Leveraging Machine Learning Models and Wearable Technology to Identify and Predict Stress Levels in Healthcare Professionals

Presented by – Mandeep Kaur

Connect with me:

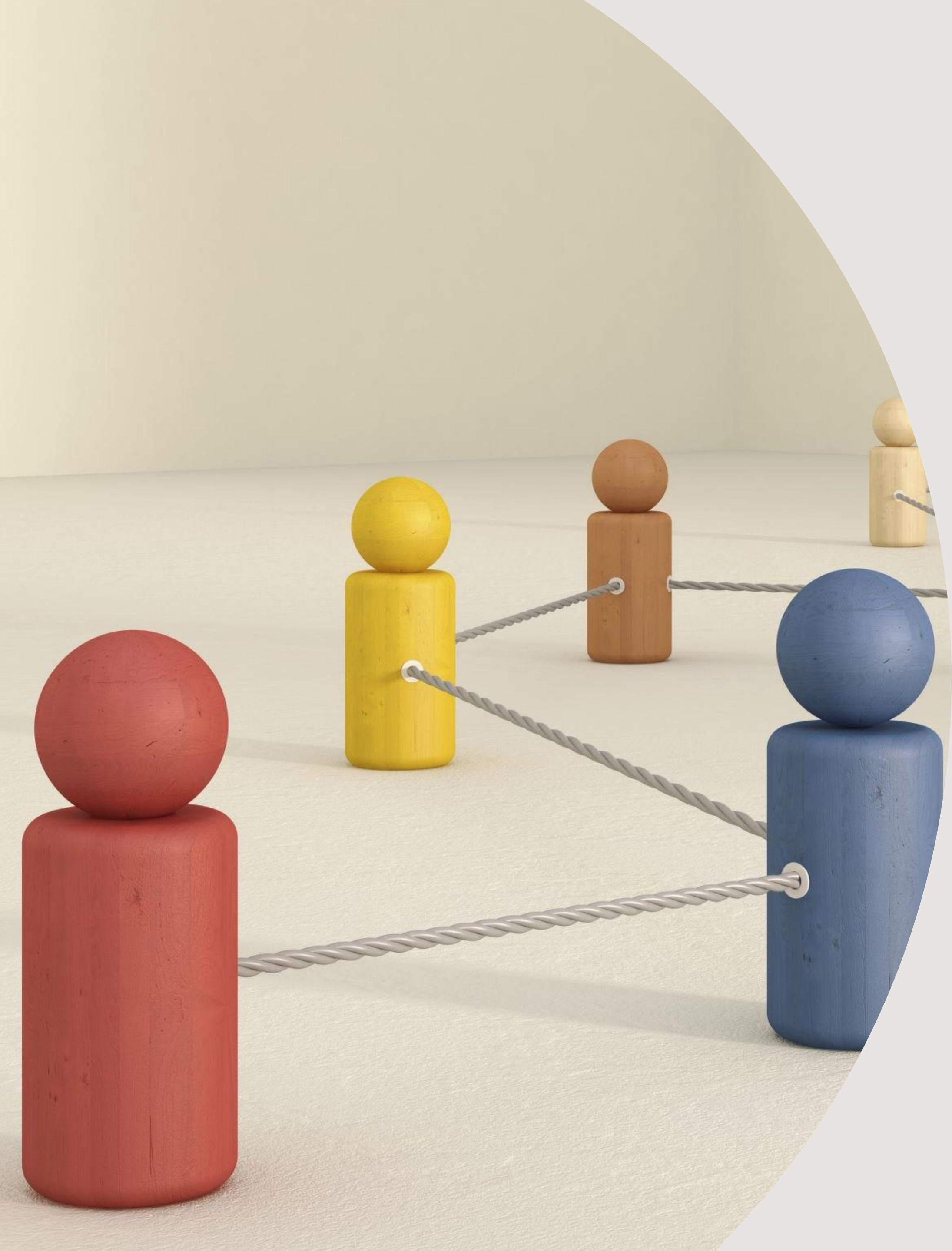
Curtin University (Institute of Data)

LinkedIn Profile: www.linkedin.com/in/mandeep-kaur-20a6a6181

GitHub Account: KaurMandeep870 (github.com)

Agenda:

- Introduction and Problem Statement
- Data Exploration and Preprocessing
- Feature Engineering
- Model Selection & Development
- Model Evaluation
- Model Interpretation & Insights
- Recommendations & Conclusion



Stakeholders:

1. Healthcare Workers (End Users)
2. Healthcare Administrators and Management
3. Occupational Health and Safety Teams
4. IT and Data Security Teams
5. Wearable Device Manufacturers
6. Data Scientists and Researchers
7. Healthcare Policy Makers and Regulators
8. Mental Health Professionals
9. Investors and Funding Agencies
10. Patients

Introduction:

Healthcare professionals are frequently exposed to high-stress environments, which can significantly impact their mental health, decision-making abilities, and patient care. Monitoring and predicting stress levels in real-time could provide actionable insights to alleviate stress and prevent long-term negative health effects.

This project aims to utilize wearable technology to collect a combination of physiological and motion data, such as ***heart rate (HR)****,* ***electrodermal activity (EDA)****,* ***body temperature****, and* ***orientation*** *(body posture or movement). Using this multimodal data, machine learning models will be developed to classify and predict stress levels in healthcare professionals, categorized as low, medium, or high stress.*

Expected Outcomes:

***Accurate Stress Prediction**: Develop models capable of accurately predicting stress levels based on real-time multimodal data.*

***Early Intervention**: Enable real-time stress detection and alerts, allowing healthcare professionals to take proactive steps to reduce stress, such as brief relaxation techniques or workload adjustments.*

***Improved Well-Being**: Provide insights into the relationship between physical activity, physiological responses, and stress, contributing to healthier work environments for healthcare professionals.*

Data Selection & Exploration: Source: DRYAD/Kaggle

Orientation Data: Analyze posture, movement, and activity level using accelerometer/gyroscope data from wearable device Empatica E4.

For instance, certain postures or patterns of movement could indicate increased physical stress or fatigue.

Electrodermal Activity (EDA): Measure sweat gland activity as a direct indicator of sympathetic nervous system activation, which is closely associated with stress.

Heart Rate (HR): Track changes in heart rate and variability to assess physiological arousal.

Temperature: Monitor body temperature as it may correlate with stress-induced thermoregulation.

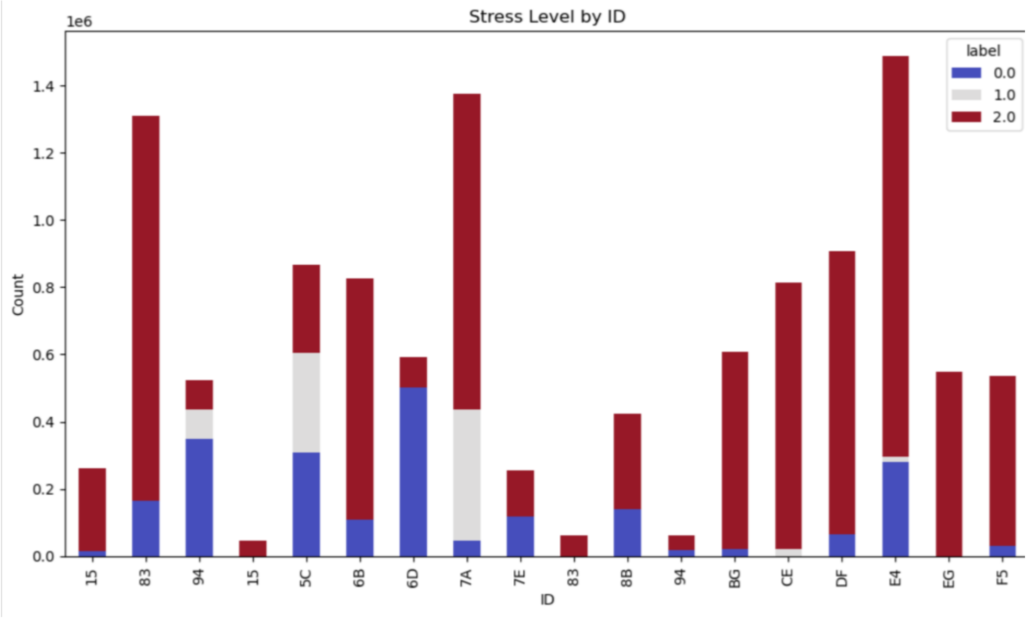
	X	Y	Z	EDA	HR	TEMP	label
count	1.150905e+07	1.150905e+07	1.150905e+07	1.150905e+07	1.150905e+07	1.150905e+07	1.150905e+07
mean	-2.823775e+01	-9.091032e-01	2.382449e+01	3.502127e+00	8.576090e+01	3.223435e+01	1.554202e+00
std	3.142310e+01	3.343382e+01	2.985317e+01	5.656541e+00	1.419642e+01	2.260516e+00	7.891827e-01
min	-1.280000e+02	-1.280000e+02	-1.280000e+02	0.000000e+00	5.100000e+01	2.409000e+01	0.000000e+00
25%	-5.200000e+01	-1.900000e+01	4.000000e+00	2.242060e-01	7.672000e+01	3.019000e+01	1.000000e+00
50%	-3.300000e+01	1.000000e+00	2.500000e+01	1.157407e+00	8.390000e+01	3.257000e+01	2.000000e+00
75%	-1.400000e+01	1.700000e+01	4.900000e+01	4.077436e+00	9.310000e+01	3.425000e+01	2.000000e+00
max	1.270000e+02	1.270000e+02	1.270000e+02	5.976071e+01	1.802300e+02	3.659000e+01	2.000000e+00

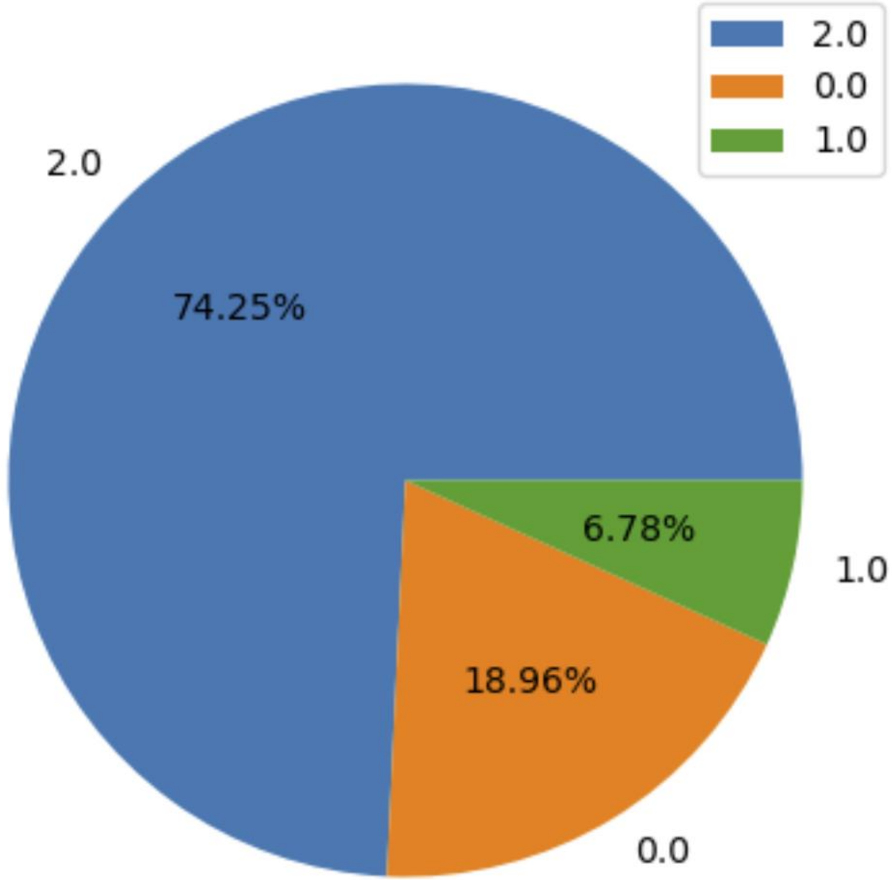
Distribution of Stress Among Participants:

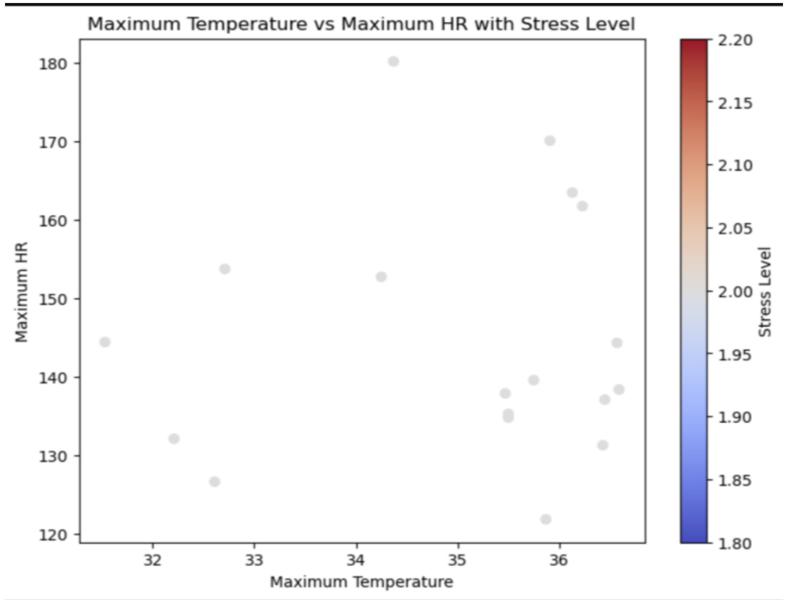
1.0 = Low Stress

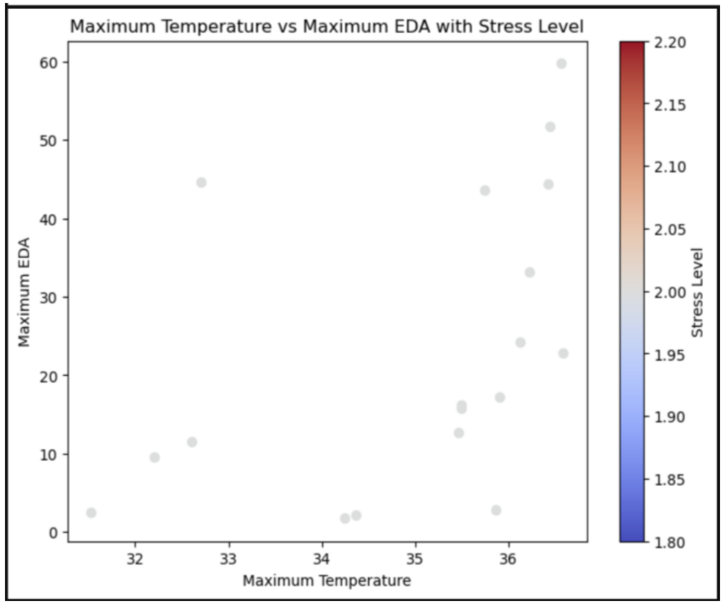
0.0 = Medium Stress

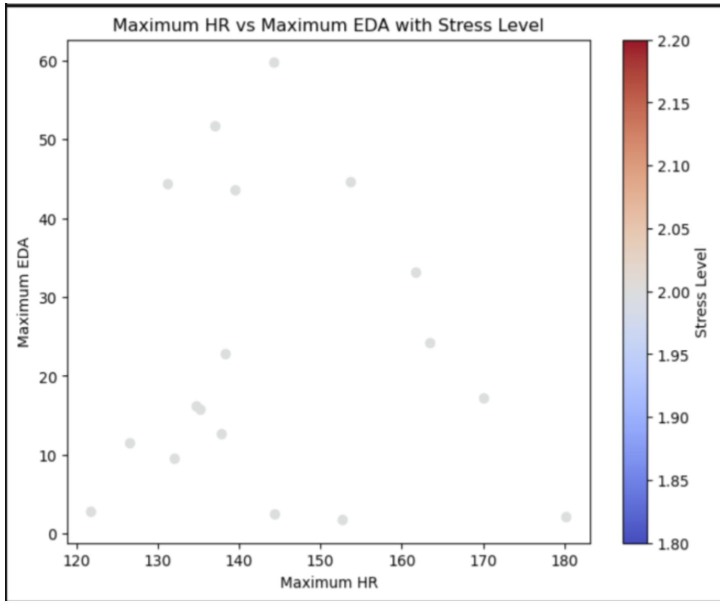
2.0 High Stress







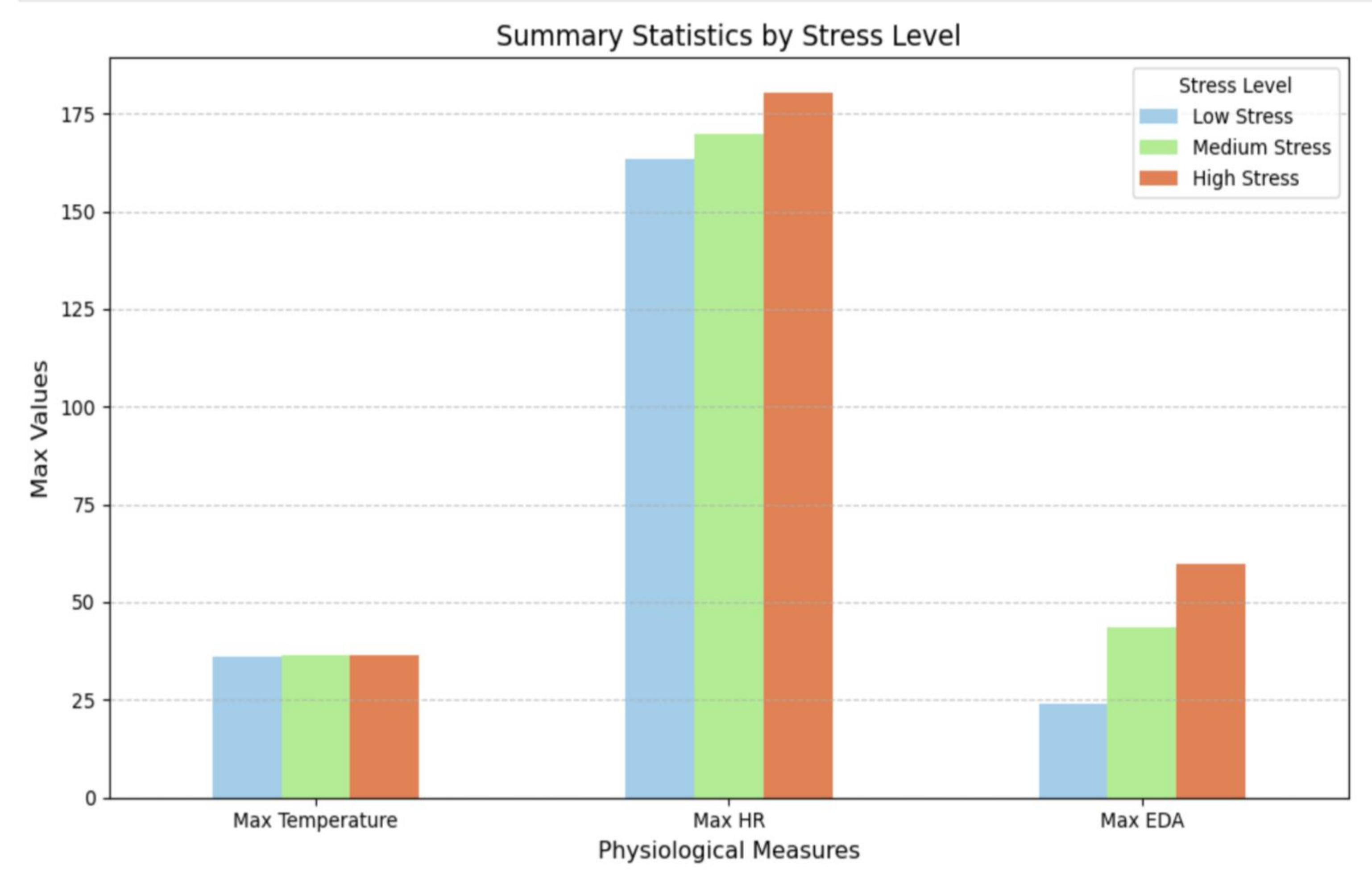




Maximum values of temperature, HR, and EDA based on Stress Levels

	Low Stress	Medium Stress	High Stress
Max Tempt.	35.910000	36.570000	36.590000
Max HR	163.500000	169.930000	180.230000
Max EDA	24.167313	43.563389	59.760712

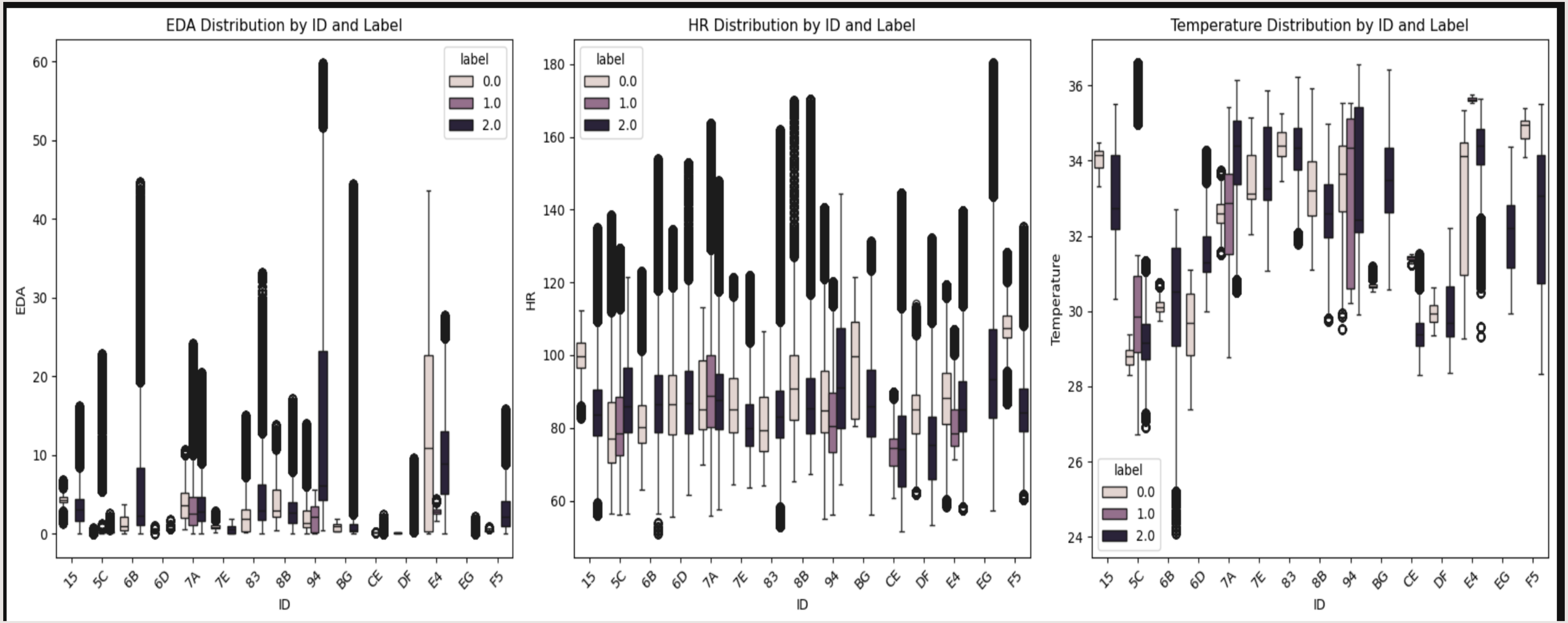
***Physiological Measures** (Max Temperature, Max HR, and Max EDA) across different stress levels (Low, Medium, High).*



Heart rate (HR) and **Electrodermal Activity (EDA)** have a positive correlation with increasing stress levels

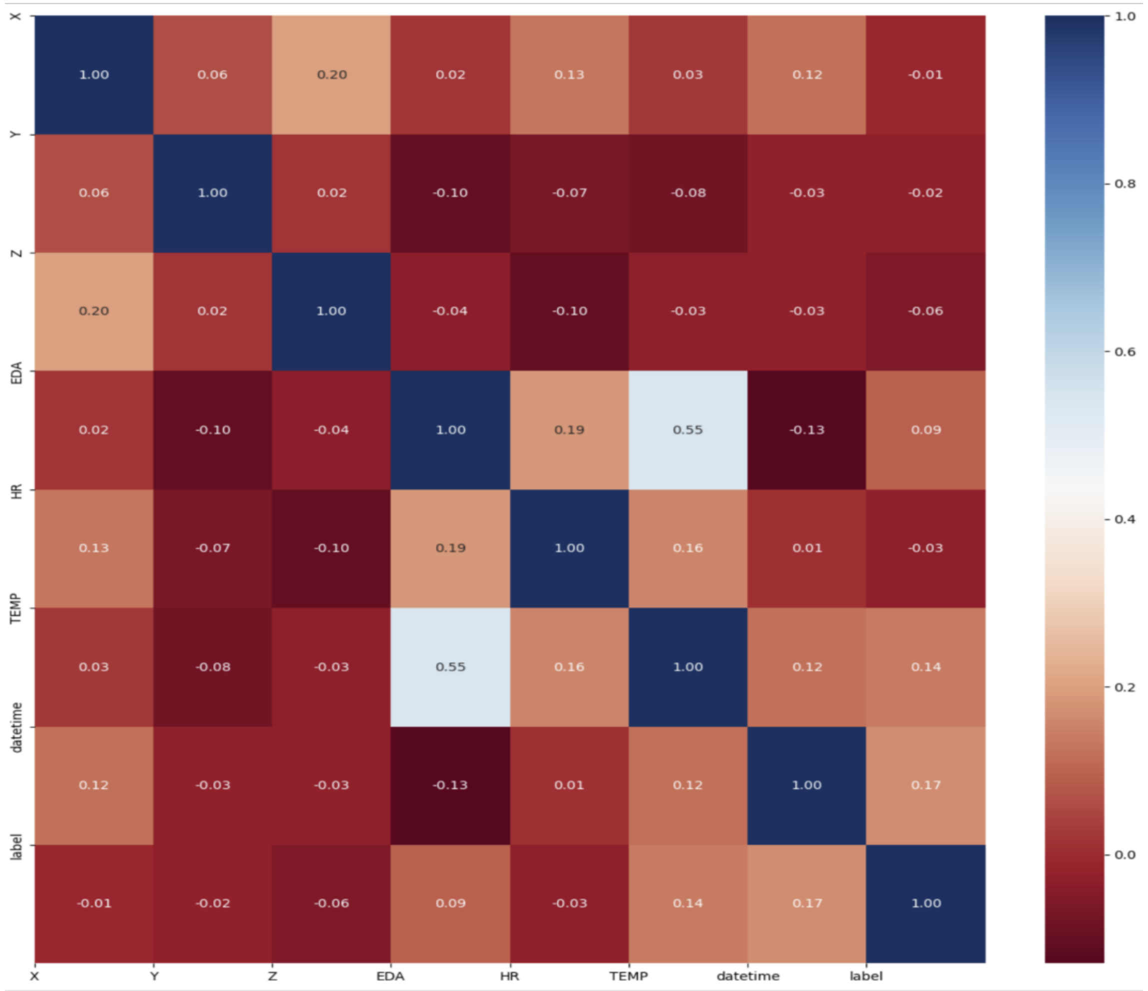
Temperature remains relatively stable across different stress levels

Variations in Physiological Measurements Across different ID’s and Stress Levels:

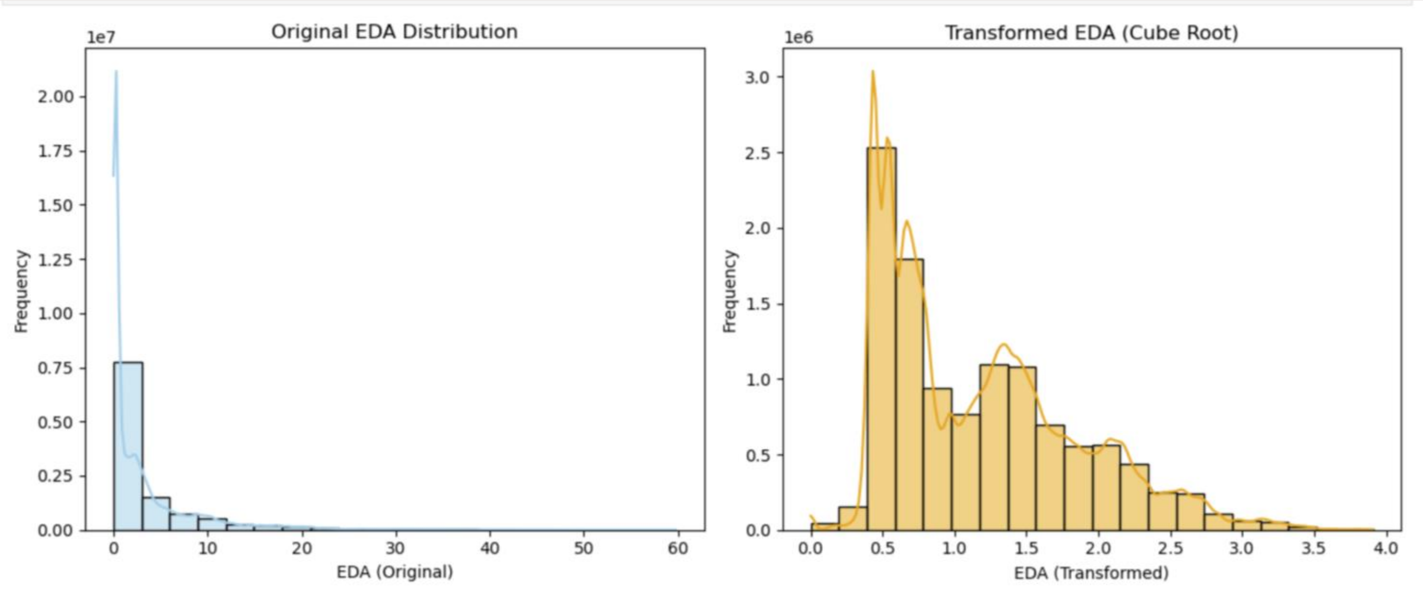


Correlation of Variables with Stress Levels:

TEMP and EDA shows the strongest correlation (0.55)



Distribution of EDA (Electrodermal Activity) *before and after Cube Root transformation*



Right Plot (Transformed EDA Distribution - Cube Root):

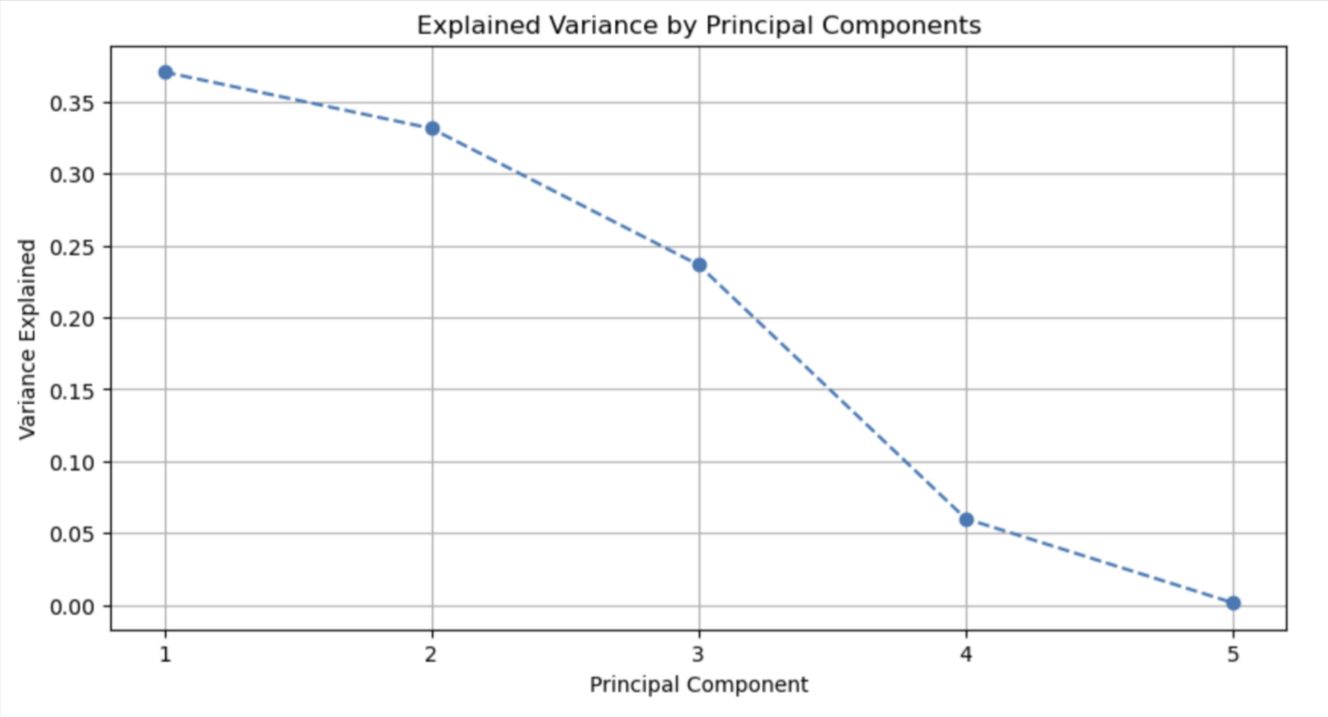
•This plot shows the distribution after applying a **cube root transformation** to the original EDA values. The cube root transformation is commonly used to reduce skewness in data with extreme values.

•After transformation, the distribution is more **spread out** and **less skewed**, with values more evenly distributed across the range. The KDE curve is smoother, indicating a more normalized shape.

The transformation successfully reduces the skewness of the original EDA data to make it less skewed.

•	EDA Original value	EDA Adjusted value (Cube Root Transformation)
•	Skewness: 3.014666	Skewness: 0.819291
•	Kurtosis: 12.053117	Kurtosis: -0.019175

Principal Component Analysis (Explained Variance):



- The first principal component explains around **35%** of the variance.
- The second component explains slightly less than **30%**.
- The subsequent components explain progressively smaller portions of the variance.
- Indicating that 3 components may be sufficient to capture most of the important variance in the data.

Model Training: XGBClassifier, LGBM, Tensor Flow:

XGBClassifier

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
```

Model: "sequential_1"

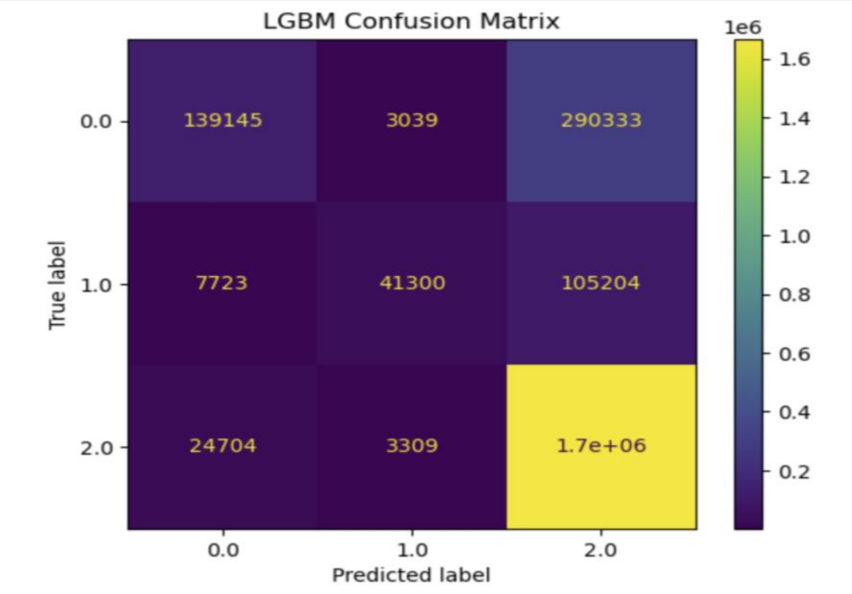
Layer (type)	Output Shape	Param #
dense_4 (Dense)	(None, 64)	512
dense_5 (Dense)	(None, 32)	2,080
dense_6 (Dense)	(None, 1)	33

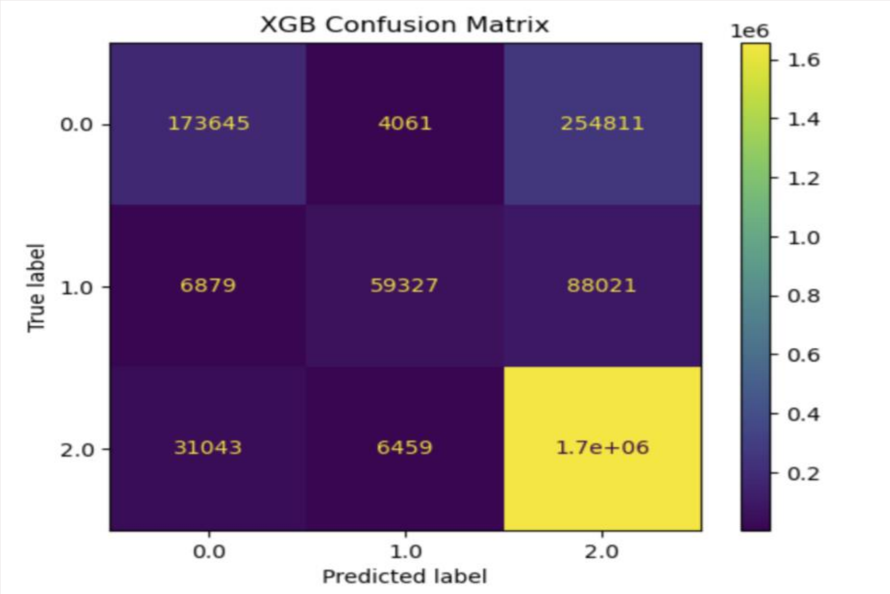
Total params: 2,625 (10.25 KB)
Trainable params: 2,625 (10.25 KB)
Non-trainable params: 0 (0.00 B)

LGBMClassifier

```
LGBMClassifier(objective='multiclass')
```

Model Performance:





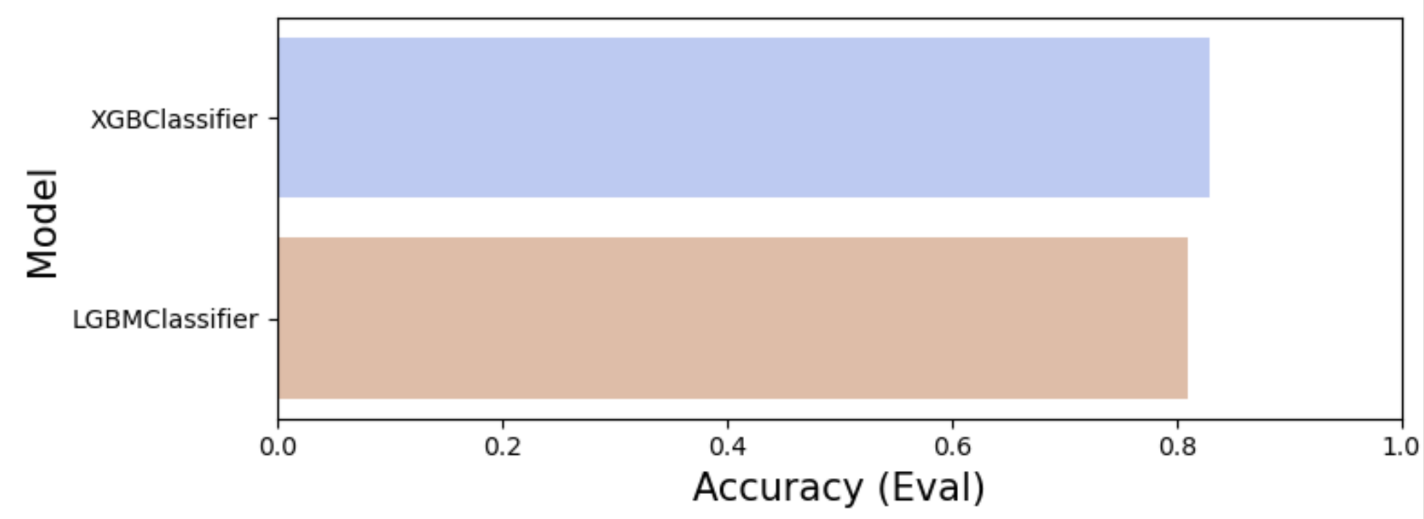
Accuracy: 80.957%
This indicates that approximately **81%** of the total predictions made by the model are correct.

Accuracy: 82.844%
This indicates that approximately **82.8%** of the total predictions made by the model are correct.

	Predicted: Class 0	Predicted: Class 1	Predicted: Class 2
Actual: Class 0	139145	3039	290333
Actual: Class 1	7723	41300	105204
Actual: Class 2	24704	3309	1665932

	Predicted: Class 0	Predicted: Class 1	Predicted: Class 2
Actual: Class 0	173645	4061	254811
Actual: Class 1	6879	59327	88021
Actual: Class 2	31043	6459	1656443

Model Accuracy:



Model	Accuracy
LGBMClassifier	0.80957
XGBClassifier	0.82844

XGBoost handles the class imbalance better through built-in options like **scale_pos_weight** and better sensitivity to class distribution, resulting in improved accuracy.

LightGBM may struggle slightly with imbalanced classes unless specifically configured (e.g., using is_unbalanced=True or scale_pos_weight).

Results

This analysis demonstrates that machine learning models, particularly XGBoost and LightGBM, can effectively predict stress levels based on physiological data collected from wearable technology. While XGBoost showed better handling of class imbalance, LightGBM can be optimized to achieve similar results.

Conclusion:

Proactive Stress Management: By monitoring stress levels continuously and intervene early, before stress escalates into burnout or mental health issues.

Tailored Interventions: Design personalized stress-relief interventions, like short breaks or mindfulness exercises, based on detected stress patterns.

Task Redistribution: The model could help redistribute tasks to professionals experiencing lower stress levels, reducing the risk of errors caused by stress in high-stakes situations like surgeries or emergency care

Chronic Stress Patterns: Over time, the model could reveal chronic stress patterns in certain individuals, helping identify professionals who may be at higher risk of long-term issues like burnout, depression, or anxiety.

Employee Well-Being Programs: The data collected can be used to justify the introduction of well-being programs (e.g., mental health days, meditation sessions, gym memberships) for high-risk groups based on quantifiable stress data.

Recommendations:

Continuous Improvement and Learning

Model Evolution: As more data is collected over time, models can be retrained to become more accurate and adaptable to new conditions. For example, the model could account for changing stress responses due to factors like COVID-19 or seasonal workloads.

User Feedback Loop: Integrating the model with healthcare professionals' feedback (e.g., subjective stress reports) can improve the model's sensitivity and allow it to learn from individual differences in stress responses.

Recommendation: Track long-term stress data to identify trends and patterns that could lead to chronic stress or burnout.

Action: Use the predictive capabilities of stress detection models to analyze historical stress data and forecast periods of high stress. This information can guide resource allocation, staffing, and preventive care.

More computing power is required to handle such big data.

Thank You ☺