Stress Detection Using Machine Learning

Ayesha Nusrat
Department of Statistics and O.R.
Aligarh Muslim University
Aligarh, U.P
ayesha1247801@gmail.com

***Abstract*—Stress is a major factor affecting mental and physical health, impacting productivity and overall well-being. Traditional stress assessment methods rely on self-reports and clinical evaluations, which can be subjective and time-consuming. With advancements in machine learning, automated stress detection using physiological data has gained importance.**

This study aims to classify stress levels using biometric features such as heart rate, respiration rate, snoring rate, body temperature, blood oxygen levels, eye movement, limb movement, and sleep duration. The dataset was sourced from Kaggle and included balanced stress levels ranging from 0 (No Stress) to 4 (High Stress). Data augmentation was applied to increase dataset size but was not needed for class balancing.

We implemented and evaluated six machine learning models: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, Naive Bayes, Random Forest, and CatBoost. Each model was trained and tested to determine its effectiveness in stress detection.

Data preprocessing included handling missing values, managing outliers, correlation analysis, and computing descriptive statistics. The models were evaluated using accuracy, precision, recall, and F1-score. Results were summarized and visualized for easy comparison.

Findings showed that KNN achieved the highest accuracy, followed by Random Forest and CatBoost. Traditional models like Naive Bayes, Logistic Regression, and SVC performed slightly lower, indicating that tree-based and distance-based methods were more effective for this task.

The study demonstrates the potential of machine learning for real-time stress monitoring, paving the way for AI-driven mental health solutions and wearable health technologies.

I. Introduction

A. Background

Stress is a growing health concern that affects both mental and physical well-being. Chronic stress has been linked to serious health conditions such as heart disease, anxiety disorders, immune system dysfunction, and sleep disturbances. Traditional stress assessment methods, such as self-reported surveys and psychological questionnaires, rely on subjective responses, which can be inconsistent, biased, and impractical for real-time monitoring.

Advancements in machine learning and physiological monitoring now enable automated stress detection using biometric data collected from wearable sensors and medical devices. Physiological indicators such as heart rate, respiration rate, snoring rate, body temperature, blood oxygen levels, eye movement, limb movement, and sleeping patterns provide valuable insights into a person's stress level. Machine learning models can analyze these signals to detect and classify stress more objectively and accurately than traditional methods.

B. Problem Description

The goal of this project is to develop a machine learning-based stress detection system that predicts stress levels ranging from 0 (No Stress) to 4 (High Stress) based on biometric data. The dataset, obtained from Kaggle, consists of 630 observations and 9 features, including physiological indicators such as heart rate, respiration rate, snoring rate, body temperature, blood oxygen levels, eye movement, limb movement, and sleeping patterns. Various machine learning models will be implemented and evaluated to determine their effectiveness in stress detection.

C. Objective

The key objectives of this project are:

- Train SVC, KNN, Logistic Regression, Random Forest, Naive Bayes, and CatBoost model for fitting approach.
- Perform hyperparameter tuning to optimize selected models.
- Apply cross-validation for a generalized model evaluation.
- Conduct data analysis including missing value handling, outlier management, correlation analysis, and descriptive statistics.
- Summarize and visualize final results for easy comparison.

D. Layout of Rest of the Document

The rest of the document is structured as follows:

- **Section II: Literature Review about Existing System** – Discusses traditional and modern approaches to stress detection, their limitations, and the role of machine learning.
- **Section III: Methodology** – Details the dataset, preprocessing techniques, applied machine learning models, evaluation metrics, and hyperparameter tuning.
- **Section IV: Results and Discussion** – Presents the outcomes of the trained models, performance comparisons, and insights from the analysis.
- **Section V: Conclusion** – Summarizes the key findings, contributions, limitations, and potential future improvements.
- **References** – Lists all sources cited in the report.

II. LITERATURE REVIEW ABOUT EXISTING SYSTEM

Stress detection has been an area of interest in healthcare and psychology for decades. Traditional methods primarily relied on self-reported questionnaires and clinical assessments, which, although useful, were often subjective and lacked real-time applicability. Some of the existing systems for stress detection include:

- **Physiological Signal-based Systems**: These methods utilize sensors to measure heart rate, blood pressure, skin conductance, and other biometric parameters. While accurate, they require specialized hardware and can be intrusive.
- **Behavioral Analysis Methods**: These include monitoring facial expressions, speech patterns, and physical activity. Although non-intrusive, they are susceptible to external influences and may not always provide reliable stress indicators.
- **Wearable Devices and Mobile Applications**: The advancement in technology has led to the use of smartwatches and mobile apps to track physiological changes over time. These systems provide continuous monitoring but can suffer from data inaccuracies and user compliance issues.
- **Machine Learning and AI-based Approaches**: Recent studies focus on using machine learning models to analyze large datasets comprising physiological and behavioral markers. These methods improve accuracy and automation but require extensive data collection and computational resources.

A. Limitations of Existing Systems

- Many traditional methods are subjective and rely on self-reporting.
- Physiological monitoring requires specialized equipment, making it less accessible.
- Behavioral analysis methods can be inconsistent due to external factors.
- Machine learning models require large amounts of high-quality data for accurate predictions.

III. METHODOLOGY

A. Dataset Description

The dataset utilized in this study was obtained from **Kaggle** and comprises **630 observations** with multiple biometric features. It is specifically designed for **stress level classification**, where the target variable represents **five distinct stress levels**:

- **0** – No Stress
- **1** – Low Stress
- **2** – Moderate Stress
- **3** – High Stress
- **4** – Severe Stress

The dataset is **balanced**, ensuring that each stress level has a proportionate number of samples, reducing the risk of model bias toward any particular class.

The features used for prediction include various **physiological and behavioral metrics**, such as:

- **Heart Rate (bpm)** – Measures cardiovascular activity under stress.
- **Limb Movement** – Indicates restlessness, which can be associated with stress.
- **Respiration Rate (breaths/min)** – Affected by stress-induced changes in breathing patterns.
- **Body Temperature (°F)** – Can fluctuate due to stress responses.
- **Eye Movement** – Tracks focus and stress-related eye behavior.
- **Blood Oxygen** – Changes under stress due to altered breathing patterns.
- **Snoring Rate** – Can indicate irregular sleep patterns often linked to stress.
- **Sleeping Hours** – Sleep deprivation is a key indicator of stress.

To improve model performance, **data augmentation** was applied **only to the training set** by introducing **small noise and scaling variations to numerical features**. However, augmentation was performed **only to increase the size of the data set** and was **not required for class balancing**, since the data set already contained an equal distribution of stress levels.

This well-balanced dataset enables **fair evaluation** of machine learning models, ensuring that no single stress level is underrepresented.

B. Data Preprocessing

1) Handling Missing Values: Handling missing values is crucial to ensure data consistency and prevent bias in machine learning models. In this dataset, missing values were checked using exploratory data analysis (EDA). Since the data set contains biometric characteristics, **the median was chosen as the imputation method for the numerical attributes**. This is because the median is less sensitive to extreme values compared to the mean, making it a robust choice for handling missing data.

2) Managing Outliers: Outliers can significantly affect model performance by skewing predictions. The outliers were first checked and analyzed through a boxplot and conclusion was made that since physiological features usually vary, these outliers play a vital role for the model; therefore, they were kept except for some features like body temperature, eye movement, blood oxygen, and heart rate because they have some unrealistic values. To manage these outliers, the **Interquartile Range (IQR)** method was used:

- The first quartile ($Q1$) and the third quartile ($Q3$) were calculated for each numeric feature.
- The interquartile range was calculated as $IQR = Q3 - Q1$.
- A threshold was set: Any data point outside the range $[Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR]$ was considered an outlier.

- Outliers were replaced with the median value of the respective feature.

This approach ensures that extreme values do not disproportionately influence model predictions while maintaining data consistency.

C. Feature Analysis

1) Descriptive Statistics: Descriptive statistics help to understand the distribution and central tendencies of each feature in the data set. The following statistical measures were computed:

- **Mean:** Provides the average value of each feature.
- **Median:** Represents the middle value, useful for skewed distributions.
- **Standard Deviation:** Indicates the spread of values around the mean.
- **Quartiles:** Divide a dataset into four equal parts, with Q1 (25th percentile), Q2 (median, 50th percentile), and Q3 (75th percentile) marking the values below which 25%, 50%, and 75% of the data fall, respectively.
- **Minimum and Maximum Values:** Show the range of each feature.

This analysis provided insights into potential data imbalances, deviations, and general trends in the dataset.

2) Correlation Analysis: Correlation analysis was performed to examine relationships between different features. A **correlation matrix** was generated using Pearson's correlation coefficient (ρ), which measures the strength and direction of relationships between numerical variables:

- $\rho > 0.7$ indicates a strong positive correlation.
- $\rho < -0.7$ indicates a strong negative correlation.
- $\rho \approx 0$ suggests no correlation.

D. Dataset Split

To train and evaluate the machine learning models, the dataset was split into training and testing sets:

- **75% of the data** was used for training.
- **25% of the data** was reserved for testing.

This split ensures that models generalize well on unseen data and do not overfit.

E. Data Augmentation

To improve model robustness and generalization, **data augmentation** was applied only to the training set. The augmentation techniques used were:

- **Jittering:** Small random **Gaussian noise** was added to numerical features to introduce slight variations in data values.
- **Scaling:** Each feature was multiplied by a random factor within a specified range to create slight distortions.

Important Considerations:

- Augmentation was applied **only to numerical features**, ensuring that categorical variables and labels remained unaffected.
- The process increased the dataset size, enhancing the model's ability to generalize across unseen variations.
- The augmented data was combined with the original training set to create a richer dataset for model training.

This method helped improve the model's robustness by introducing slight variations that mimic real-world biometric fluctuations.

F. Model Training and Evaluation

1) Model Fitting: To train and compare different machine learning models, a unified approach was adopted where all models were trained together in a single execution loop. The models used in this study include:

- **Support Vector Machine (SVM)**: A robust classifier that finds the optimal hyperplane to separate different stress levels.
- **k-Nearest Neighbors (KNN)**: A distance-based algorithm that classifies an instance based on its closest neighbors.
- **Logistic Regression**: A linear model commonly used for classification tasks.
- **Naive Bayes**: A probabilistic model based on Bayes' theorem, often effective for small datasets.
- **Random Forest**: An ensemble learning method using multiple decision trees to improve accuracy.
- **CatBoost**: A gradient boosting algorithm optimized for categorical data.

Each model was trained using the preprocessed dataset, and their performances were compared based on standard evaluation metrics.

2) Hyperparameter Tuning: To enhance model performance, hyperparameter tuning was conducted for selected models using:

- **Grid Search**: Exhaustively searches through a specified set of hyperparameters to find the best combination.
- **Random Search**: Randomly samples hyperparameters within a given range to efficiently find optimal values.

Hyperparameter tuning was applied particularly to models that benefit significantly from parameter optimization, such as SVM, Random Forest, and CatBoost. The best hyperparameter configurations were selected based on cross-validation performance.

3) Cross-Validation: To ensure that the models generalize well to unseen data, **5-fold cross-validation** was implemented. This method works as follows:

- The dataset is randomly divided into five equal-sized folds.
- Each model is trained on four folds and tested on the remaining fold.
- The process repeats five times, ensuring that each data point is used for both training and testing.
- The average performance across all folds is reported as the final evaluation metric.

This approach reduces overfitting and ensures that the results are not biased by a particular train-test split.

4) Performance Metrics: To assess model performance, several evaluation metrics were used:

- **Accuracy**: Measures the overall proportion of correctly classified stress levels.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- **Precision**: Indicates the proportion of correct positive predictions out of all predicted positives.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- **Recall (Sensitivity)**: Measures the model's ability to identify stressed individuals.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

- **F1-Score**: A harmonic mean of precision and recall, balancing false positives and false negatives.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

These metrics were chosen to provide a comprehensive evaluation of model performance, particularly in distinguishing between different stress levels.

IV. Results and Discussion

After training and evaluation, the models were systematically compared based on their overall performance across key metrics, including accuracy, precision, recall, F1-score. The results were visualized using bar charts and confusion matrices to highlight differences in classification performance. Additionally, a summary table was created for easy interpretation, presenting a side-by-side comparison of each model's strengths and weaknesses. This approach helped identify the most effective classifier for stress detection while considering trade-offs between precision and recall, particularly in distinguishing higher stress levels.

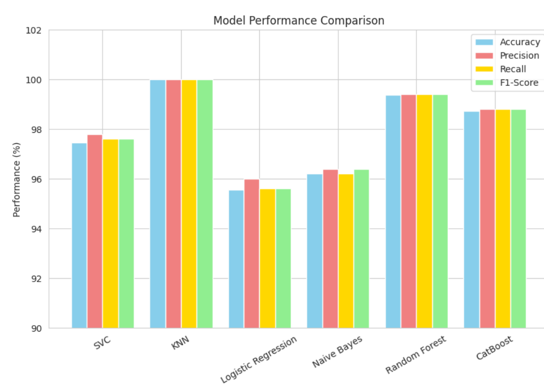
TABLE I
SUMMARY

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
KNN	100	100	100	100
Random Forest	99.37	99.4	99.4	99.4
Logistic Regression	95.57	96	95.6	95.6
SVC	97.47	97.8	97.6	97.8
Naive Bayes	96.2	96.4	96.2	96.4
CatBoost	98.73	98.8	98.8	98.8

A. Observations

The following key insights were derived from the model comparison:

- **KNN achieved the highest accuracy (100%)**, outperforming all other models. This suggests that the stress detection dataset may contain well-separated classes, making KNN's distance-based classification highly effective.
- **Random Forest and CatBoost closely followed, with accuracies of 99.37% and 98.73% respectively.** These models leverage ensemble learning, which combines multiple decision trees to reduce overfitting and improve robustness. The strong performance of these models confirms that stress levels can be effectively identified using tree-based methods.
- **Naive Bayes performed slightly lower than ensemble-based models (96.2%).** This could be due to the assumption of feature independence, which may not hold true for biometric data. Since stress-related features like heart rate, respiration rate, and limb movement are often correlated, Naive Bayes may struggle with capturing complex feature interactions.
- **Support Vector Machine (SVM) showed strong performance (97.47%) but was slightly behind tree-based models.** SVM works well in high-dimensional spaces, but its performance can be affected by feature scaling and parameter selection.
- **Logistic Regression achieved the lowest accuracy (95.57%) among the models.** Since logistic regression is a linear model, it might struggle with capturing the non-linear relationships between stress and biometric features. This suggests that more complex models are better suited for stress classification.
- **Overall, ensemble learning and distance-based methods performed best.** KNN, Random Forest, and CatBoost outperformed traditional models like Logistic Regression, Naive Bayes, and SVM. The results indicate that for biometric-based stress classification, models that can handle complex, non-linear relationships and interactions among features tend to yield superior performance.



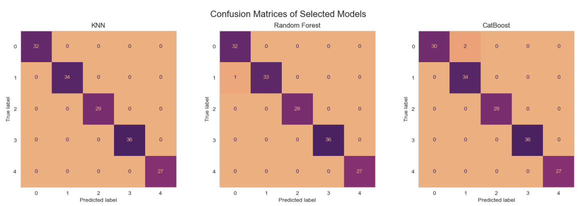
- **Trade-offs between models should be considered based on application needs.** While KNN achieved perfect accuracy, it can be computationally expensive for large datasets. Random Forest and CatBoost offer a balance between accuracy and efficiency, making them practical choices for real-time applications.

The results highlight the importance of selecting appropriate models based on dataset characteristics and the intended application of stress detection systems.

V. Conclusion

This study explored biometric-based stress detection using machine learning models. Through comprehensive data analysis, preprocessing, and model evaluation, valuable insights were obtained regarding the effectiveness of various classification techniques. The key findings of the study are summarized as follows:

- Data preprocessing steps such as **handling missing values, outlier management, and correlation analysis** improved data quality, ensuring reliable model performance.
- **KNN achieved the highest accuracy (100%)**, followed by **Random Forest (99.37%)** and **CatBoost (98.73%)**, indicating that non-linear models and ensemble methods are highly effective for stress detection.



- The use of **cross-validation** ensured that model performance was consistent across different data splits, preventing overfitting and improving generalization.

TABLE II
Model Accuracy Comparison

Model	Train Accuracy (%)	Test Accuracy (%)	Cross Validation (%)
KNN	99.36	100.00	99.36
Random Forest	99.47	99.37	99.36
Logistic Regression	97.88	95.57	96.50
SVC	99.36	97.47	98.83
Naive Bayes	97.88	96.20	96.08
CatBoost	99.47	98.73	99.47

A. Limitations of the Study

Although the study produced promising results, several limitations must be acknowledged:

- **Dataset Dependency:** The dataset used in this research was sourced from Kaggle, and its applicability to broader, real-world populations remains uncertain. Differences in data collection conditions could impact model generalization.
- **Synthetic Data Augmentation:** Data augmentation was performed using synthetic noise addition and scaling, which may not fully capture real-world variations in biometric stress responses.
- **Feature Constraints:** The dataset primarily consists of biometric signals. Including additional behavioral and psychological factors might enhance stress prediction accuracy.

B. Future Scope

Future work can build upon this study by incorporating the following advancements:

- **Expanding Dataset:** Collecting and integrating more diverse real-world biometric data will enhance model robustness and improve generalization to varied demographics.
- **Real-Time Stress Monitoring:** Developing a real-time stress detection system that integrates with wearable devices and mobile applications can provide continuous stress assessment for practical applications.
- **Exploring Deep Learning Models:** Future studies could investigate deep learning architectures such as LSTMs and CNNs to better capture temporal and spatial dependencies in biometric data.
- **Multimodal Stress Detection:** Combining biometric data with textual, audio, or behavioral cues (e.g., speech patterns and facial expressions) may improve stress detection accuracy.

VI. Reference

- Kaggle, "Stress Detection Dataset," Available at: https://www.kaggle.com/datasets/dheerov/stress-detection-dataset.
- "Human Stress Prediction," GitHub Repository. Available at: https://github.com/ishreya09/Human-Stress-Prediction
- S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *Journal of Health and Social Behavior*,
- H. J. Baek, J. S. Kim, K. K. Kim, and K. S. Park, "Enhancing the usability of wearable devices for stress level measurement," *Journal of Medical Systems*, vol. 39, no. 12, p. 179, 2015.
- S. K. Roy, A. A. Asif, F. Ahmed, A. K. Bashar, A. M. Rahman, and M. S. Kaiser. (2020). *Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data*. Available at: https://www.researchgate.net/publication/344052913_Stress_Detection_with_Machine_Learning_and_Deep_Learning_using_Multimodal_Physiological_Data
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. (Standard textbook for understanding ML models like SVM, KNN, etc.)
- Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5–32. (Fundamental paper on the Random Forest algorithm)
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). *CatBoost: Unbiased Boosting with Categorical Features*. In *Advances in Neural Information Processing Systems (NeurIPS)*. (Original paper introducing CatBoost)
- Scikit-learn Developers, "Scikit-learn documentation," https://scikit-learn.org/stable/documentation.html, Accessed: May 1, 2025.

Appendix: Code Repository

* The complete source code for this project is available at the following GitHub repository:

https://github.com/a1s8/Stress-Detection