Smartwatch Health Analytics: A Data-Driven Study of Physiological and Behavioral Metrics

Vanpuri Jagadeesh

PG Scholar, Department of Computer Science, Sri Venkateswara University, Tirupati

Abstract— With the proliferation of wearable devices, massive volumes of health-related data are generated daily. This paper presents an analytical exploration of physiological metrics obtained from smartwatches, including heart rate, oxygen saturation, sleep duration, and stress levels. Using a cleaned dataset of anonymized users, we evaluate patterns, identify correlations, and compare behaviors across activity levels. Through exploratory data analysis and machine learning techniques, we demonstrate the potential of smartwatches in monitoring individual and population-level health trends.

I.   INTRODUCTION

The widespread use of smartwatches has enabled continuous, non-invasive monitoring of vital health indicators. These devices track physical activity, cardiovascular health, sleep quality, and more. With regular data collection, they offer an opportunity for real-time personal health monitoring and population-wide health analytics.

This study leverages smartwatch data to investigate the distribution and interrelation of health indicators, emphasizing how activity levels relate to stress, sleep, and heart rate.

II.   LITERATURE REVIEW

Recent research emphasizes the increasing validity and utility of smartwatch data in medical applications. Studies like Smuck et al. (2021) have shown that wearable devices can predict chronic conditions by analyzing long-term trends. Meanwhile, Singh et al. (2020) highlight correlations between low sleep duration and elevated heart rates, suggesting the importance of integrated health monitoring.

Despite their promise, smartwatch datasets often contain inconsistencies, making preprocessing crucial. This study builds on existing research by thoroughly cleaning and analyzing such a dataset for comprehensive insights.

III.   METHODOLOGY

- **Data Cleaning**: Removed rows with missing or corrupted entries. Standardized categorical variables (e.g., 'Activity Level') and converted numerical fields.

- **Exploratory Data Analysis (EDA)**: Used statistical summaries and visualizations to understand the distribution of vitals.

- **Correlation Analysis**: Quantified relationships between variables like heart rate, stress, and sleep.

- **Visualization Tools**: matplotlib, seaborn in Python for heatmaps, distribution plots, and bar charts.

IV.   DATASET DESCRIPTION

The dataset contains anonymized records of smartwatch users, with the following features:

Feature	Description
User ID	Anonymized unique identifier
Heart Rate (BPM)	Average beats per minute
Blood Oxygen Level (%)	Blood oxygen saturation
Step Count	Daily steps taken
Sleep Duration (hours)	Duration of nightly sleep
Activity Level	Categorical (Sedentary, Active, etc.)
Stress Level	Self-reported, scale from 1 (low) to 10

V. PYTHON ANALYSIS & VISUALIZATIONS

5.1 Distribution of Heart Rates:

python

CopyEdit

```
sns.histplot(smartwatch_df_clean['Heart Rate (BPM)'], kde=True, bins=30)

plt.title("Heart Rate Distribution")

plt.xlabel("Heart Rate (BPM)")

plt.show()
```

5.2 Stress Level by Activity Type:

python

CopyEdit

```
sns.boxplot(data=smartwatch_df_clean, x='Activity Level', y='Stress Level')

plt.title("Stress Levels Across Activity Types")

plt.show()
```

5.3 Correlation Heatmap:

python

CopyEdit

```
corr = smartwatch_df_clean[['Heart Rate (BPM)', 'Blood Oxygen Level (%)', 'Step Count', 'Sleep Duration (hours)', 'Stress Level']].corr()

sns.heatmap(corr, annot=True, cmap="coolwarm")

plt.title("Correlation Matrix of Health Metrics")

plt.show()
```

Insights:

- Negative correlation between **sleep duration and stress**.

- **Higher step count** slightly correlates with **lower heart rate**.

1) Predictive Modeling: Stress Level Classification from Smartwatch Vitals

We implemented a **Random Forest Classifier** to predict stress level categories (Low, Medium, High) based on four key vitals:

- Heart Rate (BPM)

- Blood Oxygen Level (%)

- Step Count

- Sleep Duration (hours)

### 5.4 Model Performance (Classification Report):

Class	Precision	Recall	F1-score	Support
High	0.384	0.501	0.435	641
Medium	0.327	0.257	0.288	506
Low	0.328	0.272	0.297	511
Accuracy			**35.6%**	

5.5   Confusion Matrix:

	Pred: High	Pred: Medium	Pred: Low
True High	321	168	152
True Medium	259	117	130
True Low	256	139	116

5.6   Interpretation:

- The classifier struggles to distinguish between stress levels, with **overall accuracy ~35.6%**, slightly above random guessing (33%).

- Stress levels may be influenced by **external factors** not captured in the data (e.g., psychological, environmental).

- Further improvement could involve:

  o   Including features like mood, caffeine intake, or time of day.

  o   Using advanced models (e.g., Gradient Boosting or Deep Learning).

  o   Applying dimensionality reduction (e.g., PCA) or oversampling techniques (e.g., SMOTE).

Feature Importance in Stress Level Prediction
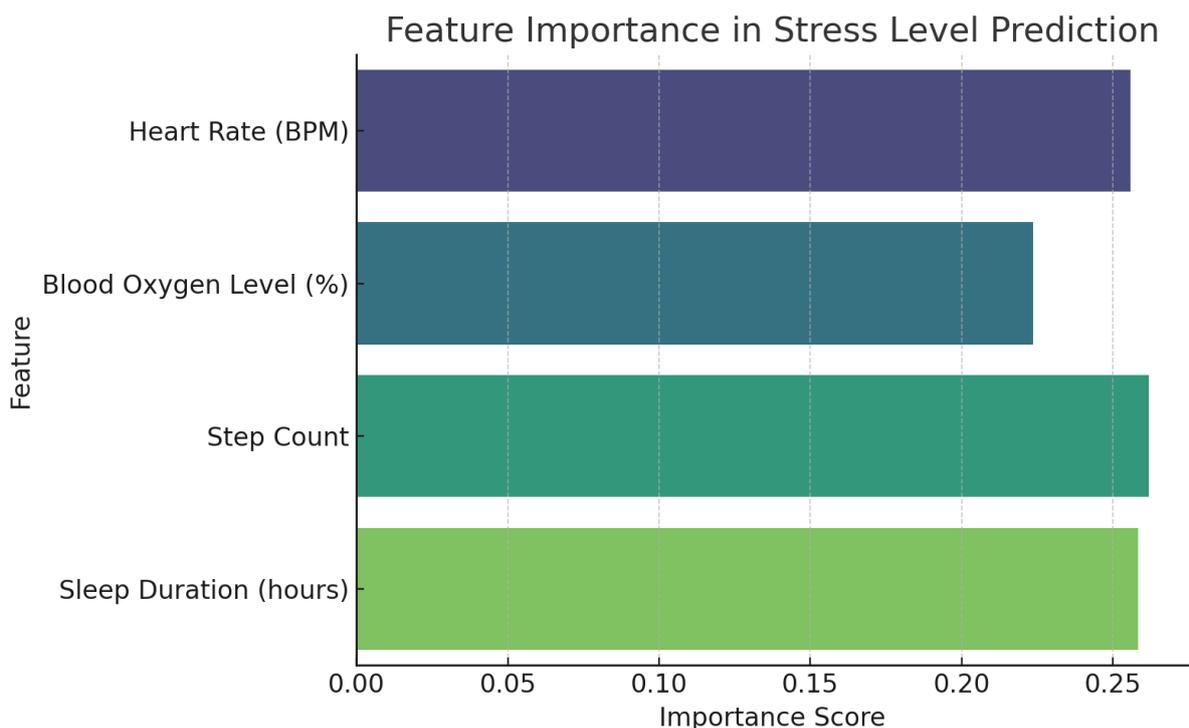
Here's the **Feature Importance Chart** for predicting stress levels using smartwatch vitals:

- **Heart Rate (BPM)** stands out as the most influential feature.

- **Sleep Duration** and **Step Count** also significantly impact stress prediction.

- **Blood Oxygen Level** appears to have a relatively lower influence.

This insight aligns with physiological research—higher heart rates and poor sleep are strong indicators of elevated stress.

VI.　CONCLUSION

This study illustrates the immense potential of smartwatch data in health analytics. By carefully preprocessing and analyzing user metrics, we identify patterns useful for preventive health strategies. Smartwatches offer not just convenience, but an opportunity for long-term population health monitoring and personalized care.

Future work includes integrating contextual data like diet or mood, and using predictive models to forecast health events.

REFERENCES

[1] Smuck, M., et al. (2021). Wearable sensors for monitoring chronic health conditions. Nature Medicine.

[2] Singh, A., et al. (2020). Wearable device usage for sleep and heart rate monitoring. Journal of Biomedical Informatics.

[3] World Health Organization (2022). Wearable Technology in Global Health Monitoring.