Nurses Stress Prediction Wearable Sensors

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Problem Statement

Nurses work in high-stress environments, facing long shifts, emotional strain, and heavy workloads. Prolonged stress leads to burnout, reduced care quality, and impacts their well-being. Early detection of stress is essential for improving both nurse health and patient care.

Wearable sensors offer a non-intrusive way to monitor stress indicators such as:

- **Heart Rate Variability (HRV)**
- **Electrodermal Activity (EDA)**
- **Respiratory Rate**

By utilizing these sensors and applying machine learning, we can predict stress in real-time, enabling timely interventions to support nurse well-being and maintain healthcare quality.

Motivation

1. ## Critical Role of Nurses:-
   Nurses are essential in healthcare, providing critical patient care under high-pressure conditions.

2. ## Consequences of Unmanaged Stress:-
   Prolonged stress can lead to burnout, reduced quality of patient care, increased errors, and high nurse turnover rates.

3. ## Improving Work Environment:-
   A stress prediction model for nurses can enhance work environments, improve well-being, and ensure high-quality patient care.

Literature Review

To ensure a strong foundation for our project, we reviewed key research that explored stress prediction using wearable sensors and machine learning in healthcare settings.

Research Paper 1:

[***Stress Detection in Working People***]

*Sriramprakash.S * , Prasanna Vadana. D, O. V. Ramana Murthy*

Research Paper 2:

[***Multi-Class Stress Detection Through Heart Rate Variability:***]

JON ANDREAS MORTENSEN, MARTIN EFREMOV MOLLOV, AYAN CHATTERJEE, DEBASISH GHOSE, (Senior Member, IEEE), AND FRANK Y.LI

Research Paper 1

- **Physiological Signals**: Data is collected from 25 subjects under three stressor conditions using **GSR** (Galvanic Skin Response) and **HRV** (Heart Rate Variability).
- **Feature Extraction**: Key features are extracted using certain algorithms, focusing on time and frequency domain characteristics for both GSR and HRV.
- **Machine Learning Models**: Extracted features are used to train **SVM** and **KNN** models.
- **Classification**: The models classify subjects as either **Normal** or **Stressed** based on the input features.
- Random Forest Achieves the highest accuracy of **91.34%**

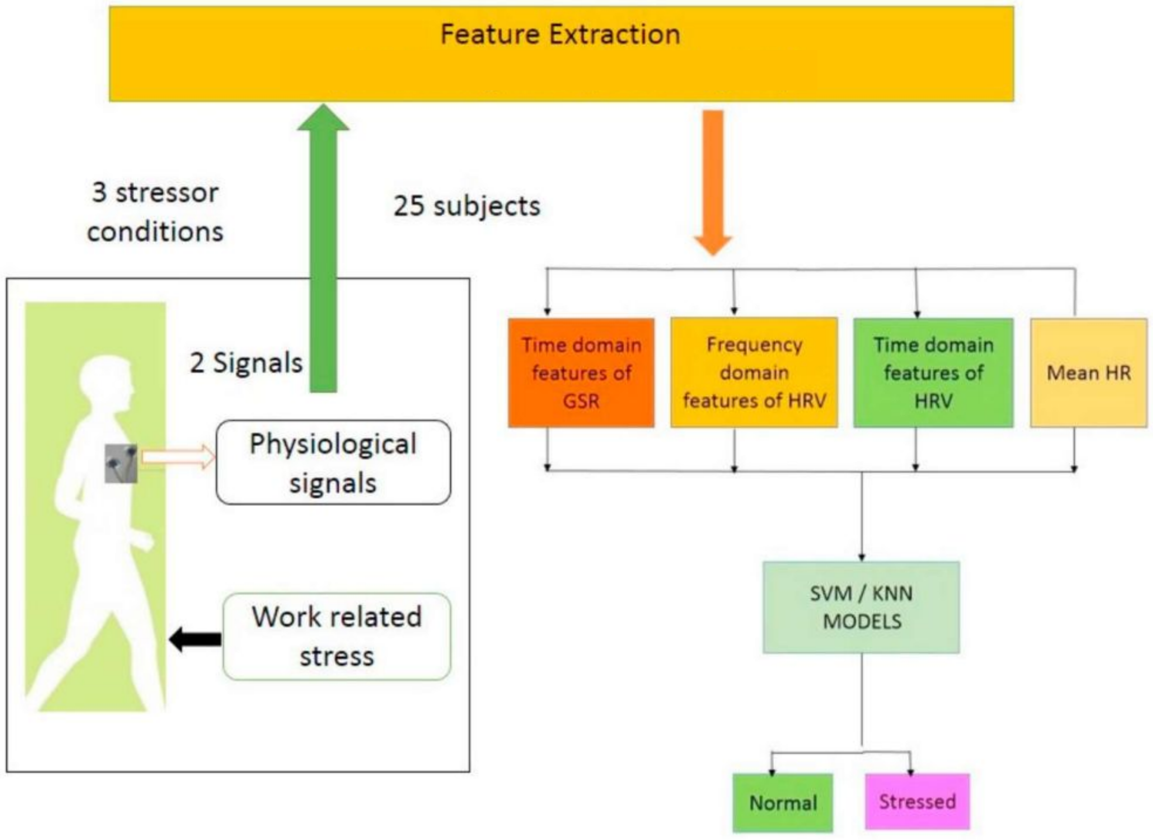


Fig. 1. Overall Framework: Stress Detection-Feature Extraction- Classification

Research Paper 2

- **Data Collection**: HRV signals were gathered from 25 participants during tasks with stressors (email interruptions, time pressure).
- **Data Preprocessing**: Time-series HRV data was cleaned, normalized, and split into training (80%) and testing (20%) sets.
- **Feature Extraction**: Time-domain and frequency-domain features were extracted, and ANOVA F-test was used for feature ranking.
- **Classification**: A 1D CNN model was developed to classify stress into three categories: no stress, time pressure, and interruption.
- **Feature Optimization**: Top features achieved **96.5%** accuracy with reduced computational load.

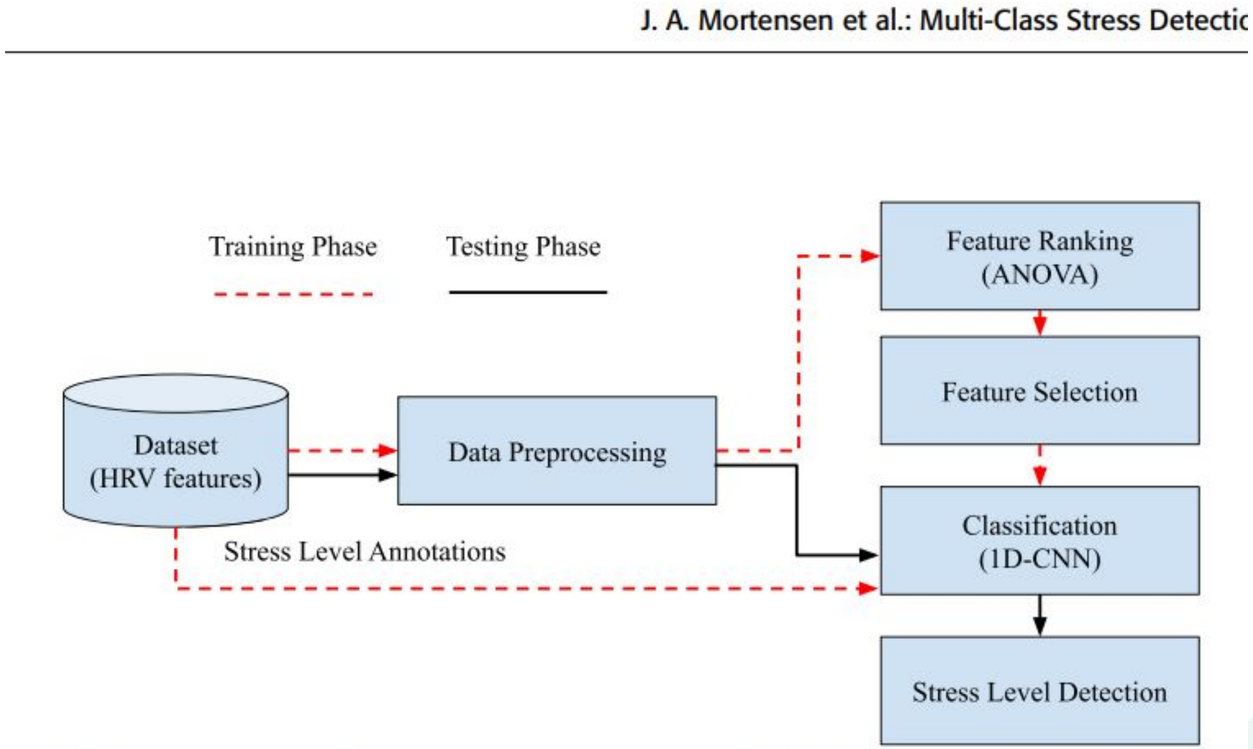
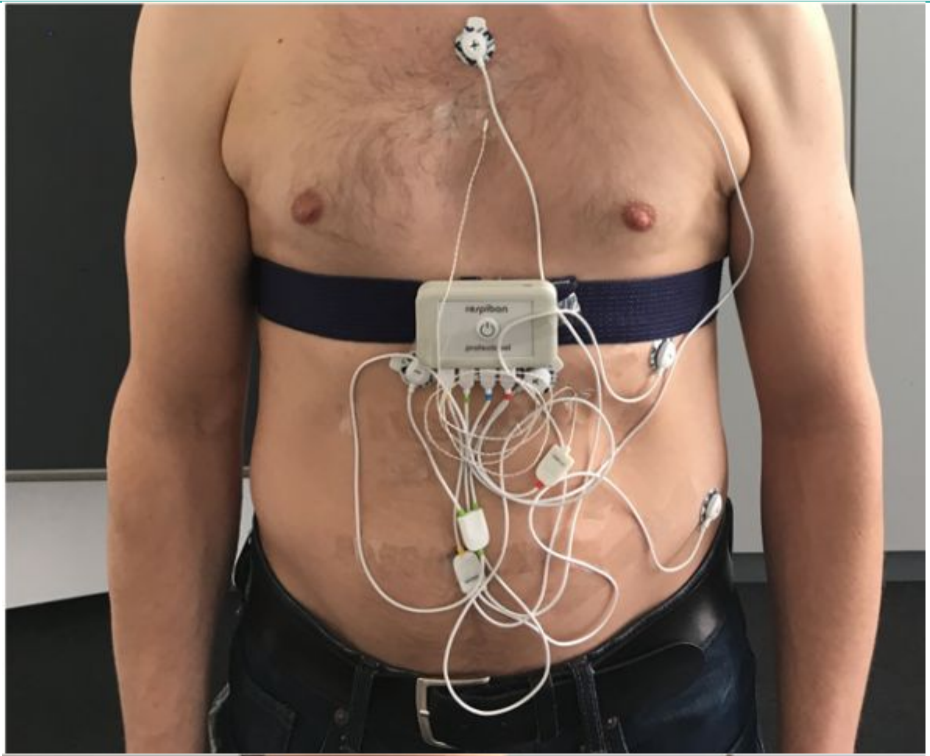


FIGURE 1. **Framework of the proposed stress status classification model: From data collection to stress level classification.**

Dataset description

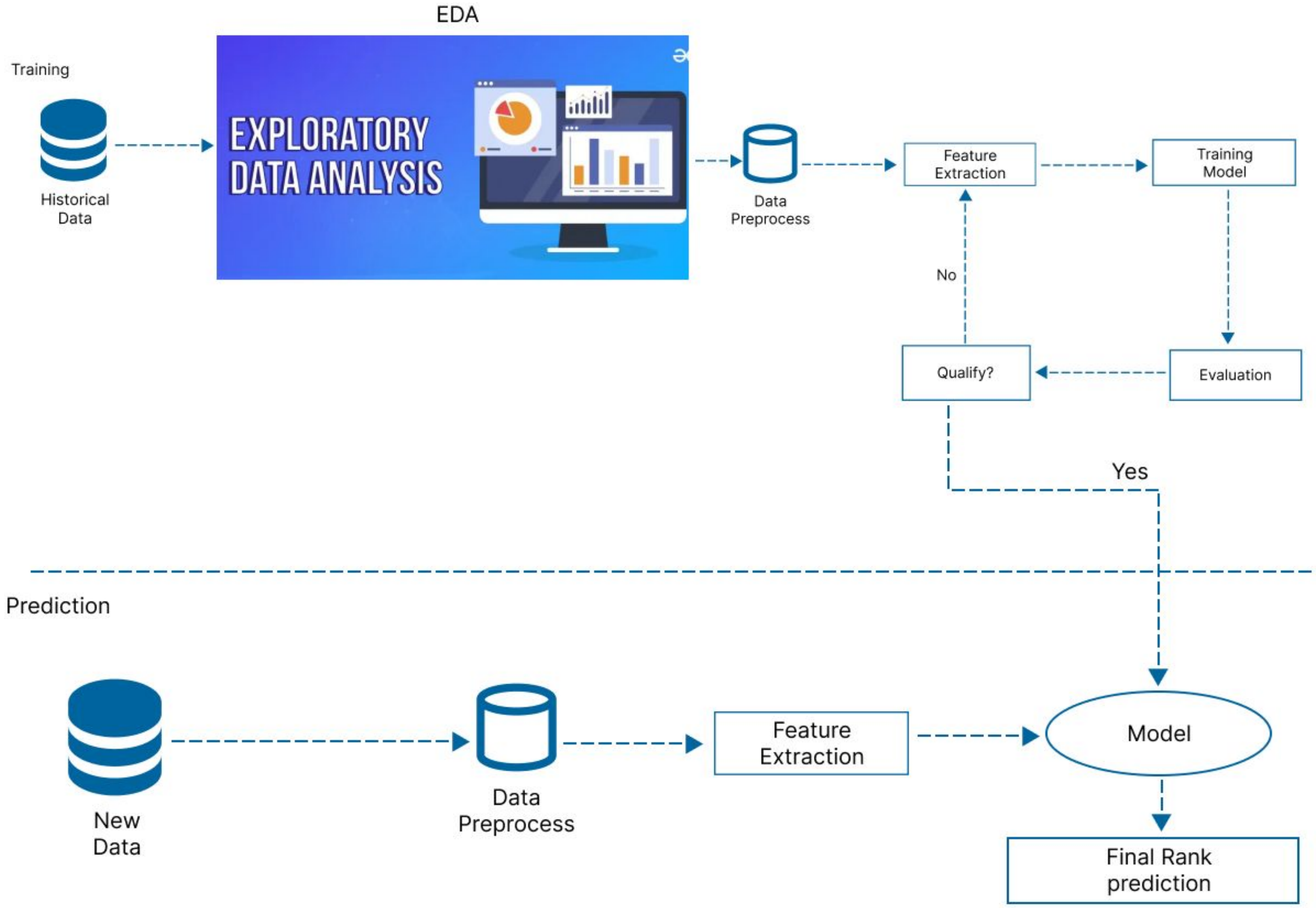
The Swell-KW dataset consists of data from **25 participants** who were knowledge workers, collected to study the effects of stress and workload in a controlled work environment. Participants were exposed to three different conditions: **low stress, time pressure stress, and interruption stress**. Physiological measurements, including heart rate variability (e.g., RMSSD, SDSD), respiration, and other metrics like KURT, SKEW, and their relative variations (REL_RR), were recorded during various work tasks. The dataset also includes self-reported stress and workload assessments, making it valuable for analyzing the relationship between physiological responses and perceived mental states.



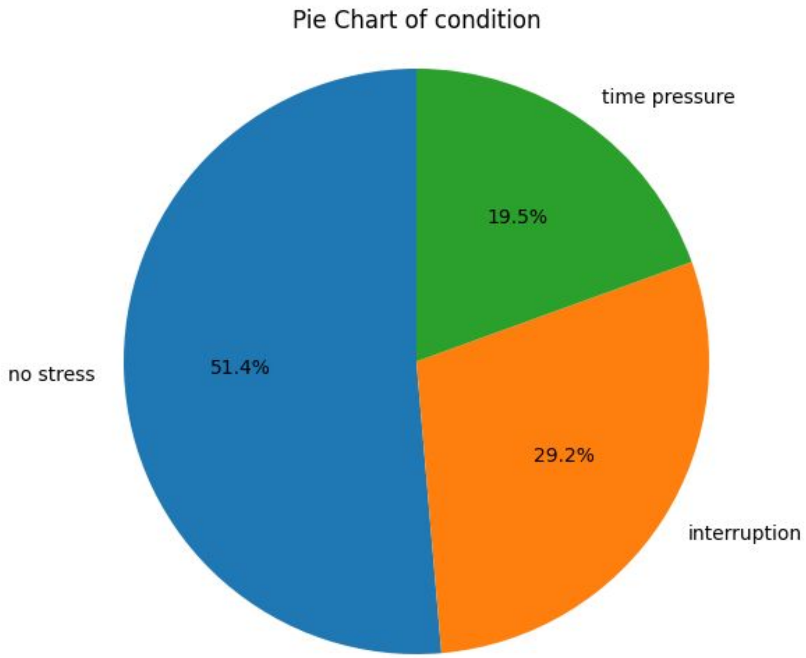
Attributes

No.	Feature	Meaning
1	MEAN_RR	Mean of RR intervals
2	MEDIAN_RR	Median of RR intervals
3	SDRR	SD of RR intervals
4	RMSSD	Root mean square of successive RR interval differences
5	SDSD	SD of successive RR interval differences
6	SDRR_RMSSD	Ratio of SDRR over RMSSD
7	HR	Heart rate
8	pNN25	Percentage of successive RR intervals that differ more than 25 ms
9	pNN50	Percentage of successive RR intervals that differ more than 50 ms
10 . . 34	SD1 . . higuci	Measures short-term HRV in ms and correlates with baroreflex sensitivity (BRS) . . Higuchi Fractal Dimension

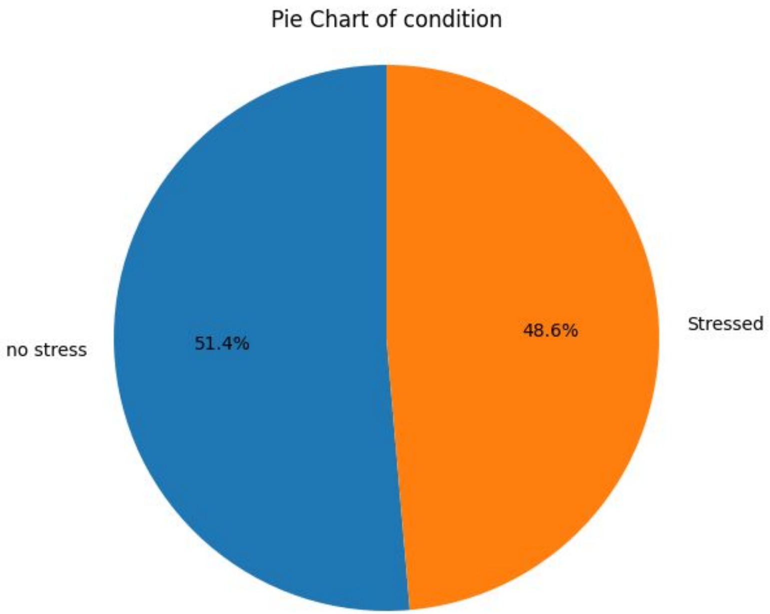
Methodology



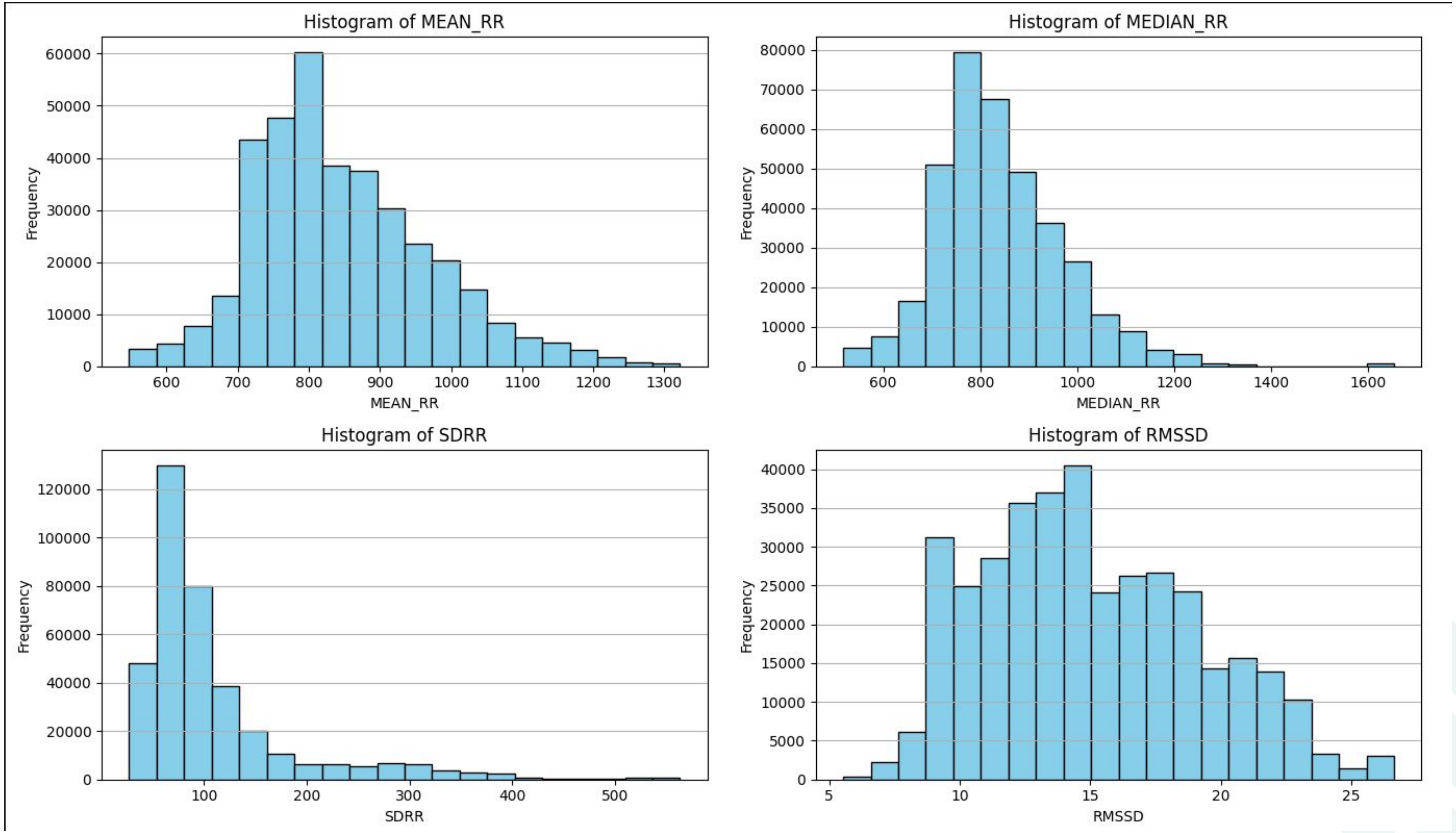
PIE CHART



combining "time pressure" and "interruption" as stressed



Histograms

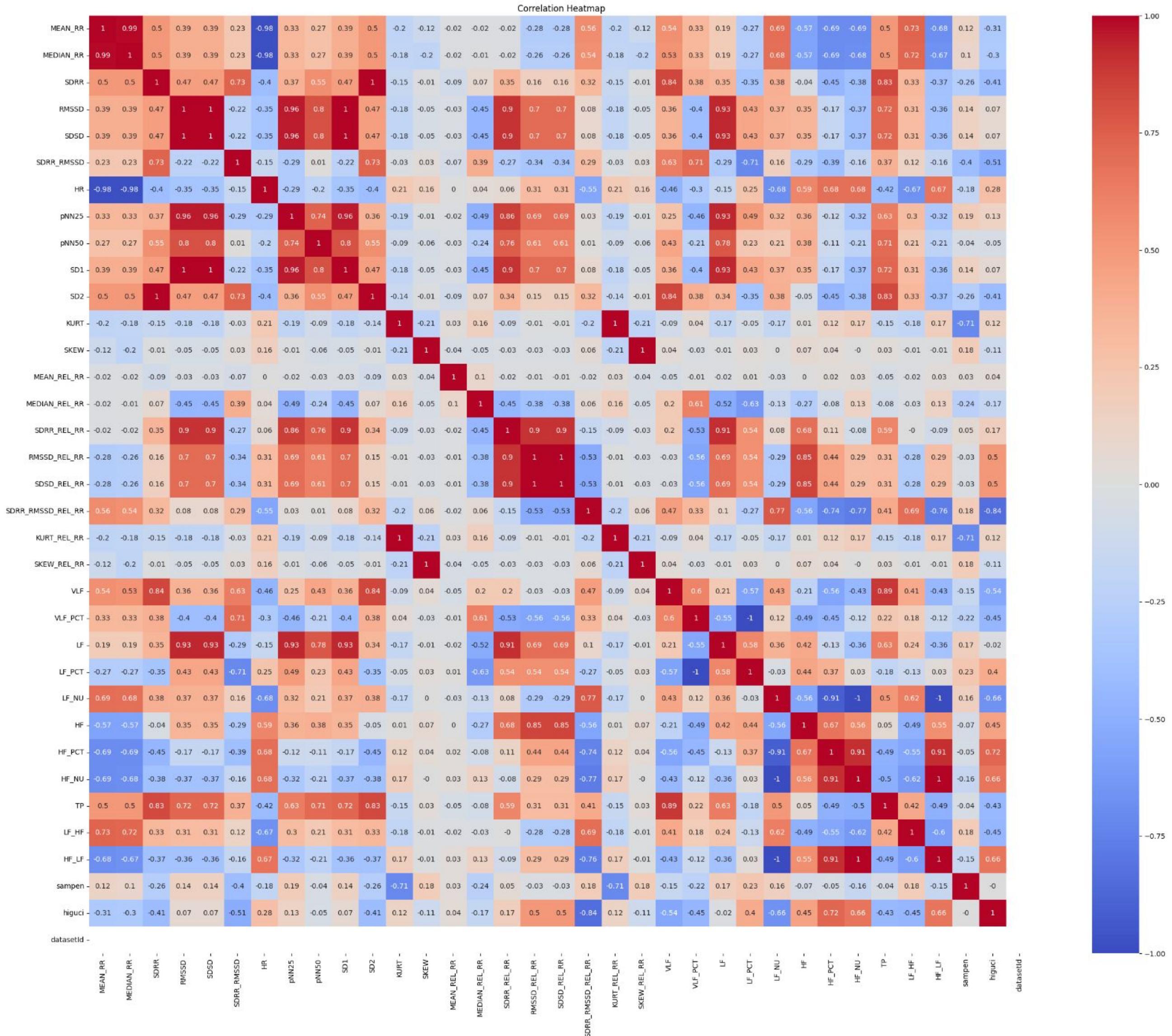

Histogram of MEAN_RR
Frequency
MEAN_RR
0
10000
20000
30000
40000
50000
60000
600
700
800
900
1000
1100
1200
1300
Histogram of MEDIAN_RR
Frequency
MEDIAN_RR
0
10000
20000
30000
40000
50000
60000
70000
80000
600
800
1000
1200
1400
1600
Histogram of SDRR
Frequency
SDRR
0
20000
40000
60000
80000
100000
120000
100
200
300
400
500
Histogram of RMSSD
Frequency
RMSSD
0
5000
10000
15000
20000
25000
30000
35000
40000
5
10
15
20
25

Heatmaps

Understanding Heatmap:

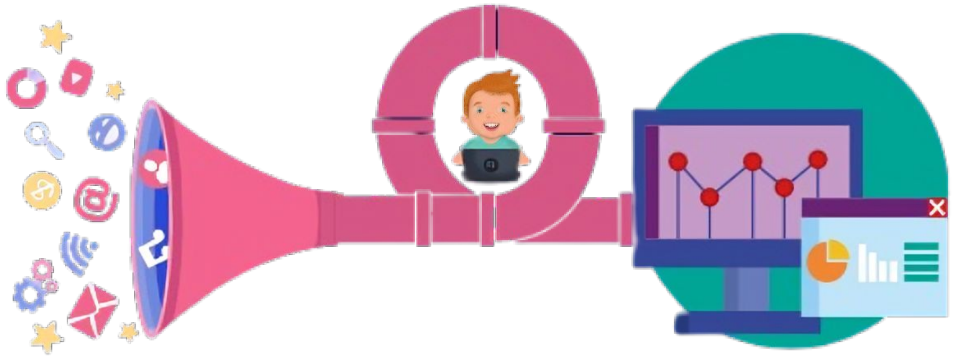
- *Some features exhibit correlations close to 0, indicating they provide independent and useful information.*

- *Some features show strong positive or negative correlations, suggesting they could be combined or reduced to create a more independent and meaningful feature.*

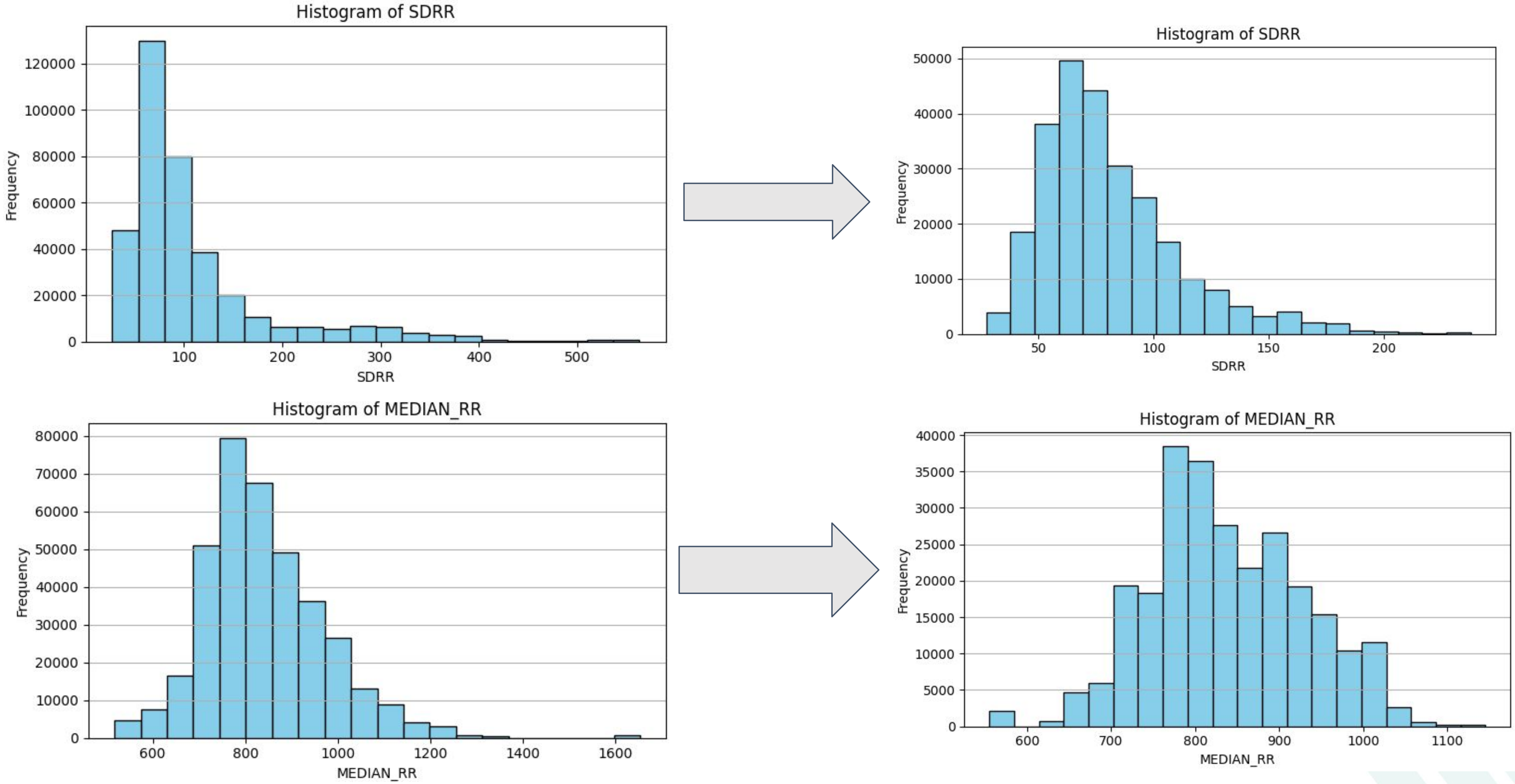


Preprocessing

- Missing Data: Handle missing sensor values.
- Normalization: Normalize physiological signals.
- Outlier Removal: Detect and remove data outliers.
- Conduct t-test: Determines the critical p-value for a two-tailed test at a 0.05 significance level and conducts t-tests.
- PCA: Reduce 34 features to 22 features.



Outlier removal



Preprocessing

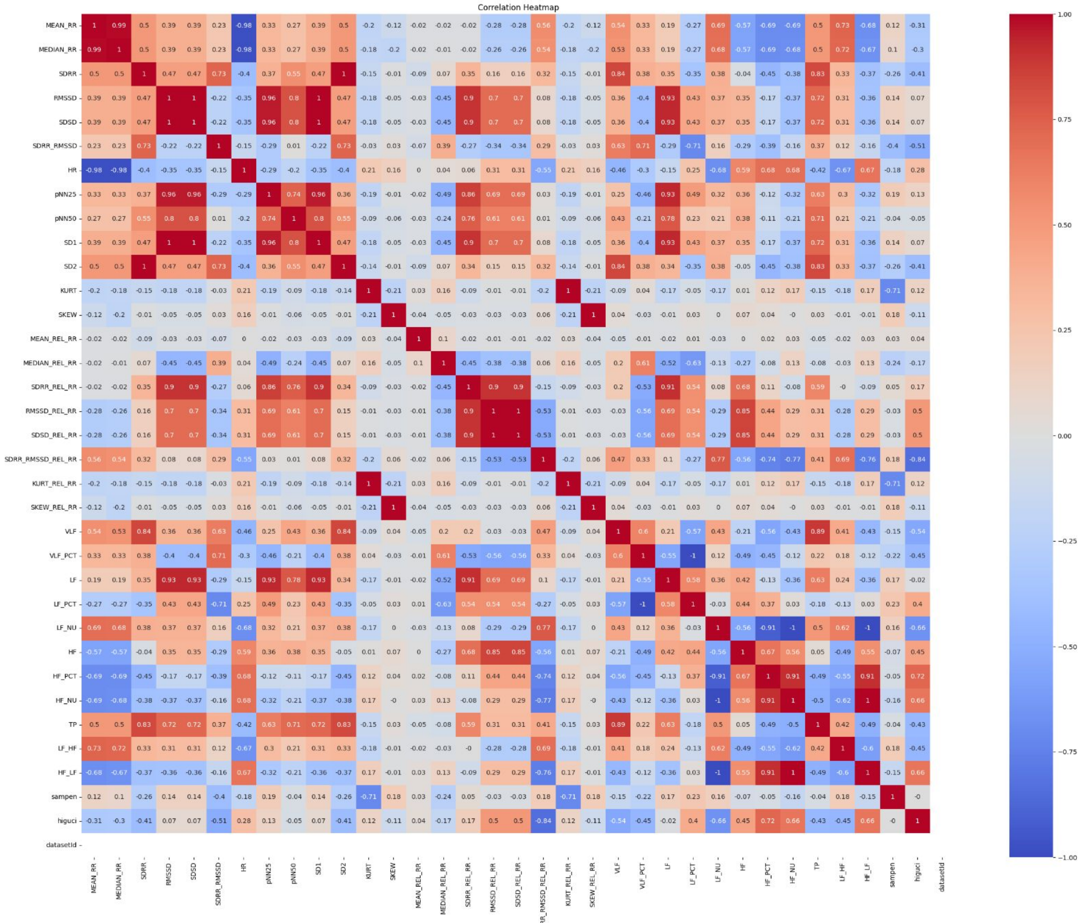
Feature Combination:

- RMSSD_SDSD = (RMSSD + SDSD) / 2
- KURT_MERGED = (KURT + KURT_REL_RR) / 2
- SKEW_MERGED = (SKEW + SKEW_REL_RR) / 2
- RMSSD_SDSD_REL_RR = (RMSSD_REL_RR + SDSD_REL_RR) / 2

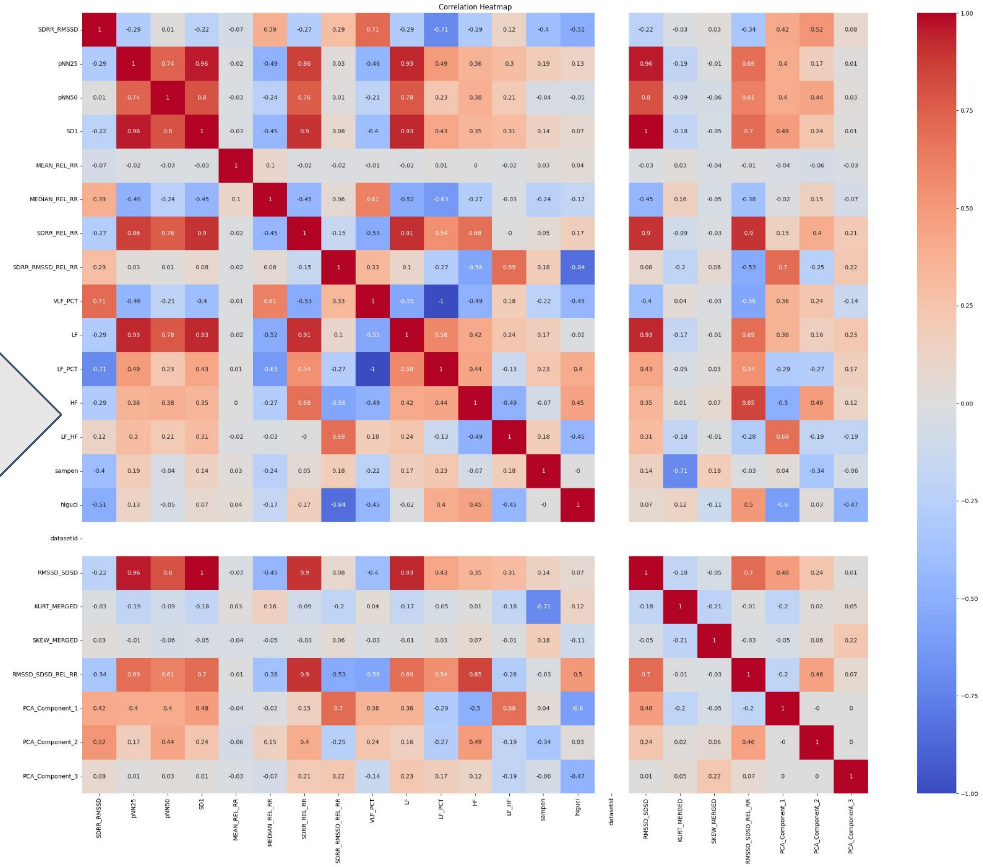
PCA(Principal Component Analysis):

- 'MEAN_RR', 'MEDIAN_RR', 'HR', 'SDRR', 'SD2', 'VLF', 'TP', 'LF_NU', 'HF_PCT', 'HF_NU', 'HF_LF'
- Combine these **11** features to get **3** PCA features
- Final number of features - **22**

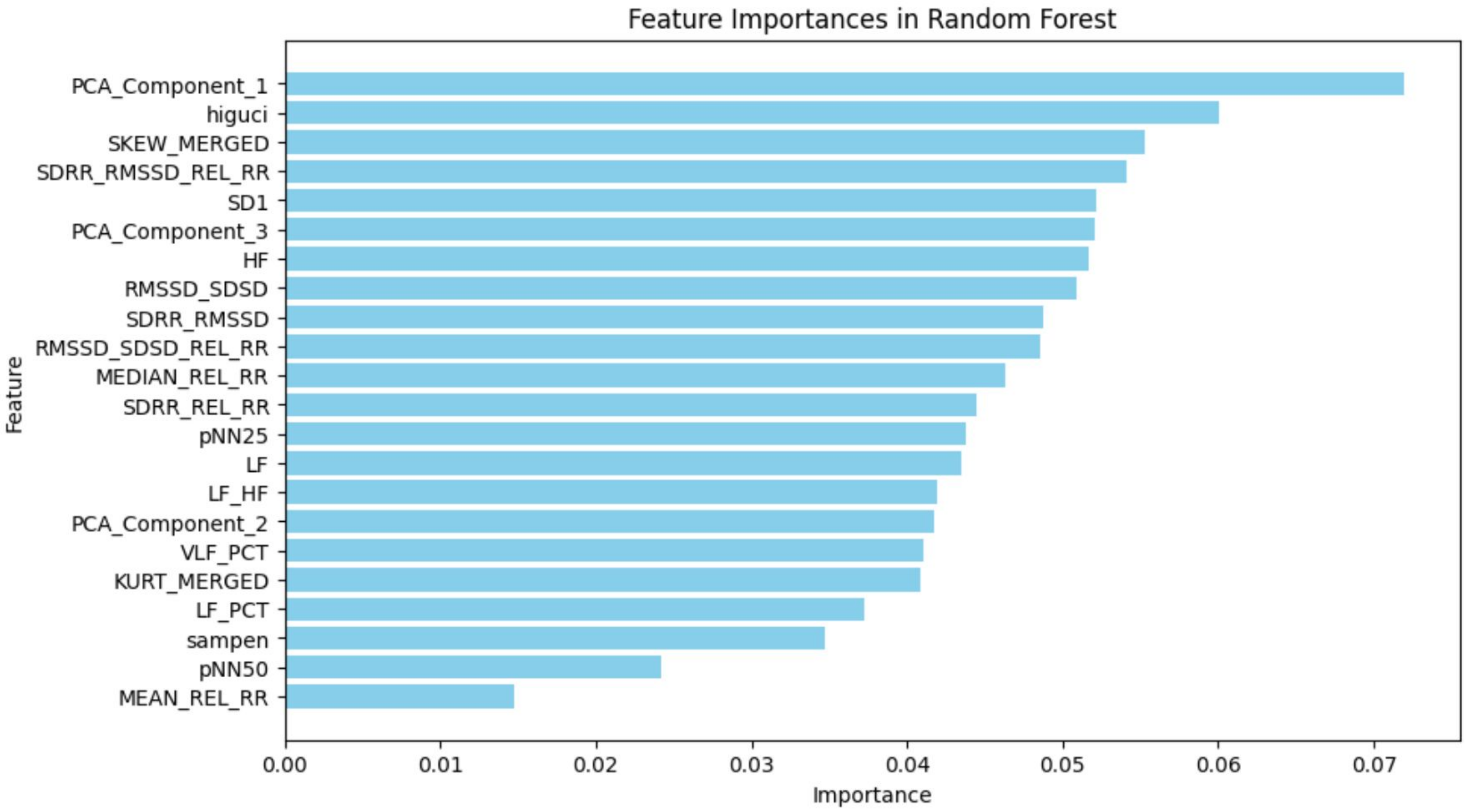
Applying PCA



Applying PCA and feature combination



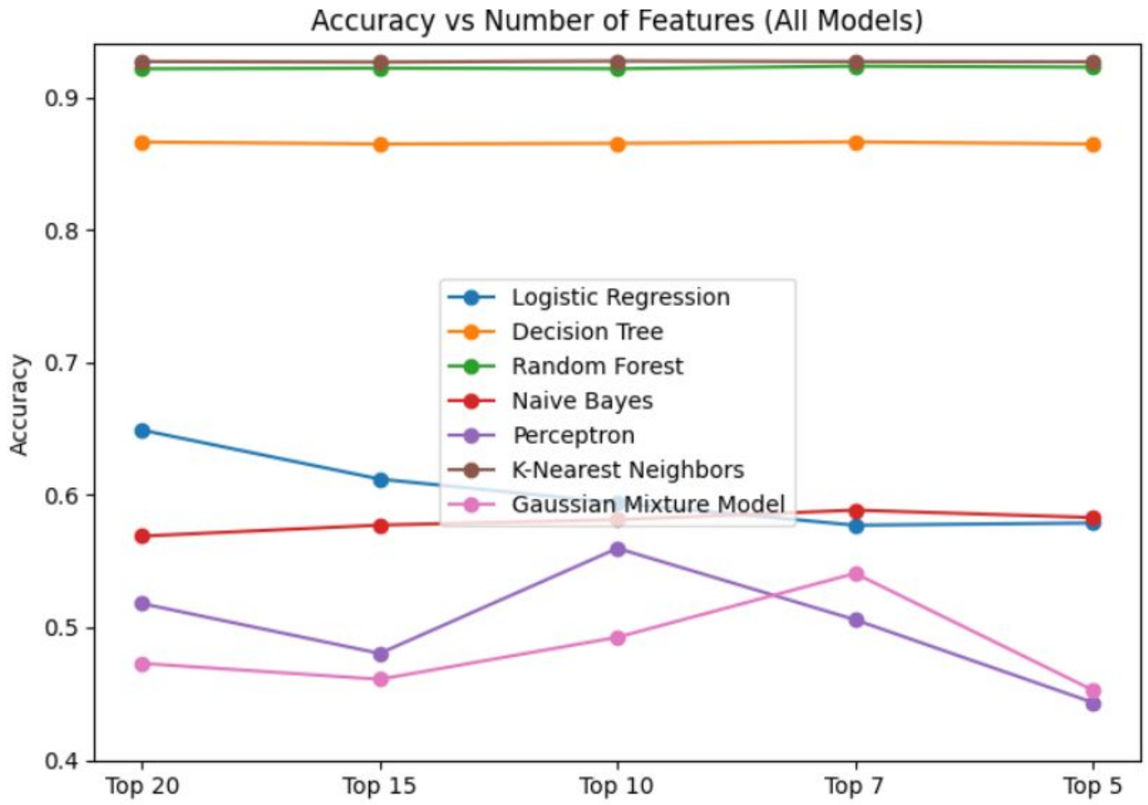
Feature Importance in Random Forest

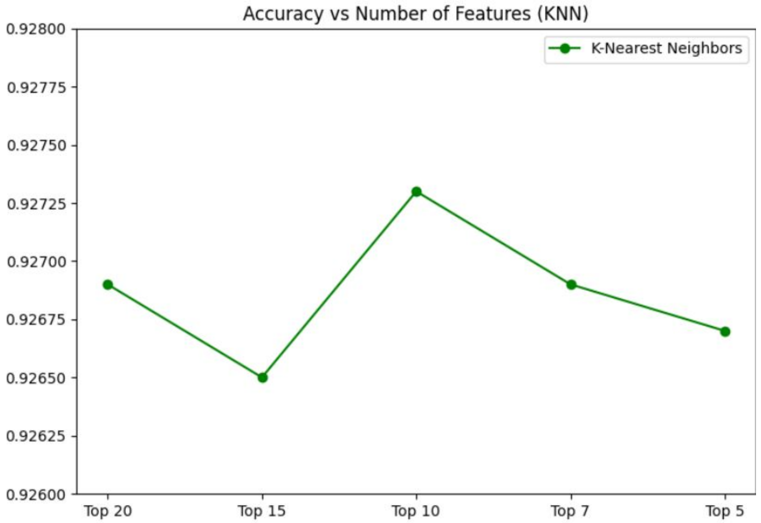


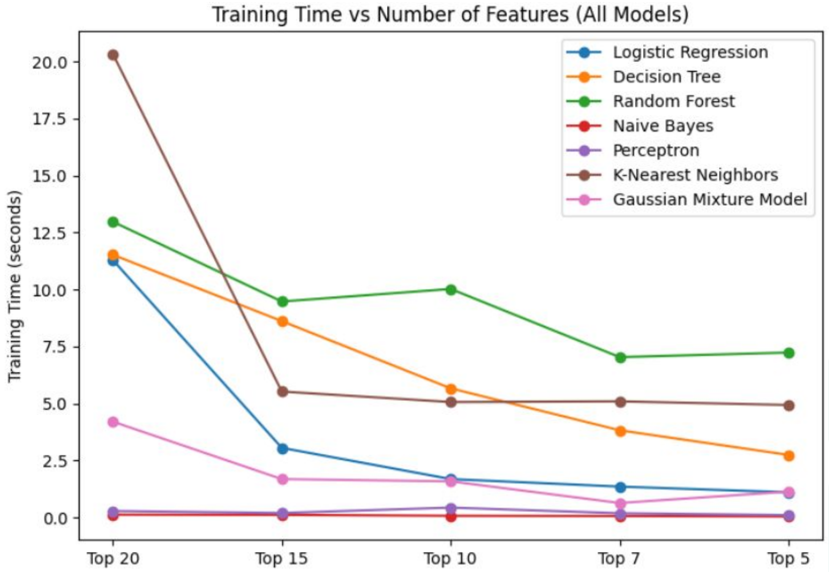
Conclusion

Features	Logistic Regression	Decision Tree	Random Forest	Naive Bayes	Perceptron	KNN	GMM
Top 20	0.6489	0.8662	0.9214	0.5689	0.518	0.9269	0.4729
Top 15	0.6119	0.8647	0.9219	0.5772	0.4804	0.9265	0.4609
Top 10	0.5939	0.8652	0.9215	0.5813	0.5597	**0.9273**	0.4926
Top 7	0.5771	0.8663	0.9234	0.5885	0.5058	0.9269	0.5408
Top 5	0.5788	0.8647	0.9226	0.5829	0.4434	0.9267	0.4527

Conclusion


Accuracy vs Number of Features (All Models)
Accuracy
0.9
0.8
0.7
0.6
0.5
0.4
Top 20
Top 15
Top 10
Top 7
Top 5
Logistic Regression
Decision Tree
Random Forest
Naive Bayes
Perceptron
K-Nearest Neighbors
Gaussian Mixture Model


Accuracy vs Number of Features (KNN)
0.92800
0.92775
0.92750
0.92725
0.92700
0.92675
0.92650
0.92625
0.92600
Top 20
Top 15
Top 10
Top 7
Top 5
K-Nearest Neighbors


Training Time vs Number of Features (All Models)
Training Time (seconds)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Top 20
Top 15
Top 10
Top 7
Top 5
Logistic Regression
Decision Tree
Random Forest
Naive Bayes
Perceptron
K-Nearest Neighbors
Gaussian Mixture Model

Analysing Results

KNN Performs Best:

- KNN consistently outperforms the other models, especially with the top 10 features, achieving the highest accuracy of **0.9273**.

Training Time Analysis:

- Decision Trees and Logistic Regression may be preferable for scenarios requiring rapid model updates, while **KNN** excels in accuracy at the cost of slower computation.

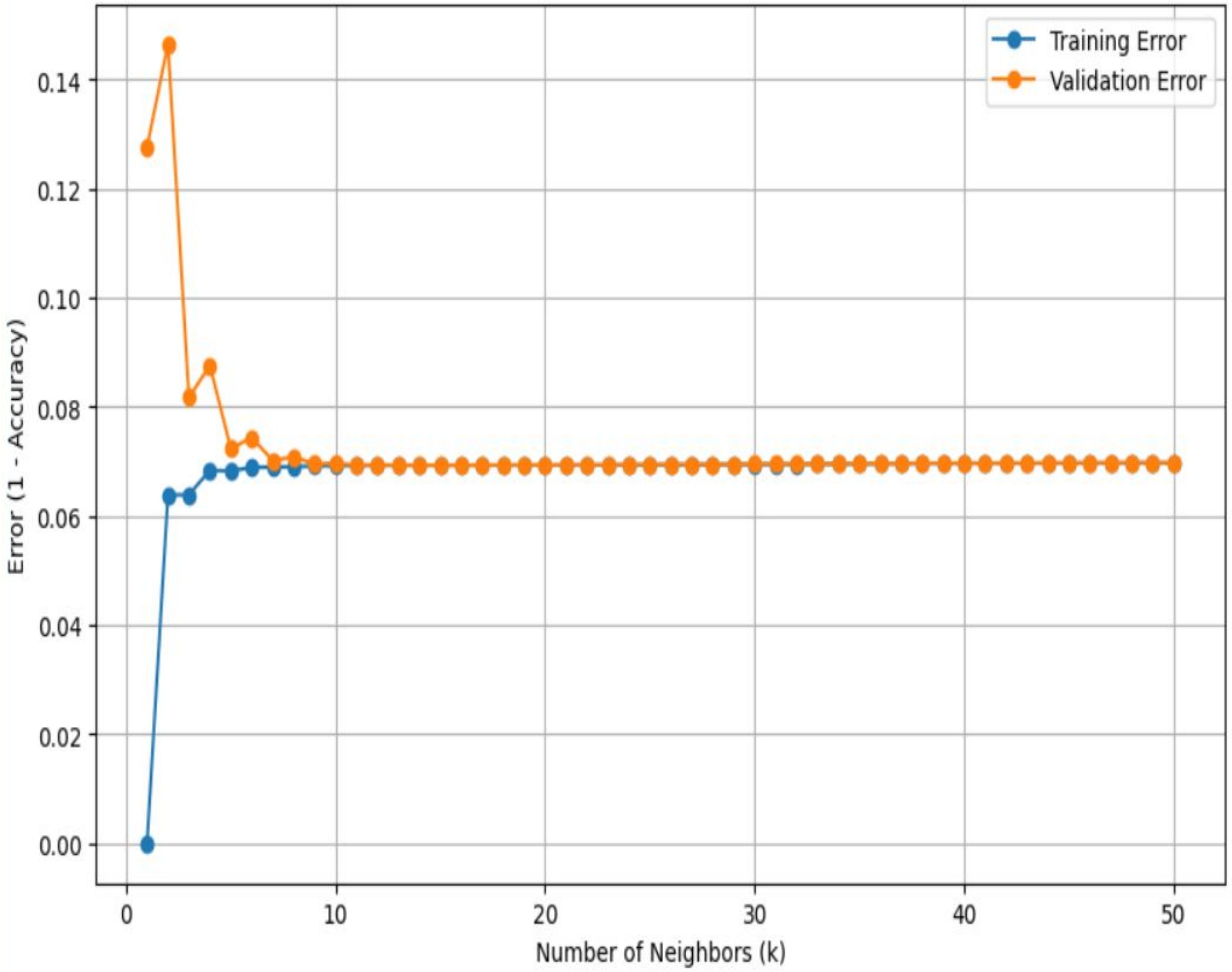
Feature Reduction Effect:

- As features are reduced, the performance of some models (like Logistic Regression and Naive Bayes) decreases slightly, but KNN maintains high performance.

Best Feature Count for Performance:

- The best accuracy is achieved with the **top 10 features** for KNN and it takes almost least time or computation.

Bias-Variance Tradeoff for KNN



Key Insights:

1. **Optimal k Range:** Around `k=10`, the validation error is minimized
2. **Overfitting for Small k:** A small value of `k(<4)` results in overfitting.
3. **Model Generalization:** Larger values of `k` lead to more generalization though slightly increasing the training error.

Power Features

- *PCA_Component_1*
- *higuci*
- *SKEW_MERGED*
- *SDRR_RMSSD_REL_RR*
- *SD1*
- *PCA_Component_3*
- *HF*
- *RMSSD_SDSD*
- *SDRR_RMSSD*
- *RMSSD_SDSD_REL_RR*

Timeline

Week 1-2
Problem Understanding & Data Preprocessing

Week 3-4
Exploratory Data Analysis & Feature Selection

Week 5-6
Model Development & Training

Week 7-8
Model Testing & Validation

Week 9-10
Documentation

Individual Contribution

- **Namit Jain (2022315):-** PPT creation, Programming
- **Saurav Haldar (2022464):-** Report Creation, Programming
- **Saarthak Saxena (2022421):-** Report Creation, Programming
- **Satwik Garg (2022461):-** PPT creation, Programming