The Fusion of Mind and Motion: Classifying Activity Levels with Biometric and Psychological Insights

MSc Research Project
Data Analytics

Ajinkya Gajananrao Chintawar
Student ID: 23113561

School of Computing
National College of Ireland

Supervisor: Dr William Clifford

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ajinkya Gajananrao Chintawar
Student ID:	23113561
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Dr William Clifford
Submission Due Date:	29/01/2025
Project Title:	The Fusion of Mind and Motion: Classifying Activity Levels with Biometric and Psychological Insights
Word Count:	6495
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Ajinkya Gajananrao Chintawar
Date:	28th January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

The Fusion of Mind and Motion: Classifying Activity Levels with Biometric and Psychological Insights

Ajinkya Gajananrao Chintawar
23113561

Abstract

This study explores the application of machine learning models for classifying physical activities using biometric measures and personality traits. A dataset from Figshare, integrated with survey responses, was analyzed using two algorithms: Gradient Boosting and a Tuned Stacking Classifier. Gradient Boosting achieved the highest accuracy at 73%, while the Stacking Classifier demonstrated balanced performance across all metrics, achieving 72% and effectively addressing class imbalances. The study highlights the models' real-world applicability through validation and visualization using a Tableau dashboard. Key predictors included Resting Heart Rate, Honesty-Humility personality trait, and BMI. Challenges such as class imbalance were addressed using SMOTE and imputation techniques. The findings emphasize the potential of machine learning in personalized health monitoring and promoting active lifestyles via wearable devices.

Keywords: Physical Activity Classification, Machine Learning, Biometric Data, Gradient Boosting, Personalized Health Interventions

1 Introduction

1.1 Background

Wearable technology and machine learning have made some leaps forward in monitoring human behavior and health outcomes. Wearable sensors have proved their ability to capture and classify the level of physical activity by biometric signals such as heart rate, electrodermal activity, and movement patterns Mannini and Sabatini (2010); Piciucco et al. (2021). Simultaneously, investigations in personality-related research, for instance, on the Big Five model, have been shown to have significant associations in tendencies toward behaviors for which critical insight into health, productivity, and the civil good is important Rochin Demong et al. (2023); Shaposhnyk et al. (2023). Studies such as Bianco and Napoletano (2019); Alsareii et al. (2022) highlighted how biometric signals fit into the person's traits, especially that part relates to effectively classifying physical activity based on input from biometrics, which is still not fully explored.

The existing literature has established the potential to combine physiological and psychological data to improve healthcare interventions. For example, biometric signals can be combined with contextual and demographic data to enhance machine-learning models that recognize emotions and predict activity patterns, as shown by Sánchez-Reolid et al. (2022); Saganowski et al. (2022). Despite these recent developments, research is still confined to controlled environments, leaving a significant gap in applying these techniques to dynamic datasets in the real world. This gap, once addressed, paves the way for more personalized, adaptive, and accurate systems for health monitoring and behavioral analysis.

1.2 Importance

Gaining insight into the correlation between these biometric data and personality traits could lead to a significant leap in predictive capabilities with machine learning models. This would enable a much more holistic approach to activity classification. This integration will provide insight into how individual traits influence physical behavior, providing actionable information for healthcare professionals, fitness enthusiasts, and behavioral scientists. This research is also part of the more significant efforts worldwide in using wearable technologies and machine learning to promote healthy lifestyle choices for better mental and physical well-being Brons et al. (2024); Stockwell et al. (2021).

1.3 Research Question

How can machine learning models use data from biometrics and personality traits to classify physical activity levels effectively and show the relationship between biometrics and personality traits?

1.4 Objectives

The following objectives are outlined to answer the research question:

1. Investigate machine learning techniques for combining biometric and personality traits data.

2. Design a generic framework that effectively integrates biometric signals with personality traits for activity classification.

3. Implementation of models to analyze the relationship between the results of biometric data and personality traits to identify the most powerful predictors.

4. Evaluate the proposed models' accuracy, reliability, and generalizability using diverse real-world datasets.

1.5 Contribution

This research presents novelty by developing a framework using an intersection of biometric and personality data in novel physical activity classification. Thus, the study hopes to help bridge the gap by demonstrating the synergistic effects of collaboration of physiological and psychological factors toward personal and effective health monitoring systems Fiedler et al. (2020); Seol et al. (2024).

1.6 Dissertation Structure

This dissertation is organized into the following sections:

1. **Introduction**: Provides an overview of the research problem, objectives, and significance of the study, alongside a summary of contributions made.

2. **Related Work**: Discusses the current state-of-the-art literature in physical activity classification using biometric and personality data, identifying research gaps this study aims to address.

3. **Methodology**: Outlines the research framework, data collection, preprocessing, and machine learning techniques for classifying physical activity levels.

4. **Design Specification**: Details the system architecture, algorithms, and tools for building the solution pipeline.

5. **Implementation**: Describes the process of implementing the proposed solution, from exploratory data analysis to developing and deploying machine learning models.

6. **Evaluation**: Presents the results and performance metrics of the models, including experiments conducted and insights derived.

7. **Conclusion and Future Work**: Summarizes the findings, discusses limitations, and proposes directions for future research and practical applications.

This structure ensures a logical flow of ideas and methods, providing clarity and cohesion throughout the dissertation.

2 Related Work

Integrating biometric data and personality traits in machine learning models to classify physical activity levels and explore their relationships has been a growing area of research. This literature review critically evaluates the key studies that have contributed to this field, highlighting their methodologies, findings, and limitations. The review identifies the existing research gaps and justifies the need for the proposed study.

2.1 Biometric Data and Physical Activity Classification

Wearable technologies have done a lot in modern times to monitor physical activity levels further. Mannini and Sabatini (2010); Harmouche-Karaki et al. (2023) demonstrated the machine learning classification applied on-body accelerometers with support vector machines that attained high accuracy, but this study had several limitations due to being done in controlled conditions.

Dinh et al. (2019); Butte et al. (2012) extended machine learning for health monitoring by integrating multiple biometric signals to predict diabetes and cardiovascular disease. This work showed promising fusion results from various physiological signals but did not focus on physical activity classification. On the other hand, Sánchez-Reolid et al. (2022)

used CNNs to classify arousal levels based on electrodermal activity with a high F1-Score. However, their work did not cover personality traits as predictors, leaving a significant gap.

Stockwell et al. (2021) estimated the influence of the COVID-19 lockdown on physically active measures by self-report and device-based measures. Their findings showed significant declines in activity, which calls for adaptive, personalized interventions, which this study aims to address.

2.2 Personality Traits in Health and Behavior Analysis

Personality traits have increasingly been recognized for influencing behavior and health outcomes. Rochin Demong et al. (2023) classified students' social well-being using machine learning algorithms based on personality traits, showing a strong predictive relationship. On the other hand, Shaposhnyk et al. (2023) investigated the relationship between cognitive load and personality traits, providing a basis for exploring such relationships in physical activity contexts.

Garbarino et al. (2014); Cristi-Montero (2018); Islam (2024) identified wearables as real-time feedback devices in behavioral modification. Their work points to a promising avenue of merging physiological and psychological data, even though they did not relate personality traits to activity classification.

2.3 Integration of Biometric and Personality Data

The integration of biometric and personality traits data for activity classification remains underexplored. Ahmadi et al. (2020) conducted the exercise of integrating the group and fully personalized models in human activity classification for children who had cerebral palsy. Their findings emphasize the importance of a tailored approach but do not generalize to broader populations and lack personality traits.

Fiedler et al. (2020) systematically reviewed eHealth interventions focusing on integrating biometric data, pointing to the potential for mHealth solutions. However, an overall lack of personality traits within their study suggests more significant opportunities for further research into the development of additional studies.

2.4 Strengths and Limitations of Reviewed Studies

The literature has pointed out several strengths, including high accuracy achieved in controlled environments (Mannini and Sabatini; 2010; Sánchez-Reolid et al.; 2022), the feasibility of integrating biometric signals for predictive modeling (Dinh et al.; 2019), and the significant role personality traits play in behavior prediction (Rochin Demong et al.; 2023). However, notable limitations exist, such as generalizability being restricted because of controlled settings, biometric data being integrated with personality traits in a limited way, and the lack of focus on real-world applications and adaptive approaches.

2.5 Gaps and Justification for Research

Despite advancements, none of the available studies have fully integrated biometric and personality data toward activity classification. A literature review shows a gap in understanding how such factors interact and their collective impact on behavior. The present study attempts to fill these gaps by developing machine learning models that combine biometric signals and personality traits to classify physical activity levels and explore their relationships.

2.6 Conclusion

In conclusion, while considerable research has been done on biometric signal analyses and personality traits, their application to the classification of physical activities remains unexploited. Therefore, this paper will bridge this critical gap, extending our knowledge about where these domains overlap and providing actionable insights for personalized health monitoring.

3 Research Methodology

The present research has adopted the CRISP-DM framework, a Cross-Industry Standard Process for Data Mining, to ensure a structured, repeatable, and verifiable investigation. Each step should consider the six phases of CRISP-DM to agree with the research objectives.

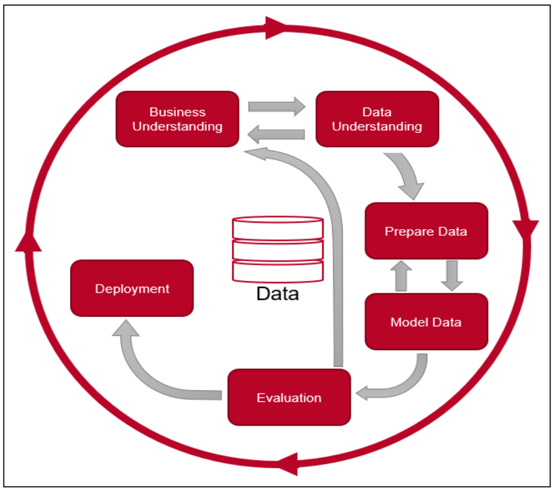

Figure 1: CRISP-DM Methodology

3.1 Business Understanding

The main aim of this study is to classify the level of physical activity using machine learning models and find the relationship between biometric data and personality traits. This study aims to provide insights for tailored health interventions, encourage active lifestyles, and improve well-being Brons et al. (2024); Fiedler et al. (2020).

Addressing gaps in previous studies, this study will try to show the practical utility of combining personality traits with biometric indicators in predictive models.

3.2 Data Understanding

Ensuring data quality and suitability includes cleaning, addressing missing values and inconsistencies, and removing duplicates. Feature engineering involves creating new features, such as categorizing physical activity levels from raw biometric data. The final training and testing datasets were split exclusively from the Figshare dataset for model performance evaluation.

3.3 Data Preparation

Several steps were implemented to ensure data quality and suitability for modeling. To maintain data integrity, missing values, inconsistencies, and duplicates were used to evaluate model performance and effectively address it during data cleaning. Feature engineering involves creating new features based on raw biometric data, such as evaluating model performance effectively; the training and testing datasets were split exclusively from the Figshare dataset.

3.4 Modeling

Different machine-learning algorithms were utilized to classify physical activity and explore the relationship between biometric data and personality traits. The choice of algorithms was guided by their efficacy in handling structured datasets and finding complex patterns in multimodal data.

Algorithms Used:

1. **Logistic Regression:** Known for its simplicity and interpretability, this algorithm provides a baseline for comparison in classification tasks.
2. **K-Nearest Neighbors (KNN):** Effective for smaller datasets, KNN captures local relationships between data points, aiding in classifying similar activities based on biometric and personality features.
3. **Decision Tree:** This algorithm was chosen for its intuitive structure and ability to capture non-linear relationships.
4. **Support Vector Machine (SVM):** Utilized to maximize the separation between classes, beneficial for handling high-dimensional data like biometric metrics.
5. **Random Forest:** A robust ensemble method aggregating predictions from multiple decision trees reduces overfitting.
6. **Gradient Boosting:** Leveraging iterative learning, this model builds on weak learners to achieve high accuracy, particularly on complex datasets.
7. **AdaBoost:** By weighting data points and focusing on difficult-to-classify samples, AdaBoost improves classification performance.
8. **Stacking Classifier:** Combines predictions from multiple models through a meta-learner, enabling the integration of strengths from various algorithms.

Relevance of Algorithms: Shaposhnyk et al. (2023); Mesanza et al. (2020) have successfully used ensemble methods such as Random Forest and Stacking Classifier in analyzing cognitive load based on physiological and personality traits. Their findings have also underlined the appropriateness of algorithms such as Support Vector Machine and Gradient Boosting for capturing complex relationships in multimodal data, thus addressing the objectives of this study.

Hyperparameter Tuning: For each algorithm, where necessary, Shaposhnyk et al. (2023) recommend the HalvingGridSearchCV, which was used to optimize parameters to develop models that were customized to capture the specific characteristics of the dataset.

Cross-Validation: To make the model more reliable, k-fold cross-validation was used. This prevents overfitting and ensures the models generalize well on new, unseen data, following best practices identified by Shaposhnyk et al. (2023).

3.5 Evaluation

Model performance was measured against accuracy, precision, recall, and F1-Score parameters. Rochin Demong et al. (2023) demonstrated the utility of such metrics in assessing the relevance of activity classification models. A relative analysis of models using these metrics was conducted to identify the best-performing classifier, following the benchmarking principles demonstrated by Stockwell et al. (2021).

3.6 Deployment

The deployment phase translated the research findings into practical applications. Further, interactive results visualization using Tableau was developed to highlight the relationships between physical activity level, personality traits, and demographic factors. The best-performing model was then deployed on survey data analysis to generate predictions and actionable insights for personalized health interventions.

Ethical Considerations: Data collection was anonymous, and data was gathered through Microsoft Forms; thus, all responses were non-identifiable. Participants were 18 years and above, and 40 individuals contributed to the dataset. These measures ensure adherence to ethical guidelines and the protection of privacy.

This methodology reflects the rigor of the CRISP-DM framework, contributing to the broader understanding of how biometric and personality data interact in activity classification models.

4 Design Specification

This section identifies and presents the techniques and architecture that underlie the implementation and its associated requirements. This study employs a modular approach integrating machine learning techniques for classifying physical activity levels within an end-to-end pipeline, from data collection to visualization.

4.1 System Architecture

The system architecture utilizes machine learning techniques for classifying physical activity levels, as illustrated in Figure 2. The process begins with data sources, including the Figshare dataset for training and testing and survey data collected via Microsoft Forms for validation and deployment. Data understanding involved statistical analysis and visualization using *Pandas*, *Matplotlib*, *Seaborn*, and *Plotly* libraries to uncover relationships and patterns. The preparation of the data involved dealing with missing values, feature engineering, and handling imbalanced classes using *SMOTE*, while preprocessing and pipeline setup were handled by *Scikit-Learn*.

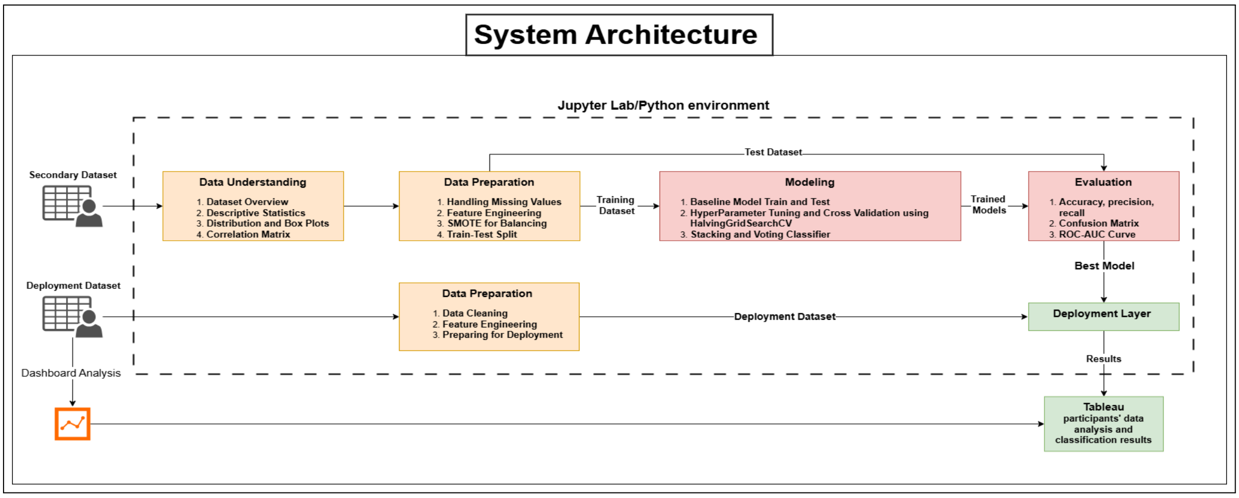

Figure 2: System Architecture

The modeling phase used machine learning techniques in the form of Logistic Regression, Random Forest, and Gradient Boosting, optimized through *HalvingGridSearchCV* with cross-validation using *Scikit-Learn* and *XGBoost*. Models were evaluated using accuracy, precision, recall, F1-score, confusion matrices, and ROC-AUC curves generated from *Matplotlib* and *Seaborn*. The best model was then applied to the survey dataset, saved using *Pickle File*, and loaded with *Joblib* library, after which Tableau was used in building an interactive dashboard for exploring results.

This modular design assures that the tasks are efficiently integrated and Python tools and compatibility with Tableau go smoothly to provide actionable insight.

4.2 Techniques and Algorithms

The different machine learning models used in the study are Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, and Stacking Classifier. For better accuracy, hyperparameter optimization was performed using *HalvingGridSearchCV* to ensure that classification performance is sound. SMOTE handled class imbalance, ensuring that training is well-balanced for better generalization. Ensemble methods used the Stacking Classifier to combine predictions for higher accuracy.

4.3 System Requirements

1. **Hardware Requirements:** A system with high enough processing power to train machine learning and deploy. Preferably, a GPU-enabled system should be used to handle large datasets.

2. **Software Requirements:** Python libraries such as *Scikit-Learn, Pandas, NumPy, Matplotlib, Plotly* for modeling; Tableau for visualization.

3. **Data Requirements:** Public datasets on physical activity and personality traits, complemented by a survey for the deployment and validation of the model.

This architecture ensures accurate classification of physical activity levels, offering actionable insights into participants' activity levels.

5 Implementation

This section describes the proposed implementation solution, focusing on the final stages of the development process. It also covers the output produced, such as transformed datasets, models developed, and tools used.

5.1 Dataset Description and Exploratory Data Analysis (EDA)

5.1.1 Dataset Description

The primary dataset from Figshare had 12 columns and 2,580 rows, including biometric measures and HEXACO personality traits. This dataset contained demographic and psychological variables such as sex, Honesty-Humility, and Extraversion, plus wearable-derived measures like steps from Fitbit, body mass index, and resting heart rate.

5.1.2 Exploratory Data Analysis (EDA)

A. Distribution & Box Plot of Fitbit Steps, BMI, and Resting Heart Rate:

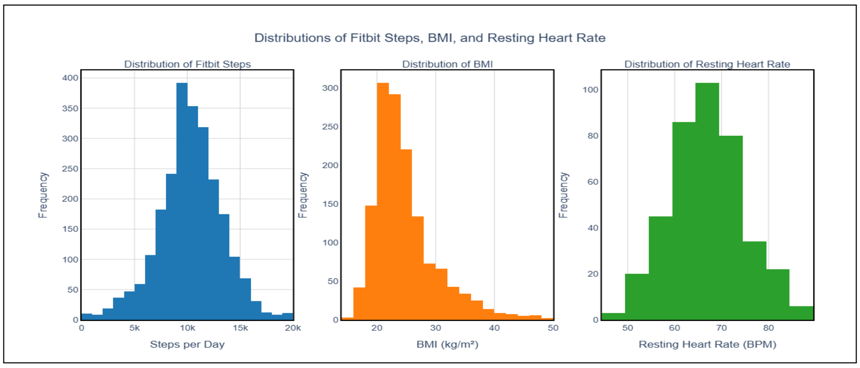

Figure 3: Distribution of Fitbit Steps, BMI, and Resting Heart Rate (HEXACO Dataset)

Figure 3 represents the Distribution of Fitbit steps, BMI, and resting heart rate. The Distribution of steps from Fitbit is approximately symmetric, around 10,000 steps/day, showing an active cohort in general but with variability. The Distribution of BMI is right-skewed, with most participants within a range of 20 to 30, although some outliers indicate obesity. The Distribution of resting heart rate is normal, centered at approximately 67 beats per minute, indicating the cardiovascular fitness level of the participants.

Figure 4 shows the box plots of the dispersion of values within the data sets, considering the outliers in BMI and resting heart rate. In the case of Fitbit steps, this range falls for most people between 8,000 and 12,000, while outliers below and above indicate persons who are either sedentary or highly active.

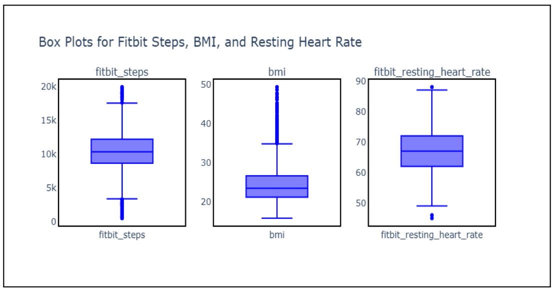

Figure 4: Box Plots for Fitbit Steps, BMI, and Resting Heart Rate (HEXACO Dataset)

B. Distribution & Box Plot of Personality Traits:

Figure 5 illustrates the distribution plots for personality traits, including personality traits such as Honesty/Humility and Extraversion: their Distribution comes out close as expected since most are 'very normally distributed,' and a wide range can be obtained inside the trends of participants expressing every side of traits, thus setting new grounds to understand why variability in physical activity was demonstrated after the tests.

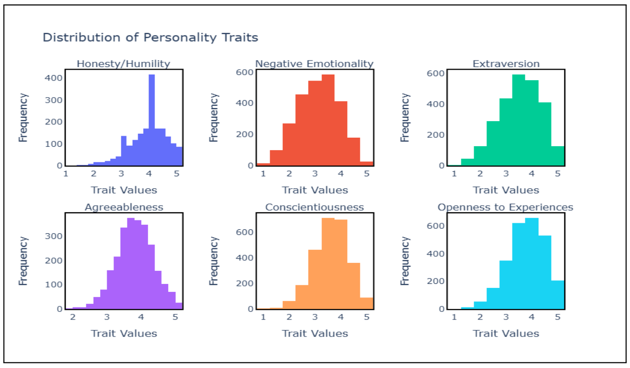

Figure 5: Distribution of Personality Traits (HEXACO Dataset)

Figure 6 shows that some sets have outliers, such as Honesty/Humility and Agreeableness, which may indicate atypical behavioral patterns for some participants.

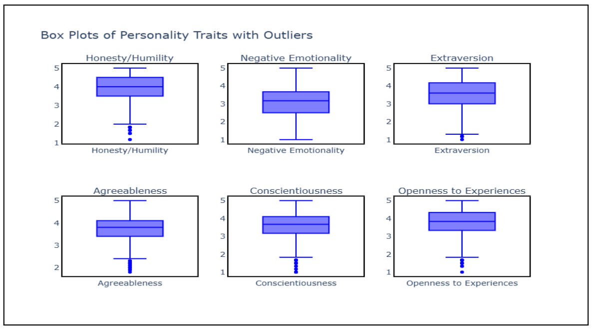

Figure 6: Box Plots of Personality Traits with Outliers (HEXACO Dataset)

C. Correlation Matrix:

Figure 7 illustrates the correlation matrix, showing the variables with each other. As will be seen, the correlations between personality traits and metrics of physical activity are very poor, probably because substantial amounts of data were missing from this data set. Hence, the imputation of further data or further analysis of these variables is warranted. For example, the correlation of lifestyle scores with the metrics of physical activity is moderate, which means that a good lifestyle is associated with better physical fitness. In contrast, personality traits correlate weakly with all the physical activity metrics, which may indicate complex relationships that require further investigation.

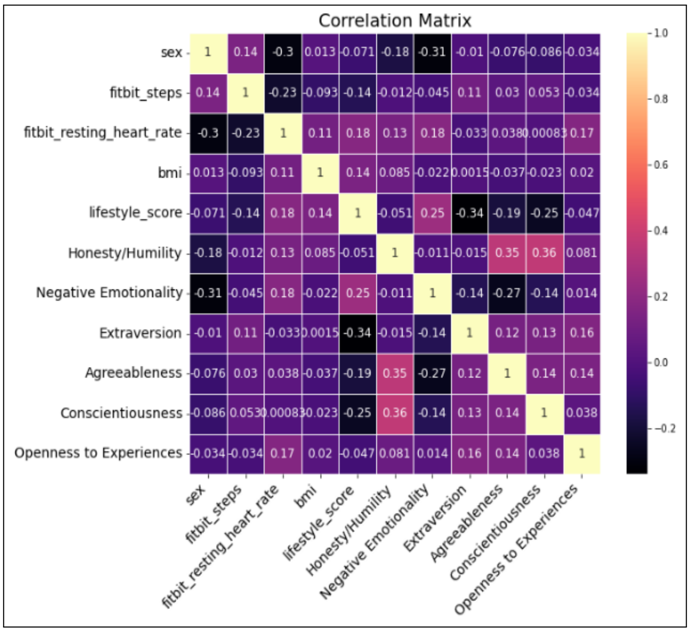

Figure 7: Correlation Matrix (HEXACO Dataset)

5.1.3 Conclusion of EDA

The EDA's conclusion revealed huge data variability in the process, where most of the characteristics of their biometric and personality traits included outliers. The process involved the identification of missing values in key columns such as 'bmi' and 'fitbit resting heart rate', to mention just a few, which indeed need consideration during data pre-processing. This study involves health-related data; hence, removing the outliers may lead to the loss of crucial information Gress et al. (2018); Kaur et al. (2023) , argued that this might be a risk or lead to losing important details. We will, therefore, retain the outliers so as not to affect the integrity of the data. With this, the next steps in Data Preprocessing will include handling missing values, removing features, feature engineering, splitting into train and test, and balancing data using SMOTE in preparation for modeling.

5.2 Data Preprocessing

5.2.1 Handling Missing Values

Iterative Imputer applied to impute missing values for columns like 'fitbit steps', 'fitbit resting heart rate', 'bmi', 'lifestyle score', and 'Honesty/Humility'. It kept the relationship between the variables and handled the missing values accordingly.

Figure 8 illustrates the After Imputation distribution. Iterative imputation has filled the missing values rather well, keeping the pattern of the original data. Comparing the Distribution before and after imputation, one can find that imputed distributions show a similar shape; hence, the underlying characteristics did not change much. This is one of the ways to reduce the impact of missing data on further modeling steps.

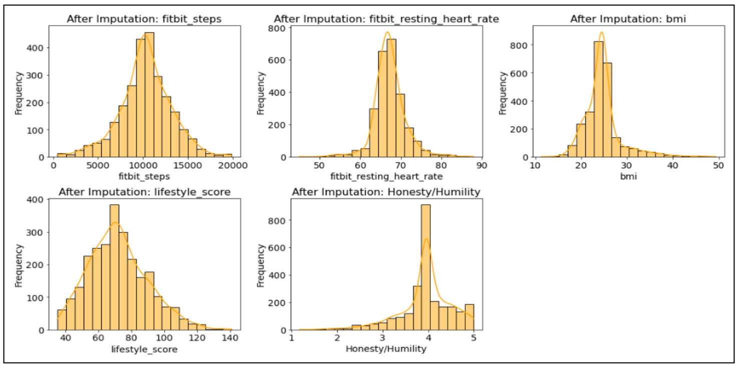

Figure 8: Distribution of Missing Values Columns After Imputation (HEXACO Dataset)

5.2.2 Feature Engineering

A new target variable, "activity level" is created based on Fitbit steps using thresholds from established health guidelines Pallavi Suyog Uttekar (2024):

- **Sedentary:** Less than 5,000 steps/day.
- **Moderately Active:** 5,000–9,999 steps/day.
- **Active:** Greater or Equal to 10,000 steps/day.

Figure 9 illustrates the Distribution of activity levels, showing class imbalance; therefore, SMOTE would have to be used to prepare the training set.

5.2.3 Train-Test Split

The proper work of model evaluation necessitated correctly splitting this dataset into training and test portions. Stratification was carried out so that each class associated with an activity level contributed to forming a test set on an equal basis. Exactly 10 percent of every class was set aside for the test set. Training set size (2,320 samples, 8 features) and test set size (260 samples, 8 features).

5.2.4 Data Balancing with SMOTE

After the initial train-test split, SMOTE was applied to the training dataset to handle class imbalance. Figure 10 below shows this:

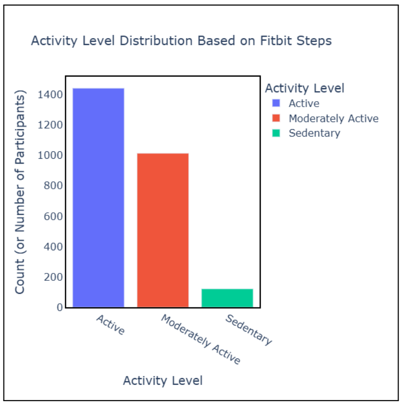

Figure 9: Activity Level Distribution (HEXACO Dataset)

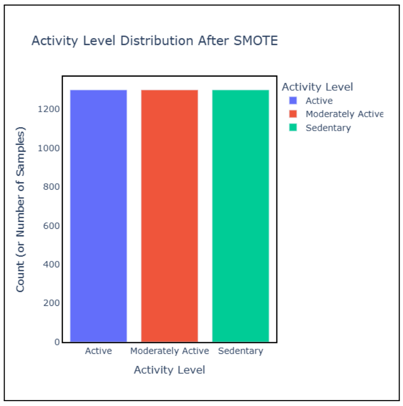

Figure 10: Activity Level after SMOTE (HEXACO Dataset)

After applying SMOTE, the training set was balanced to contain 3,897 samples and 8 features, while the test set remained unchanged with 260 samples and 8 features.

5.3 Modeling

The following steps involve the training and testing of baseline models, including Logistic Regression, KNN, Decision Trees, Random Forest, Gradient Boosting, SVM, and AdaBoost, on different metrics such as precision, recall, F1-score, and accuracy. Hyperparameter tuning has been performed with the help of HalvingGridSearchCV using 5-fold cross-validation to optimize performance in every model. In the case of ensemble methods, the Stacking Classifier combined multiple base models with meta-models like Gradient Boosting. This process was iterative to reach the best combination, balancing the accuracy-robustness trade-off.

5.4 Deployment

Participant survey data were pre-processed identically to the Figshare dataset, calculating BMI from height and weight and creating the Activity Level variable with the same thresholds. The best-performing Stacking Classifier was saved using the pickle module and deployed to predict the activity level using the survey data. Model accuracy has been rechecked to assess robustness.

Predicted activity levels are combined with the original survey data for complete analysis. An interactive Tableau dashboard was created for better usability, where stakeholders could investigate activity levels, personality traits, and biometric data..

This pipeline implementation underlines a structured approach to ensuring integrity in the data, robust modeling, and smooth deployment for actual applications of the insights derived from models.

6 Evaluation

This section is for the comprehensive evaluation of models and techniques. The analysis is structured, focusing on key experiments and case studies that address the research objectives. These findings are critically evaluated with rigorous statistical analysis and visualizations to assess their significance and implications. Each subsection explores some aspects of the experimental setup and its results.

6.1 Experiment 1: Baseline Classifier Results

The experiment aimed to investigate the results of individual classifiers and set a baseline upon which further improvements will be made. Below is the summary of key performance measures for each classifier as follows in Table 1:

Table 1: Baseline Models Results

Classifier	**Accuracy**	**Macro F1-Score**	**Weighted F1-Score**
Logistic Regression	0.42	0.35	0.46
K-Nearest Neighbors	0.54	0.48	0.56
Decision Tree	0.62	0.50	0.63
Random Forest	0.70	0.56	0.70
SVM	0.58	0.48	0.59
Gradient Boosting	0.67	0.57	0.68
AdaBoost	0.55	0.45	0.56

The results show that the algorithms of Random Forest and Gradient Boosting ensemble methods seem to run much more potent than any standalone classifier on their own and generalize well between all classes. The modeling of Logistic Regression and the AdaBoost requires substantive optimizations or might be unsuitable for this dataset in general. These findings add potential scope to the ensemble models that have been further explored in subsequent experiments.

6.2 Experiment 2: Model Training and Testing with Hyperparameter Tuning and Cross-Validation

This section presents the training and testing of models based on the best hyperparameters achieved from HalvingGridSearchCV using 5-fold cross-validation. It tries to analyze the contribution of hyperparameter tuning to improving model performance to establish the best-performing models.

6.2.1 Optimized Parameters

The hyperparameter tuning done using HalvingGridSearchCV with 5-fold cross-validation significantly enhanced the performance to match the dataset characteristics of the models best. Table 2 shows the best parameters for each model.

Table 2: Model and Best Parameters

Model	Best Parameters
Logistic Regression	C=0.1, max_iter=500, solver='lbfgs', class_weight="balanced"
K-Nearest Neighbors	metric='manhattan', n_neighbors=3, weights='distance'
Decision Tree	criterion='entropy', max_depth=20, min_samples_split=2
Random Forest	max_depth=20, min_samples_split=2, n_estimators=50, class_weight="balanced"
Gradient Boosting	learning_rate=0.2, max_depth=5, n_estimators=100
AdaBoost	learning_rate=0.1, n_estimators=200
Support Vector Classifier	C=10, gamma='auto', kernel='rbf', probability=True, class_weight="balanced"

6.2.2 Evaluation of Tuned Models

The tuned models were trained on the training set and tested on the test set. Table 3 summarizes the results of the best models:

Table 3: Evaluation of Tuned Models

Model	Accuracy	Macro Accuracy	Weighted Accuracy
Logistic Regression	0.60	0.50	0.62
K-Nearest Neighbors	0.60	0.50	0.61
Decision Tree	0.67	0.57	0.69
Random Forest	0.68	0.53	0.69
Gradient Boosting	0.73	0.61	0.74
AdaBoost	0.55	0.44	0.56
SVM	0.63	0.54	0.64

Gradient Boosting turned out to be the best model with an accuracy of 73%, along with the highest macro and weighted F1-scores. The accuracy of this performance underlined its robustness in dealing with class imbalances and classifying physical activity accordingly. In this regard, Random Forest and Decision Tree were comparably impressive, with accuracies of 68% and 67%, respectively, though they had slight weaknesses concerning the memorization of minority classes. On the other side, AdaBoost showed hardly any enhancement, underperforming at an accuracy of 55% even after tuning.

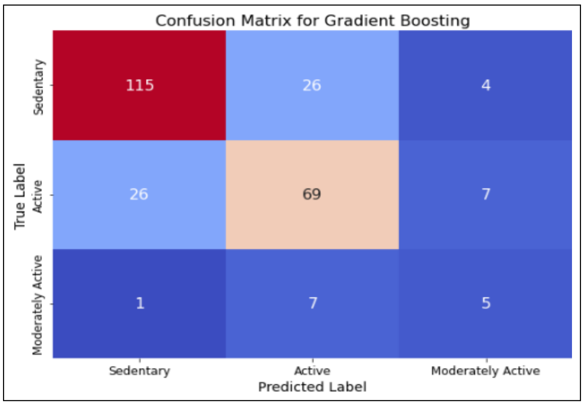

Figure 11: Confusion Matrix for Gradient Boosting (HEXACO Dataset)

Figure 11 shows that the confusion matrix indeed points toward a strong performance of Gradient Boosting, with high precision and recall for the "Sedentary" and "Active" classes and a more moderate performance for the "Moderately Active" class, indicating further room for improvement.

6.3 Experiment 3: Ensemble Techniques – Stacking Classifiers

This section reflects the performance of the Stacking Classifier, which uses several models as base models and performs with a meta-model to improve accuracy. Extensive tuning has been conducted after testing diverse combinations of base models coupled with meta-models, which showed optimality.

6.3.1 Best Configuration

The best improvement in configuration for the used Stacking Classifier included the use of Random Forest, Gradient Boosting, and AdaBoost as base models, while Gradient Boosting was the meta-model. It had reached an accuracy of 77.69%, the weighted F1-score measured at 0.77, and a macro-F1-score of 0.65, reflecting considerable balance regarding the various cuts at which its performance will be assessed.

6.3.2 Final Tuning Results

Class weighting was done to fine-tune the stacking classifier for imbalance; this further improved performance. The summary of the Classification Report is shown in Table 4.

Table 4: Classification Report of Tuned Stacking Model

Class	**Precision**	**Recall**	**F1-Score**
Active	0.83	0.81	0.82
Moderately Active	0.73	0.60	0.66
Sedentary	0.23	0.62	0.33
Overall			0.72
Macro Average	0.59	0.67	0.60
Weighted Average	0.76	0.72	0.73

The Multi-Class ROC Curve (Figure 12) gives insight into how the model classes are separated; basically, this model does exceptionally well on recall for the "Sedentary" class. Class-level predictions are better visualized on a confusion matrix (Figure 13). The model generalizes way more to minority classes and, especially, does a far better job distinguishing between the "Moderately Active" and "Sedentary" classes.

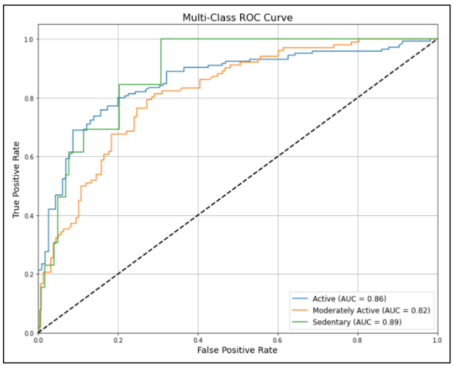

Figure 12: Multi-Class ROC Curve for Stacking Classifier (HEXACO Dataset)

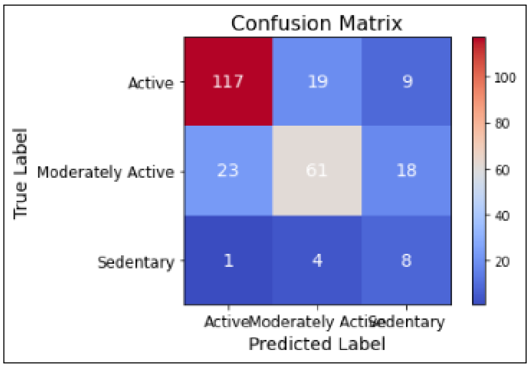

Figure 13: Confusion Matrix for Stacking Classifier (HEXACO Dataset)

The feature importance from the Gradient Boosting model offers an insight into the decision-making process of the meta-model. Figure 14 reveals that Fitbit Resting Heart Rate emerged as the strongest predictor, while BMI and the personality trait Honesty-Humility also figured in high. Behavioral traits also had a valuable contribution, specifically in sharpening the prediction for borderline cases, which pinpoints their role in the classification process.

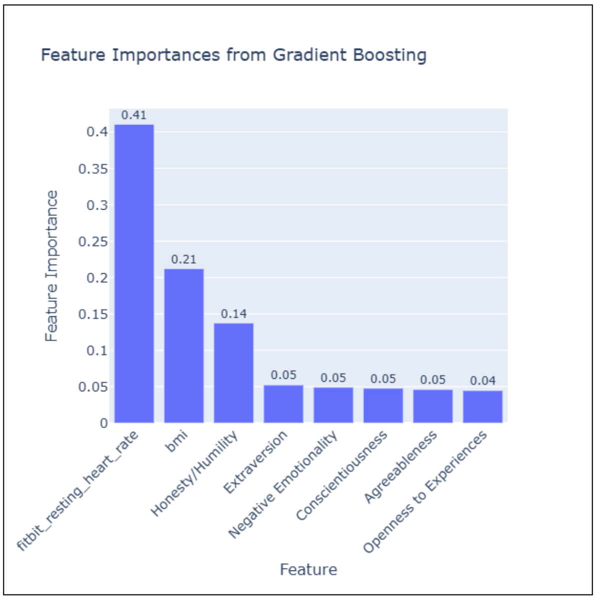

Figure 14: Feature Importance from Gradient Boosting (HEXACO Dataset)

6.3.3 Key Observations

1. Stacking Classifier performed at 72% accuracy, impressive weighted F1-score at 0.73, and a recall for the minority "Sedentary" class at 62%.

2. Weighting by class addressed the class balance well, improving those underrepresented classes without losing too much overall accuracy.
3. The analysis reflected the importance of physiological metrics, represented by resting heart rate and BMI, while behavioral traits also contributed greatly to the predictions.

6.4 Deployment Results

The last step was to deploy the best-performing model, the Stacking Classifier, to the survey data of real-world participants. This was essential in determining the model's robustness and applicability to real-world scenarios beyond those seen during training and testing.

Table 5: Classification Report of Participants Data

Class	Precision	Recall	F1-Score
Active	0.69	0.65	0.67
Moderately Active	0.55	0.69	0.61
Sedentary	0.75	0.43	0.55
Overall			0.62
Macro Average	0.66	0.59	0.61
Weighted Average	0.64	0.62	0.62

As shown in Table 5, the deployed model's overall accuracy improved to 62%, reflecting stronger performance in a real-world setting. The Active group had the highest F1-score of 0.67 among the activity classes, showing a reliable balance between precision and recall. In contrast, the Moderately Active class obtained higher recall, 69%, which indicated that for this group, the model was able to detect more subjects. The Sedentary group showed high precision, 0.75, but lower recall, 0.43, reflecting difficulties in the identification of all sedentary participants.

Further, an interactive Tableau Dashboard was created to see and analyze the deployment results, with the help of some predictions and their relationships to biometric and personality traits. Follow this link to the dashboard: Activity Level Dashboard on Tableau

These results emphasize the robustness and practical utility of the model in personalized health monitoring.

6.5 Discussion

The current study demonstrates that machine learning techniques, in particular, Gradient Boosting and Tuned Stacking Classifier, perform very well in classifying subjects according to physical activity based on physiological and psychological data. Figure 15 compares the Best Models from each experiment. The best performance was 73% using Gradient Boosting, while the Stacking Classifier presented a balanced result of 72%, managing class imbalances very well, especially regarding the minority "Sedentary" class. These findings also comply with previous studies that employ ensemble methods for handling imbalance and improving generalization.

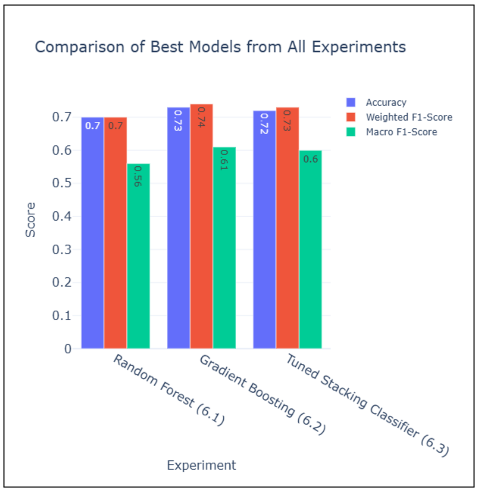

Figure 15: Comparison of Best Models from All Experiments (HEXACO Dataset)

The deployment phase confirmed the practical viability of the Stacking Classifier by showing reasonable accuracy and good generalization in most activity classes. While the model was good at detecting active participants, difficulties regarding the sedentary participants' classification highlighted that real-world data variability presents many challenges that should be targeted with enhanced preprocessing and tuning.

Despite all these promising results, the limitations include, among other factors, the presence of missing data and class imbalance, besides which the modest size of the dataset is a limiting factor that affects performance and generalization. Further improvements are likely to be made by addressing these shortcomings with more and larger datasets and better-advanced augmentation techniques.

The role of this study in emphasizing the application of machine learning in wearable health technologies and personalized fitness interventions is excellent, bringing a gap between theoretical modeling and practical applications. Generally, this contributes to elaborating an emerging branch of health that involves person-specific health using machine learning with multidimensional data in actionable insights.

7 Conclusion and Future Work

The current study has answered the following research question: *How can machine learning models use biometric and personality traits data to classify the level of physical activity effectively, showing the relationship between biometric and personality traits?* This research combined **physiological measures** such as Fitbit Resting Heart Rate and BMI with **HEXACO personality traits** using machine learning classification techniques that have classified physical activity and shed light on the interaction between biometric and personality data. The current analysis revealed that, among the biometric features, **Resting Heart Rate** was the major contributor. In contrast, among personality traits, **Honesty-Humility** played a key role in activity level predictions, thus affirming their complementary role in influencing physical behaviors.

While the best performance of **73%** was provided by **Gradient Boosting**, the **Tuned**

Stacking Classifier balanced all the metrics relatively well, treating **class imbalance** nicely, especially for the "Sedentary" class. Deployment of the Stacking Classifier further verifies **practical applicability** and shows **reasonable accuracy and generalizability** on diverse datasets. The results obtained significantly extend machine learning applications to provide **personalized health monitoring** and create a backbone for developing **adaptive interventions** to support active lifestyle and wellness.

Future work may extend the dataset to more diverse populations and additional features, such as **contextual and environmental features** of physical activity. **Real-world deployment and user feedback** could further refine the models. Deploying such machine learning models within **wearable devices or health platforms** may create commercial opportunities by providing **personalized recommendations** to foster healthier behaviors and deepening our insight into the relationship between biometrics and personality traits. The current study stands at the seam between **machine learning** and the **science of health behavior** to advance the subdiscipline of personalized intervention.

Acknowledgement

I, Ajinkya Chintawar, a student in MSc Data Analytics, would like to humbly take this opportunity to express my gratitude to my project supervisor, Dr. William Clifford, for his vast and continued guidance and support during this research project. His expertise and insightful suggestions have greatly helped bring the present study to execution.

I am also very grateful to the respondents who have taken the time to contribute to the survey; their contribution was valuable for the success of this research.

References

Ahmadi, M., O'Neil, M., Baque, E., Boyd, R. and Trost, S. (2020). Machine learning to quantify physical activity in children with cerebral palsy: Comparison of group, group-personalized, and fully-personalized activity classification models, *Sensors* **20**(14): 3976.

Alsareii, S., Awais, M., Alamri, A., AlAsmari, M., Irfan, M., Aslam, N. and Raza, M. (2022). Physical activity monitoring and classification using machine learning techniques, *Life* **12**(8): 1103.

Bianco, S. and Napoletano, P. (2019). Biometric recognition using multimodal physiological signals, *IEEE Access* **7**: 83581–83588.

Brons, A., Wang, S., Visser, B., Kröse, B., Bakkes, S. and Veltkamp, R. (2024). Machine learning methods to personalize persuasive strategies in mhealth interventions that promote physical activity: Scoping review and categorization overview, *Journal of Medical Internet Research* **26**: e47774.

Butte, N., Ekelund, U. and Westerterp, K. (2012). Assessing physical activity using wearable monitors measures of physical activity, *Medicine and Science in Sports and Exercise* **44**(1 Suppl 1): S5–S12.

Cristi-Montero, C. (2018). An integrative methodology for classifying physical activity level in apparently healthy populations for use in public health, *Revista Panamericana de Salud Pública* **41**: e161.

Dinh, A., Miertschin, S., Young, A. and Mohanty, S. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning, *BMC Medical Informatics and Decision Making* **19**(1): 1–15.

Fiedler, J., Eckert, T., Wunsch, K. and Woll, A. (2020). Key facets to build up ehealth and mhealth interventions to enhance physical activity, sedentary behavior and nutrition in healthy subjects – an umbrella review, *BMC Public Health* **20**: 1–21.

Garbarino, M., Lai, M., Bender, D., Picard, R. and Tognetti, S. (2014). Empatica e3—a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition, *2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, IEEE, pp. 39–42.

Gress, T. W., Denvir, J. and Shapiro, J. I. (2018). Effect of removing outliers on statistical inference: implications to interpretation of experimental data in medical research, *Marshall journal of medicine* **4**(2).

Harmouche-Karaki, M., Mahfouz, M., Salameh, P. and El Helou, N. (2023). Physical activity levels and predictors during covid-19 lockdown among lebanese adults, *Healthcare* **11**(14): 2080.

Islam, M. (2024). *Contextual Gamification Platform based Big5 Personality Trait and Preference Selection for User Recommendation*, PhD thesis, UIU.

Kaur, S., Bansal, K. and Kumar, Y. (2023). Machine learning-based approaches for accurately diagnosing and detecting hypertension disease, *Proceedings of the 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Vol. 10, IEEE, pp. 169–173.

Mannini, A. and Sabatini, A. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers, *Sensors* **10**(2): 1154–1175.

Mesanza, A., Lucas, S., Zubizarreta, A., Cabanes, I., Portillo, E. and Rodriguez-Larrad, A. (2020). A machine learning approach to perform physical activity classification using a sensorized crutch tip, *IEEE Access* **8**: 210023–210034.

Pallavi Suyog Uttekar, M. (2024). How many steps a day is considered active?, `https://www.medicinenet.com/how_many_steps_a_day_is_considered_active/article.htm`. Accessed: 2024-12-08.

Piciucco, E., Di Lascio, E., Maiorana, E., Santini, S. and Campisi, P. (2021). Biometric recognition using wearable devices in real-life settings, *Pattern Recognition Letters* **146**: 260–266.

Rochin Demong, N., Shahrom, M., Abdul Rahim, R., Omar, E. and Yahya, M. (2023). Personalized recommendation classification model of students' social well-being based on personality trait determinants using machine learning algorithms, *Journal of Information and Communication Technology* **22**(4): 545–585.

Saganowski, S., Perz, B., Polak, A. and Kazienko, P. (2022). Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review, *IEEE Transactions on Affective Computing* **14**(3): 1876–1897.

Seol, J., So, R., Murai, F. and Matsuo, T. (2024). Association between physical activity patterns of working-age adults and social jetlag, depressive symptoms, and presenteeism, *Journal of Occupational Health* p. uiae068.

Shaposhnyk, O., Yanushkevich, S., Babenko, V., Chernykh, M. and Nastenko, I. (2023). Inferring cognitive load level from physiological and personality traits, *2023 International Conference on Information and Digital Technologies (IDT)*, IEEE, pp. 233–242.

Stockwell, S., Trott, M., Tully, M., Shin, J., Barnett, Y., Butler, L., McDermott, D., Schuch, F. and Smith, L. (2021). Changes in physical activity and sedentary behaviours during the covid-19 pandemic lockdown, *BMJ Open Sport & Exercise Medicine* **7**(1): e000960.

Sánchez-Reolid, R., de la Rosa, F., López, M. and Fernández-Caballero, A. (2022). One-dimensional convolutional neural networks for low/high arousal classification from electrodermal activity, *Biomedical Signal Processing and Control* **71**: 103203.