machine learning models to wearable-derived physiological data, highlighting the feasibility of assessing positive affect through passively collected signals. Our work builds upon this by not only predicting positive states but also identifying specific physiological features—such as accelerometer and skin conductance metrics—as influential predictors, thereby enhancing interpretability.

Other studies have linked affective states to physical behavior patterns using wearable technology[55], demonstrating that activity and physiological signals are meaningful indicators of psychological states. Our study complements this work by incorporating explainable AI tools such as SHAP and LIME to uncover feature-level insights, making the predictions more transparent and interpretable for real-world applications.

Furthermore, research in personalized affect forecasting[56] has shown that multimodal deep learning models can effectively predict affective states up to a week in advance. Our work introduces new evaluation metrics, such as Top-3 accuracy and threshold accuracy, to provide a more nuanced assessment of model performance, revealing that CNN and LSTM architectures can rank predictions effectively, even when exact accuracy remains modest. In contrast to these previous works, our study makes several key contributions:

- Shifting the focus from stress/negative affect prediction to positive psychological states, contributing to a more holistic approach in wearable-based mental health monitoring.
- Leveraging deep learning models (LSTM, Transformer) with a two-dimensional temporal feature representation, improving the ability to model time-dependent physiological patterns.
- Introducing new evaluation frameworks (Top-3 accuracy, threshold accuracy) to provide a richer assessment of predictive performance beyond conventional accuracy metrics.
- Enhancing interpretability through explainable AI methods (SHAP, LIME), making wearable-based mental health applications more transparent and user-friendly.

6.2 Limitations

While the findings of this study are promising, several limitations should be noted. First, the dataset, though rich in sensor modalities, was limited in size and diversity, as it was collected from a small, homogenous sample (primarily university students from a WEIRD—Western, Educated, Industrialized, Rich, Democratic—population). This restricts the generalizability of the results to broader populations.

Second, data quality issues such as missing values or sensor noise presented challenges during preprocessing, which may have affected model performance. Imputation methods were applied to reduce the impact, but advanced methods using deep learning for imputation was not applied, like MICE[57]. Additionally, the modest predictive accuracy of the models underscores the complexity of psychological state prediction, suggesting that current feature engineering and model selection methods may not fully capture the underlying physiological signals.

Another limitation of this study lies in its current inability to translate prediction intervals into actionable insights for end-users. While MAPIE provided robust uncertainty quantification, its application to guide personalized interventions has not yet been explored. For example, predictions with wide intervals could indicate periods of physiological or behavioral instability, potentially warranting targeted interventions, such as mindfulness exercises or reminders to rest. Incorporating such insights into the modeling process or real-time applications could enhance the utility of these predictions.

This study primarily aimed to explore the contribution of different physiological signals in predicting psychological states rather than optimizing the selection or combination of multimodal features. While feature importance was analyzed to gain insights into relevant physiological markers, a more systematic investigation of feature selection strategies could further refine predictive performance.

Another point is accelerometer features used in this study, including features in the Y and Z directions, are sensitive to sensor orientation and placement. Variations in how the sensor was worn (e.g.different angles) may have influenced the extracted features and model predictions. While alternative approaches, such as using the norm of the accelerometer data, could mitigate this issue by making features orientation-independent, this study prioritized other movement-related features. Future work could explore a combination of directional and orientation-invariant features to improve robustness.

6.3 Future Works

To enhance the generalizability of findings, future studies should consider utilizing larger and more diverse datasets. This approach has been shown to improve model robustness and applicability across varied populations[58]

Implementing real-time data processing and prediction pipelines is crucial for deploying these models in practical settings, thereby increasing their utility for end-users. The development of such systems can facilitate timely interventions and support.

Exploring advanced feature selection techniques may better capture synergies between different data modalities, potentially enhancing predictive performance. Methods such as hybrid fusion, which integrate structured and unstructured data, have demonstrated improved accuracy in clinical prediction models.[59]

Expanding research to include multimodal data fusion—for instance, combining wearable data with contextual factors like location or activity—could provide deeper insights into the dynamics of positive psychological states. Multimodal fusion approaches have been effective in uncovering complex relationships in mental health studies.[60]

Integrating uncertainty quantification methods, such as MAPIE, into real-time applications can enhance user trust by conveying confidence levels in predictions. MAPIE offers a model-agnostic approach to estimating prediction intervals, which is valuable for assessing the reliability of machine learning models.

Further refinement of explainable AI techniques is also recommended to improve interpretability and user understanding. Leveraging explainable AI in conjunction with multimodal data has been shown to provide transparent and scalable approaches to mental health diagnostics.[61]

Additionally, future research should investigate whether data from specific periods of the day suffice for accurate predictions. Focusing on particular time windows, such as the first six hours after waking, could enable proactive interventions earlier in the day, thereby optimizing user experience and mental health outcomes.

Chapter 7

Conclusions

This study demonstrates that wearable device data can be used to predict positive psychological states, contributing to the broader goal of leveraging digital biomarkers for mental health monitoring. Through the integration of physiological signals, machine learning models, and explainable AI techniques, this work provides a structured approach to understanding the relationship between physical indicators and psychological well-being.

Our results indicate that while absolute predictive accuracy remains modest, models such as LSTMs, Random Forest and Transformers show improved performance using threshold-based accuracy metrics, achieving up to 62% accuracy in specific cases. The analysis revealed that accelerometer-based features contributed most significantly to prediction performance, while heart rate variability (HRV) and electrodermal activity (EDA) played supporting roles in specific psychological constructs, such as self-esteem and meaning in life. The use of SHAP and LIME analysis further highlighted the interpretability of the models, identifying key physiological features that influence predictions.

Despite these advancements, challenges remain. The variability in data quality, potential sensor noise, and individual differences in physiological responses contribute to limitations in model generalizability. Furthermore, while uncertainty quantification methods such as MAPIE helped assess prediction confidence, further research is needed to translate these insights into practical, real-world applications, such as real-time interventions.

Future work should focus on expanding the dataset to improve generalizability, optimizing feature selection strategies to enhance predictive power, and integrating real-time processing capabilities for deployment on wearable devices. Additionally, multimodal data fusion—including contextual factors such as sleep patterns, physical activity levels, and environmental conditions—could provide a more comprehensive understanding of psychological states.

Overall, this study lays the groundwork for using wearable technology in mental health applications, particularly in understanding and predicting positive psychological states. By refining machine learning approaches and enhancing interpretability, future research can bridge the gap between theoretical findings and practical, user-centered applications.

Bibliography

[1] Keith M. Diaz et al. "Fitbit®: An Accurate and Reliable Device for Wireless Physical Activity Tracking". In: *International Journal of Cardiology* 185 (Mar. 15, 2015), pp. 138–140. ISSN: 0167-5273. DOI: 10.1016/j.ijcard.2015.03.038.

[2] Kelly R. Evenson, Miya M. Goto, and Robert D. Furberg. "Review of Validity and Reliability of Garmin Activity Trackers". In: *Journal of Physical Activity and Health* 12 (June 1, 2015), S157–S168. ISSN: 1543-3080. DOI: 10.1123/jpah.2014-0264.

[3] Blake Anthony Hickey et al. "Smart Devices and Wearable Technologies to Detect and Monitor Mental Health Conditions and Stress: A Systematic Review". In: *Sensors* 21 (Jan. 1, 2021), pp. 1–29. ISSN: 1424-8220. DOI: 10.3390/s21010025.

[4] Nuno Gomes et al. "A Survey on Wearable Sensors for Mental Health Monitoring". In: *Sensors (Basel, Switzerland)* 23.3 (Jan. 25, 2023), p. 1330. ISSN: 1424-8220. DOI: 10.3390/s23031330. pmid: 36772370.

[5] Berrenur Saylam and Özlem Durmaz İncel. "Quantifying Digital Biomarkers for Well-Being: Stress, Anxiety, Positive and Negative Affect via Wearable Devices and Their Time-Based Predictions". In: *Sensors* 23.21 (21 Jan. 2023), p. 8987. ISSN: 1424-8220. DOI: 10.3390/s23218987.

[6] Pramod T Borghare, Disha A Methwani, and Aniket G Pathade. "A Comprehensive Review on Harnessing Wearable Technology for Enhanced Depression Treatment". In: *Cureus* 16.8 (), e66173. ISSN: 2168-8184. DOI: 10.7759/cureus.66173. pmid: 39233951.

[7] Robert P Hirten et al. "A Machine Learning Approach to Determine Resilience Utilizing Wearable Device Data: Analysis of an Observational Cohort". In: *JAMIA Open* 6.2 (July 1, 2023), ooad029. ISSN: 2574-2531. DOI: 10.1093/jamiaopen/ooad029.

[8] Simon Föll et al. "FLIRT: A Feature Generation Toolkit for Wearable Data". In: *Computer Methods and Programs in Biomedicine* 212 (Nov. 2021), p. 106461. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2021.106461.

[9] Jennifer J. Newson, Daniel Hunter, and Tara C. Thiagarajan. "The Heterogeneity of Mental Health Assessment". In: *Frontiers in Psychiatry* 11 (Feb. 27, 2020), p. 76. ISSN: 1664-0640. DOI: 10.3389/fpsyt.2020.00076. pmid: 32174852.

[10] Elena Smets et al. "Large-Scale Wearable Data Reveal Digital Phenotypes for Daily-Life Stress Detection". In: *npj Digital Medicine* 1.1 (Dec. 2018), pp. 1–10. ISSN: 2398-6352. DOI: 10.1038/s41746-018-0074-9.

[11] Paola Pedrelli et al. "Monitoring Changes in Depression Severity Using Wearable and Mobile Sensors". In: *Frontiers in Psychiatry* 11 (2020), p. 584711. ISSN: 1664-0640. DOI: 10.3389/fpsyt.2020.584711.

[12] Daniel A. Adler et al. "Beyond Detection: Towards Actionable Sensing Research in Clinical Mental Healthcare". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8.4 (Nov. 2024), 160:1–160:33. DOI: 10.1145/3699755.

[13] Gideon Vos et al. "Ensemble Machine Learning Model Trained on a New Synthesized Dataset Generalizes Well for Stress Prediction Using Wearable Devices". In: *Journal of Biomedical Informatics* 148 (Dec. 2023), p. 104556. ISSN: 1532-0464. DOI: 10.1016/j.jbi.2023.104556.

[14] C. Jyotsna et al. "PredictEYE: Personalized Time Series Model for Mental State Prediction Using Eye Tracking". In: *IEEE Access* 11 (2023), pp. 128383–128409. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2023.3332762.

[15] Laurent Servais et al. "First Regulatory Qualification of a Digital Primary Endpoint to Measure Treatment Efficacy in DMD". In: *Nature Medicine* 29.10 (Oct. 2023), pp. 2391–2392. ISSN: 1546-170X. DOI: 10.1038/s41591-023-02459-5.

[16] Ujunwa Madububambachu, Augustine Ukpebor, and Urenna Ihezue. "Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review". In: (). DOI: 10.2174/0117450179315688240607052117.

[17] Suppawong Tuarob et al. "How Are You Feeling?: A Personalized Methodology for Predicting Mental States from Temporally Observable Physical and Behavioral Information". In: *Journal of Biomedical Informatics* 68 (Apr. 2017), pp. 1–19. ISSN: 1532-0464. DOI: 10.1016/j.jbi.2017.02.010.

[18] Sirat Samyoun et al. "M3Sense: Affect-Agnostic Multitask Representation Learning Using Multimodal Wearable Sensors". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6.2 (July 2022), 73:1–73:32. DOI: 10.1145/3534600.

[19] Isao Kurebayashi et al. "Mental-State Estimation Model with Time-Series Environmental Data Regarding Cognitive Function". In: *Internet of Things* 22 (July 2023), p. 100730. ISSN: 2542-6605. DOI: 10.1016/j.iot.2023.100730.

[20] Clemens Stachl et al. "Predicting Personality from Patterns of Behavior Collected with Smartphones". In: *Proceedings of the National Academy of Sciences* 117.30 (July 2020), pp. 17680–17687. DOI: 10.1073/pnas.1920484117.

[21] Ruiqi Wang et al. *Husformer: A Multi-Modal Transformer for Multi-Modal Human State Recognition*. Apr. 2023. DOI: 10.48550/arXiv.2209.15182. arXiv: 2209.15182 [cs].

[22] Gerard Anmella et al. "Exploring Digital Biomarkers of Illness Activity in Mood Episodes: Hypotheses Generating and Model Development Study". In: *JMIR mHealth and uHealth* 11.1 (May 2023), e45405. DOI: 10.2196/45405.

[23] Filippo Corponi et al. "Automated Mood Disorder Symptoms Monitoring from Multivariate Time-Series Sensory Data: Getting the Full Picture beyond a Single Number". In: *Translational Psychiatry* 14.1 (Mar. 2024), pp. 1–9. ISSN: 2158-3188. DOI: 10.1038/s41398-024-02876-1.

[24] Emese Sükei et al. "Predicting Emotional States Using Behavioral Markers Derived From Passively Sensed Data: Data-Driven Machine Learning Approach". In: *JMIR mHealth and uHealth* 9.3 (Mar. 2021), e24465. ISSN: 2291-5222. DOI: 10.2196/24465.

[25] Ali Tazarv et al. "Personalized Stress Monitoring Using Wearable Sensors in Everyday Settings". In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Nov. 2021, pp. 7332–7335. DOI: 10.1109/EMBC46164.2021.9630224.

[26] Hossein Hamidi Shishavan et al. "Unsupervised Bayesian Change Point Detection Model to Track Acute Stress Responses". In: *Biomedical Signal Processing and Control* 95 (Sept. 2024), p. 106415. ISSN: 1746-8094. DOI: 10.1016/j.bspc.2024.106415.

[27] Aoyu Li et al. "Synergy through Integration of Digital Cognitive Tests and Wearable Devices for Mild Cognitive Impairment Screening". In: *Frontiers in Human Neuroscience* 17 (Apr. 2023). ISSN: 1662-5161. DOI: 10.3389/fnhum.2023.1183457.

[28] Aoyu Li et al. "Unearthing Subtle Cognitive Variations: A Digital Screening Tool for Detecting and Monitoring Mild Cognitive Impairment". In: *International Journal of Human–Computer Interaction* 0.0 (Apr. 2024), pp. 1–21. ISSN: 1044-7318. DOI: 10.1080/10447318.2024.2327179.

[29] Rui Wu et al. "Data Imputation for Multivariate Time Series Sensor Data With Large Gaps of Missing Data". In: *IEEE Sensors Journal* 22.11 (June 2022), pp. 10671–10683. ISSN: 1558-1748. DOI: 10.1109/JSEN.2022.3166643.

[30] Shikha, Dr. Divyashikha Sethia, and S. Indu. "Optimization of Wearable Biosensor Data for Stress Classification Using Machine Learning and Explainable AI". In: *IEEE Access* (2024), pp. 1–1. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3463742.

[31] Žiga Stržinar et al. "Stress Detection Using Frequency Spectrum Analysis of Wrist-Measured Electrodermal Activity". In: *Sensors* 23.2 (Jan. 2023), p. 963. ISSN: 1424-8220. DOI: 10.3390/s23020963.

[32] Martin Maritsch et al. "Improving Heart Rate Variability Measurements from Consumer Smartwatches with Machine Learning". In: *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*. Sept. 9, 2019, pp. 934–938. DOI: 10.1145/3341162.3346276. arXiv: 1907.07496 [cs].

[33] Mara Naegelin et al. "An Interpretable Machine Learning Approach to Multimodal Stress Detection in a Simulated Office Environment". In: *Journal of Biomedical Informatics* 139 (Mar. 2023), p. 104299. ISSN: 1532-0464. DOI: 10.1016/j.jbi.2023.104299.

[34] Martin Karl Moser, Maximilian Ehrhart, and Bernd Resch. "An Explainable Deep Learning Approach for Stress Detection in Wearable Sensor Measurements". In: *Sensors* 24.16 (Jan. 2024), p. 5085. ISSN: 1424-8220. DOI: 10.3390/s24165085.

[35] Anastasios N. Angelopoulos and Stephen Bates. *A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification*. Dec. 2022. DOI: 10.48550/arXiv.2107.07511. arXiv: 2107.07511 [cs, math, stat].

[36] Martin Seligman. "PERMA and the Building Blocks of Well-Being". In: *The Journal of Positive Psychology* 13.4 (July 2018), pp. 333–335. ISSN: 1743-9760, 1743-9779. DOI: 10.1080/17439760.2018.1437466.

[37] Enipher Chisale and Fletcher Phiri. "PERMA Model and Mental Health Practice". In: *Asian Journal of Pharmacy, Nursing and Medical Sciences* 10 (Sept. 2022). DOI: 10.24203/ajpnms.v10i2.7015.

[38] Margaret L. Kern et al. "A Multidimensional Approach to Measuring Well-Being in Students: Application of the PERMA Framework". In: *The Journal of Positive Psychology* 10.3 (May 2015), pp. 262–271. ISSN: 1743-9760. DOI: 10.1080/17439760.2014.936962.

[39] Melissa K. Kovich et al. "Application of the PERMA Model of Well-being in Undergraduate Students". In: *International Journal of Community Well-Being* 6.1 (2023), pp. 1–20. ISSN: 2524-5295. DOI: 10.1007/s42413-022-00184-4.

[40] Debbie S. Moskowitz and Simon N. Young. "Ecological Momentary Assessment: What It Is and Why It Is a Method of the Future in Clinical Psychopharmacology". In: *Journal of Psychiatry and Neuroscience* 31.1 (Jan. 2006), pp. 13–20. ISSN: 1180-4882. pmid: 16496031. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1325062/.

[41] Jinhyuk Kim et al. "Potential Benefits of Integrating Ecological Momentary Assessment Data into mHealth Care Systems". In: *BioPsychoSocial Medicine* 13.1 (Aug. 9, 2019), p. 19. ISSN: 1751-0759. DOI: 10.1186/s13030-019-0160-5.

[42] *EmbracePlus | The World's Most Advanced Smartwatch for Continuous Health Monitoring*. Empatica. URL: https://www.empatica.com/embraceplus.

[43] Merijn Mestdagh et al. "M-Path: An Easy-to-Use and Highly Tailorable Platform for Ecological Momentary Assessment and Intervention in Behavioral Research and Clinical Practice". In: *Frontiers in Digital Health* 5 (Oct. 18, 2023). ISSN: 2673-253X. DOI: 10.3389/fdgth.2023.1182175.

[44] *Raw Data in the Empatica Cloud*. Empatica Support. URL: https://support.empatica.com/hc/en-us/articles/17727336158365-Raw-data-in-the-Empatica-Cloud.

[45] *How to Access Avro Files with Python*. Empatica Support. URL: https://support.empatica.com/hc/en-us/articles/17405877853981-How-to-access-Avro-files-with-Python.

[46] Leo Breiman. "Random Forests". In: *Machine Learning* 45.1 (Oct. 1, 2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324.

[47] Keiron O'Shea and Ryan Nash. *An Introduction to Convolutional Neural Networks*. Dec. 2, 2015. DOI: 10.48550/arXiv.1511.08458. arXiv: 1511.08458 [cs]. Prepublished.

[48] *Long Short-Term Memory | Neural Computation*. URL: https://dl.acm.org/doi/10.1162/neco.1997.9.8.1735.

[49] Ashish Vaswani et al. "Attention Is All You Need". In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017, pp. 5998–6008. URL: https://papers.nips.cc/paper/7181-attention-is-all-you-need.

[50] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, pp. 785–794. URL: https://dl.acm.org/doi/10.1145/2939672.2939785.

[51] Guolin Ke et al. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017, pp. 3146–3154. URL: https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.

[52] Vianney Taquet et al. *MAPIE: an open-source library for distribution-free uncertainty quantification*. 2022. arXiv: 2207.12274 [stat.ML]. URL: https://arxiv.org/abs/2207.12274.

[53] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. 2016. arXiv: 1602.04938 [cs.LG]. URL: https://arxiv.org/abs/1602.04938.

[54] Scott Lundberg and Su-In Lee. *A Unified Approach to Interpreting Model Predictions*. 2017. arXiv: 1705.07874 [cs.AI]. URL: https://arxiv.org/abs/1705.07874.

[55] Gergely Biri et al. "Predicting Affective States Using Wearable Technology: A Machine Learning Approach". In: *2024 IEEE International Workshop on Sport, Technology and Research (STAR)*. 2024 IEEE International Workshop on Sport, Technology and Research (STAR). July 2024, pp. 199–204. DOI: 10.1109/STAR62027.2024.10635909.

[56] Zhongqi Yang et al. "Integrating Wearable Sensor Data and Self-reported Diaries for Personalized Affect Forecasting". In: *Smart Health* 32 (June 2024), p. 100464. ISSN: 23526483. DOI: 10.1016/j.smhl.2024.100464. arXiv: 2403.13841 [cs].

[57] Stef van Buuren and Karin Groothuis-Oudshoorn. "Mice: Multivariate Imputation by Chained Equations in R". In: *Journal of Statistical Software* 45 (Dec. 12, 2011), pp. 1–67. ISSN: 1548-7660. DOI: 10.18637/jss.v045.i03.

[58] Adrienne Kline et al. "Multimodal Machine Learning in Precision Health: A Scoping Review". In: *npj Digital Medicine* 5.1 (Nov. 7, 2022), pp. 1–14. ISSN: 2398-6352. DOI: 10.1038/s41746-022-00712-8.

[59] *(PDF) Multimodal Data Hybrid Fusion and Natural Language Processing for Clinical Prediction Models*. ResearchGate. Dec. 9, 2024. DOI: 10.1101/2023.08.24.23294597.

[60] Jing Sui, Dongmei Zhi, and Vince D Calhoun. "Data-Driven Multimodal Fusion: Approaches and Applications in Psychiatric Research". In: *Psychoradiology* 3 (Nov. 22, 2023), kkad026. ISSN: 2634-4416. DOI: 10.1093/psyrad/kkad026. pmid: 38143530.

[61] Agboro Destiny. "Leveraging Explainable AI and Multimodal Data for Stress Level Prediction in Mental Health Diagnostics". In: *International Journal of Research and Innovation in Applied Science* IX.XII (2025), pp. 416–425. ISSN: 24546194. DOI: 10.51584/IJRIAS.2024.912037.

.1 Tables

TABLE 2: List of Features

Feature Name
num_ibis
hrv_mean_nni
hrv_median_nni
hrv_range_nni
hrv_sdsd
hrv_rmssd
hrv_nni_50
hrv_pnni_50
hrv_nni_20
hrv_pnni_20
hrv_cvsd
hrv_sdnn
hrv_cvnni
hrv_mean_hr
hrv_min_hr
hrv_max_hr
hrv_std_hr
hrv_total_power
hrv_vlf
hrv_lf
hrv_hf
hrv_lf_hf_ratio
hrv_lfnu
hrv_hfnu
hrv_mean
hrv_std
hrv_min
hrv_max
hrv_ptp
hrv_sum
hrv_energy
hrv_skewness
hrv_kurtosis
hrv_peaks
hrv_rms
hrv_lineintegral
hrv_n_above_mean
hrv_n_below_mean
hrv_n_sign_changes
hrv_iqr
hrv_iqr_5_95
hrv_pct_5
hrv_pct_95
hrv_entropy
hrv_perm_entropy

hrv_svd_entropy
eda_tonic_mean
eda_tonic_std
eda_tonic_min
eda_tonic_max
eda_tonic_ptp
eda_tonic_sum
eda_tonic_energy
eda_tonic_skewness
eda_tonic_kurtosis
eda_tonic_peaks
eda_tonic_rms
eda_tonic_lineintegral
eda_tonic_n_above_mean
eda_tonic_n_below_mean
eda_tonic_n_sign_changes
eda_tonic_iqr
eda_tonic_iqr_5_95
eda_tonic_pct_5
eda_tonic_pct_95
eda_tonic_entropy
eda_tonic_perm_entropy
eda_tonic_svd_entropy
eda_phasic_mean
eda_phasic_std
eda_phasic_min
eda_phasic_max
eda_phasic_ptp
eda_phasic_sum
eda_phasic_energy
eda_phasic_skewness
eda_phasic_kurtosis
eda_phasic_peaks
eda_phasic_rms
eda_phasic_lineintegral
eda_phasic_n_above_mean
eda_phasic_n_below_mean
eda_phasic_n_sign_changes
eda_phasic_iqr
eda_phasic_iqr_5_95
eda_phasic_pct_5
eda_phasic_pct_95
eda_phasic_entropy
eda_phasic_perm_entropy
eda_phasic_svd_entropy
acc_acc_x_mean
acc_acc_x_std
acc_acc_x_min
acc_acc_x_max

acc_acc_x_ptp
acc_acc_x_sum
acc_acc_x_energy
acc_acc_x_skewness
acc_acc_x_kurtosis
acc_acc_x_peaks
acc_acc_x_rms
acc_acc_x_lineintegral
acc_acc_x_n_above_mean
acc_acc_x_n_below_mean
acc_acc_x_n_sign_changes
acc_acc_x_iqr
acc_acc_x_iqr_5_95
acc_acc_x_pct_5
acc_acc_x_pct_95
acc_acc_x_entropy
acc_acc_x_perm_entropy
acc_acc_x_svd_entropy
acc_acc_y_mean
acc_acc_y_std
acc_acc_y_min
acc_acc_y_max
acc_acc_y_ptp
acc_acc_y_sum
acc_acc_y_energy
acc_acc_y_skewness
acc_acc_y_kurtosis
acc_acc_y_peaks
acc_acc_y_rms
acc_acc_y_lineintegral
acc_acc_y_n_above_mean
acc_acc_y_n_below_mean
acc_acc_y_n_sign_changes
acc_acc_y_iqr
acc_acc_y_iqr_5_95
acc_acc_y_pct_5
acc_acc_y_pct_95
acc_acc_y_entropy
acc_acc_y_perm_entropy
acc_acc_y_svd_entropy
acc_acc_z_mean
acc_acc_z_std
acc_acc_z_min
acc_acc_z_max
acc_acc_z_ptp
acc_acc_z_sum
acc_acc_z_energy
acc_acc_z_skewness
acc_acc_z_kurtosis

acc_acc_z_peaks
acc_acc_z_rms
acc_acc_z_lineintegral
acc_acc_z_n_above_mean
acc_acc_z_n_below_mean
acc_acc_z_n_sign_changes
acc_acc_z_iqr
acc_acc_z_iqr_5_95
acc_acc_z_pct_5
acc_acc_z_pct_95
acc_acc_z_entropy
acc_acc_z_perm_entropy
acc_acc_z_svd_entropy
acc_l2_mean
acc_l2_std
acc_l2_min
acc_l2_max
acc_l2_ptp
acc_l2_sum
acc_l2_energy
acc_l2_skewness
acc_l2_kurtosis
acc_l2_peaks
acc_l2_rms
acc_l2_lineintegral
acc_l2_n_above_mean
acc_l2_n_below_mean
acc_l2_n_sign_changes
acc_l2_iqr
acc_l2_iqr_5_95
acc_l2_pct_5
acc_l2_pct_95
acc_l2_entropy
acc_l2_perm_entropy
acc_l2_svd_entropy

Table 1: PCA Feature Table

PC Component	Contributing Feature	PC Component	Contributing Feature
PC1	hrv_svd_entropy	PC2	eda_tonic_max
PC3	acc_l2_std	PC4	acc_l2_mean
PC5	hrv_rms	PC6	acc_acc_x_pct_5
PC7	acc_l2_entropy	PC8	eda_tonic_pct_5
PC9	acc_acc_z_sum	PC10	hrv_lf_hf_ratio
PC11	acc_l2_iqr	PC12	eda_tonic_skewness
PC13	acc_acc_x_sum	PC14	hrv_min_hr
PC15	acc_l2_kurtosis	PC16	acc_l2_kurtosis_1
PC17	eda_phasic_skewness	PC18	hrv_nni_20
PC19	acc_acc_y_skewness	PC20	acc_acc_y_kurtosis
PC21	acc_acc_y_mean	PC22	acc_acc_y_skewness_1
PC23	acc_acc_z_n_sign_changes	PC24	acc_acc_z_std
PC25	acc_acc_x_n_above_mean	PC26	acc_acc_z_n_below_mean
PC27	eda_phasic_pct_5	PC28	acc_acc_y_iqr
PC29	acc_acc_y_n_sign_changes	PC30	acc_l2_n_sign_changes
PC31	eda_phasic_n_sign_changes	PC32	acc_l2_skewness
PC33	acc_l2_n_sign_changes_1	PC34	acc_l2_skewness_1
PC35	acc_l2_n_sign_changes_2	PC36	eda_tonic_perm_entropy
PC37	acc_l2_skewness_2	PC38	eda_tonic_perm_entropy_1
PC39	acc_acc_z_n_sign_changes_1	PC40	eda_phasic_pct_5_1
PC41	acc_acc_x_mean	PC42	eda_phasic_min
PC43	eda_tonic_n_sign_changes	PC44	eda_phasic_min_1
PC45	acc_acc_x_n_sign_changes	PC46	acc_l2_iqr_1
PC47	acc_acc_x_peaks	PC48	eda_tonic_n_sign_changes_1
PC49	acc_acc_x_n_sign_changes_1	PC50	eda_phasic_iqr
PC51	acc_acc_y_n_sign_changes_1	PC52	acc_acc_x_skewness
PC53	eda_tonic_perm_entropy	PC54	acc_acc_x_peaks_1
PC55	eda_tonic_svd_entropy_1	PC56	acc_acc_y_iqr_1

.2 Plots and Figures