Mehran University Research Journal of Engineering and Technology

https://doi.org/10.22581/muet1982.0062

2025, 44(2) 35-46

Exploring commuter stress dynamics through machine learning and double optimization

Ashar Ahmed [a, *], Mario Munoz-Organero [b], Bushra Aijaz [c]

[a] *Department of Urban and Infrastructure Engineering, NED University of Engineering and Technology, Karachi 75270, Pakistan*

[b] *Universidad Carlos III de Madrid, Av. Universidad, 30, 28911 Leganés, Madrid, Spain*

[c] *Independent Researcher, Karachi 75290, Sindh, Pakistan*

[*] Corresponding author: Ashar Ahmed, Email: aahmed@cloud.neduet.edu.pk

Received: 06 January 2025, Accepted: 27 March 2025, Published: 01 April 2025

KEYWORDS

Double optimization

Imbalanced dataset

Machine learning

Road safety

SHAP

Stress

ABSTRACT

Travel dynamics significantly impact commuter stress, influenced by traffic behavior, road conditions, travel modes, distance, and socio-demographic characteristics. Previous research on travel stress often exhibits limitations, including narrow scopes focusing on specific routes, vehicle types, or demographics. This study addresses these constraints by employing a comprehensive approach to analyze the influence of various travel attributes on commuter stress levels. An interview-based dataset was collected to capture the multifaceted experiences of road users. Five tree-based machine learning models–Decision Tree (DT), Random Forests (RF), Extra Trees (ET), Extreme Gradient Boosting (XGBoost), and k-Nearest Neighbor (k-NN)–were deployed for imbalanced multi-class classification. XGBoost demonstrated superior performance with the highest accuracy (73.33%) and precision (75.63%) with a standard deviation of ±5.9. A novel double hyperparameter optimization technique enhanced the prediction accuracy across all models, notably increasing the k-NN classifier's accuracy to 19.99%. The SHAP (SHapley Additive exPlanations) method was utilized for model interpretability, revealing distance traveled per day as the most influential factor across stress levels, followed by mode of transport, gender, and age for low, medium, and high-stress categories, respectively. The study also examines the impact of features on individual commuter stress levels through random instance selection. This research provides valuable insights into the complex interplay between travel attributes and commuter stress, paving the way for the development of effective stress mitigation strategies and improved travel experiences for all road users.

1. Introduction

Karachi, the largest city in Pakistan and ranked among the top 10 most populous metropolitan areas globally, spans approximately 560 square miles and is home to an estimated 15 million people as of 2017 [1, 2]. This sprawling urban center features a diverse demographic composition, varied infrastructure, and extreme climatic conditions. The city's extensive travel distances have profound implications for commuters' health, with the quality of road infrastructure playing a pivotal role in ensuring their physiological comfort and psychological well-being [3]. The discipline of traffic science is intricately tied to the safety and comfort of road users, emphasizing the need for systematic urban planning and effective traffic management.

Good infrastructure and a planned transportation system are vital for a city's prosperity. Conversely, inadequate infrastructure can adversely affect mental health, exacerbating stress and reducing the quality of life. Studies indicate that traffic congestion and transportation inefficiencies have resulted in significant psychological and physiological challenges for commuters [4]. Drivers are increasingly exposed to complex traffic scenarios, which complicate the prediction of stress responses [5].

Key factors contributing to these challenges include irregular and narrow roads, tight corners [6], double parking, wrong-way traffic, unsignalized intersections, speedy drivers [7], potholes, vehicle lights, CNG load-shedding, and slow or aggressive driving [8]. Additionally, road curvature and longitudinal gradients are reported to influence drivers' stress levels [9]. Passengers are also impacted by factors such as travel mode, trip duration, distance, waiting time for public transport [6], traffic congestion, vehicle acceleration or deceleration, and disruptive passenger behaviors [8].

Commuting has been identified as a significant source of stress for both working individuals and students [10]. While stress can arise from various commuting modes, some are particularly stressful compared to others. Legrain [11] determined that driving is the most stressful commuting mode among university students compared to walking and public transit. Similarly, Jahangeer [12] reported that 65.4% of medical students preferred university-provided transport over public transit due to overcrowding and discomfort, which exacerbate stress levels. In contrast, public transportation can serve as a stress-free alternative to private modes if it meets the requirements of seat availability, accessibility, safety, and cleanliness [13]. However, post-COVID-19, Singh [14] observed a notable decline in passengers using metro systems, carpooling, and buses. This trend has negatively impacted mental health by increasing social interaction anxiety and contributing to worsening traffic congestion.

Time pressure to reach the workplace, traffic noise, anxiety, and road rage are significant stress-inducing factors [15]. In this context, Montoro [16] explores the relationship between work environment, stress, and driving anger, finding that driving anger mediates the associations between traffic sanctions and driving stress and partially between driving experience, hourly intensity, and job-related stress. Furthermore, road environments and weather conditions are critical determinants of a driver's mental health [17]. Wei [18], through an experimental study involving 21 drivers, quantified the mental workload by incorporating physiological signals, traffic flow, and environmental factors. These findings revealed a positive correlation in the order of physiological signals > traffic flow > environmental factors. Nevertheless, this study was limited to a few test drives conducted on predefined routes.

Machine learning (ML) techniques have been widely applied in studies focused on predicting stress levels. For instance, a survey on mental health prediction among the working population employed Decision Tree (DT), Random Forest (RF), and Naïve Bayes models on the Open Sourcing Mental Illness (OSMI) survey dataset. The findings indicated that the DT model achieved the highest accuracy at 82% [19]. Similarly, another study utilized a multimodal dataset collected via the wearable device WESAD, proposing a stacking classifier model that demonstrated an impressive accuracy of 99.9% [20]. This model incorporated multiple classifiers, including Linear Discriminant Analysis (LDA), Adaptive Boosting (AdaBoost), RF, DT, and k-Nearest Neighbors (k-NN), to classify three mental states: Neutral, Stress, and Amusement. Additionally, a hybrid model combining Gradient Boosting Machine (GBM) and RF was proposed in another study, effectively classifying five stress levels while outperforming both ML and deep learning (DL) models in terms of accuracy and computational efficiency [21]. Moreover, a comparative study of 13 classifiers across three schemes revealed that the Gradient Boosting Tree (GBT) series achieved the highest accuracy, recorded at 77.25% [22].

Hyperparameter tuning is a crucial aspect of optimizing model performance and enhancing prediction accuracy. Without proper tuning, errors, outliers, and inconsistencies in the data can lead to overfitting. Overfitting occurs when a model learns patterns in the training data too well due to insufficient data, leading to suboptimal performance on unseen test data [23]. The challenges of small or imbalanced datasets exacerbate the risk of overfitting, further complicating the modeling process [24]. To address these challenges, this study introduces a Double Hyperparameter Optimization Technique (DHOT), an advanced method that refines the traditional hyperparameter optimization approaches. Grid search cross-validation is commonly used to find optimal parameters, but DHOT improves this process by adding an extra layer of optimization, resulting in more robust and reliable models.

Tree-based models like Random Forest (RF), XGBoost, and LightGBM are popular for many applications due to their inherent interpretability. However, when these models are combined with boosting and bagging techniques to enhance prediction performance, their interpretability and complexity diminish [25]. In this context, SHapley Additive exPlanations (SHAP) has emerged as a powerful tool for post hoc analysis, providing comprehensive methods for model interpretation and visualization. Based on cooperative game theory [26], SHAP elucidates how each feature impacts individual predictions. By attributing each prediction to the influence of different features, SHAP provides valuable insights into their contributions, making model predictions more transparent and trustworthy. Numerous studies have successfully demonstrated SHAP's effectiveness in enhancing the interpretability of complex ML models [9, 23, 27-28].

This study investigates the dynamics of commuter stress through Shapley Additive exPlanations (SHAP), a robust AI explainability tool. The analysis is based on data collected via a questionnaire survey conducted at NED University of Engineering and Technology, Pakistan. The dataset reflects real-world conditions by including diverse vehicle types, travel routes, distances, participant ages, and professions. Five tree-based machine learning classifiers were trained on this dataset, leveraging the advanced Double Hyperparameter Optimization Technique (DHOT) to enhance performance and accuracy. This research develops an explainable model to demonstrate how vehicle choice, travel distance, and socioeconomic factors influence mental health outcomes, uncovering insights often missed by traditional methods.

The paper structure is depicted in Fig. 1. Section 1 introduces the study, including a comprehensive literature review and identification of research gaps. Section 2 describes the methodology, encompassing the model development process and implementation of the DHOT approach, followed by model evaluation techniques. Section 3 presents the results and discussion, with a focus on model interpretation through SHAP explainers and a comparative analysis with prior studies. Finally, Section 4 offers concluding remarks and outlines recommendations for future research.

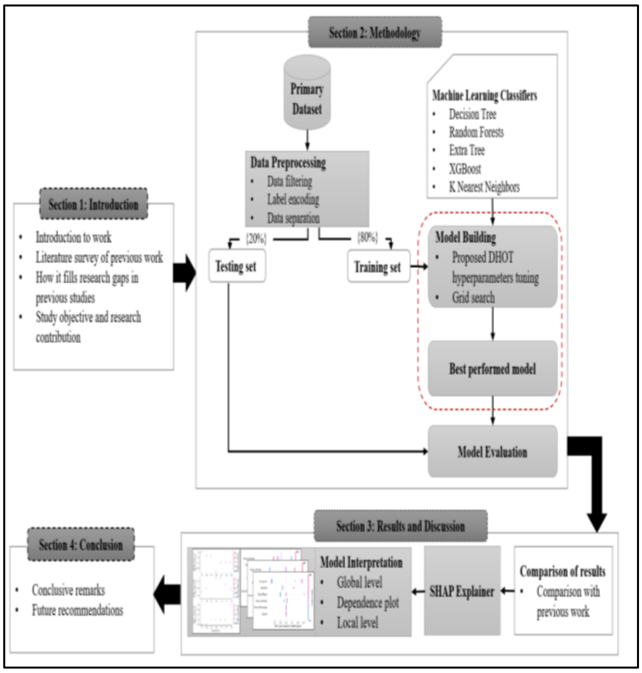

Fig. 1. Workflow and the Structure of the Paper

2. Methodology

2.1 Dataset Collection

The dataset for this study was collected through a structured questionnaire-based survey designed with two primary sections: (1) Routine mobility details and (2) Socio-economic information. Participants self-reported their stress levels using an 11-point Likert scale [29], ranging from 0 (strongly disagree) to 10 (strongly agree). Stress levels were categorized as: *High stress* when medication was required, *Medium stress* when rest sufficed, and *Low stress* when no intervention was necessary.

A total of 180 individuals voluntarily participated in the survey. Of these, the majority were female (54.4%), traveled primarily by private vehicles (53.89%), identified as students (47.8%), and reported daily travel distances categorized as Medium (14–26 km) (35.56%). The dataset comprises six input features and one multiclass target variable, Stress. The target variable is distributed across three classes: Low Stress (52.78%), Medium Stress (40%), and High Stress (7.22%), indicating a notable imbalance in data classes, as detailed in Table 1.

Class imbalance introduces challenges in machine learning model training, as models tend to favor majority classes, often at the expense of underrepresented ones. This can result in biased predictions or neglect of the minority class entirely. To address these challenges, further feature engineering and resampling techniques are employed to ensure balanced training and optimal model performance.

Table 1

General characteristics of respondents

	Category	Features	Class names and their responses
Input Attributes	Routine mobility	Distance per day	Short = 33.33%, Medium = 35.56%, Large = 31.11%
		Mode of Transport	Car = 27.78%, Bike = 26.11%, Rickshaw = 11.67%, Van = 5%, Chingchi = 9.44%, Point = 10%, Bus = 10%
		Driver or Passenger	Driver = 47.78%, Passenger = 52.22%
	Socio-Economic	Gender	Female = 54.4%, Male = 45.6%
		Age class	Young = 38.9%, Professional = 36.1%, Old = 25%
		Occupation	Student = 47.8%, Labor = 8.86%, TechStaff = 30.56%, Officer =7.78%, SnrHead = 5%
Output target	Mental stress level	Stress level	Low = 52.78%, Medium= 40%, High = 7.22%

2.2 Feature Engineering

The input feature DistancePerDay (DPD) was sorted in ascending order and subsequently divided into three classes. The short class included distances less than 14 km, the medium class ranged from 14-26 km, and the large class encompassed distances exceeding 26 km. Meanwhile, the Mode of Travel (MoT) feature was categorized into seven classes based on the diversity and privacy levels of vehicles commonly used in Karachi. These classes were assigned labels to reflect their comfort and privacy attributes. For instance, cars were classified as the most private and comfortable mode (Class 1), followed by bikes (Class 2), offering privacy but reduced comfort compared to cars. Modes such as rickshaws and van services, which cater to limited passengers along dedicated routes, were assigned Class 3 and Class 4, respectively. Chingchi, point services, and buses were categorized as Class 5, Class 6, and Class 7, respectively. Chingchi vehicles, in particular, are widely used paratransit vehicles in developing countries, offering low-cost, energy-efficient transportation over short distances and a rapid alternative to traditional public transport [30].

The Occupation (Occu) feature reflects participants' socio-economic roles, with classifications based on age, profession, and economic stability. Students, being younger and more dynamic, constituted one class. Laborers and Technical staff, representing mature, skilled personnel, formed the next two classes. The final two classes included officers and senior management, representing professionals who are socially collaborative, financially stable, and often vehicle owners. This feature introduces variability in predicting stress, as participants' road experiences and socio-economic conditions directly influence mental health outcomes.

Other input features, including gender, age class, and type of commuter, are inherently self-explanatory.

The target variable, Mental Stress (MS), was categorized into three levels based on participants' Likert-scale scores: Low (0-3), Medium (4-7), and High (8-10) [31]. Table 2 presents the statistical properties of the features, including mean and standard deviation. The dataset was complete, with no missing attributes.

Given the inherent class imbalance in the dataset, stratification was applied to maintain a proportional representation of each target class during training and testing splits, set at an 80:20 ratio. For small datasets, stratification ensures that machine learning models are exposed to all sample varieties during their learning phase, preserving the original dataset's class ratio during the training and testing phases. A detailed description of the formats of the training and testing datasets is depicted in Table 3. This is in accordance of the similar approach presented in [32].

Table 2

Mean and standard deviation of input features and target output

Feature names	Class names, [encoding]	Mean	Standard deviation
Gender, G	Female [1], Male [2]	1.45	0.49
Age Class, AC	Young [1], Professional [2],	1.86	0.79

	Old [3]		
Distance Per Day, DPD	Short [1], Medium [2], Large [3]	1.98	0.8
Mode of Travel, MoT	Car [1], Bike [2], Rickshaw [3], Van [4], Chingchi [5], Point [6], Bus [7]	3.12	2.07
Occupation, Occu	Student [1], Labor [2], Tech Staff [3], Officer [4], Senior Head [5]	2.13	1.24
Driver or Passenger, DoP	Driver [1], Passenger [2]	1.52	1.52
Stress, MS	Low [0], Medium [1], High [2]	0.54	0.63

Table 3

Shapes of Training and Test Sets

Shapes	X	y
Shape of Train	(144, 6)	(144,)
Shape of Test	(36, 6)	(36,)

2.3 Experimental Setup

This study employs five machine learning classifier algorithms: Decision Tree (DT) [33], Random Forest (RF) [34], Extra Trees (ET) [35], eXtreme Gradient Boosting (XGBoost) [36], and K-Nearest Neighbors (k-NN) [37]. Each model depends on hyperparameter tuning to optimize performance, as hyperparameters critically influence the model's structure and predictive accuracy.

To optimize performance, this study introduces a Double Hyperparameter Optimization Technique (DHOT) implemented in two distinct phases. The flowchart of the proposed model is illustrated in Fig. 2.

In phase I, grid search cross-validation is used to identify the best set of parameters. This process balances bias and variance by selecting the optimal value of *K*, thereby minimizing the risk of overfitting. Once the optimal parameters are identified, the model is discarded to prevent inefficiencies caused by repeated retraining cycles. In Phase II, a new model is trained and transformed with a standard scalar, utilizing the parameters optimized in Phase I. The stratified 4-fold cross-validation with three repetitions is employed to ensure robust generalization to unseen data. Following the grid search CV, the dataset split was adjusted to a 75:25 ratio for training and testing, in phase II. It improved the classifier performance, as detailed in Section 3.2. The DHOT framework guarantees that each classifier is fine-tuned with parameters uniquely tailored to its architecture, as different models exhibit varying performance depending on parameter configurations. Table 4 presents the optimal hyperparameters determined during Phase I of the DHOT scheme. Notably, the cross-validation configuration *(CV)* was standardized to four folds across all models.

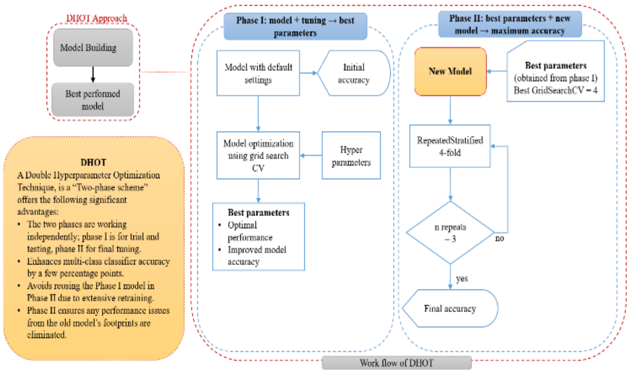

Fig. 2. Flowchart for a Proposed Model Explaining the Double Hyperparameters Optimization Technique

Table 4

Optimal hyperparameters found using a grid search CV after phase I

Models	Optimal hyperparameters obtained after grid search
Decision Trees	maximum depth = 3, min. samples split = 9, minimum sample leaves = 1
Random Forests	maximum depth = 6, min. samples split = 10, minimum sample leaves = 6, n estimators = 50
Extra Tree	maximum depth = 3, min. samples split = 7, minimum samples leaves = 8, n estimators = 10
XGBoost	maximum depth = 2, n estimators = 10, colsample_bytree = 0.1, learning rate = 0.5
k-NN	n neighbors = 18, leaf size = 7

2.4 Model Explainability Approach

Machine learning models are often perceived as "black box" systems due to their limited interpretability. This study utilizes SHapley Additive exPlanations (SHAP), an advanced framework for explainable AI, offering insights from global to local levels [38]. SHAP (version 0.44.0) is utilized to interpret the best-performing classifier among the five models. After training the selected model, SHAP explainers are generated using SHAP values associated with each feature or class, also referred to as "Expected" or

"Base" values. SHAP visualizations, such as summary and dependence plots, effectively illustrate positive and negative correlations of categorical features, offering a significant advantage in model interpretation. The explainability process is carried out in three stages, enabling a comprehensive post hoc analysis of the model's behavior.

2.4.1 Global-level interpretation (Stage I)

At the global level, the mean absolute SHAP values are used to rank the importance of features in predicting stress levels. This ranking can differ across target classes. Summary plots are generated to illustrate the general behavior of the model, highlighting the most influential features for each stress level. The global feature importance is mathematically expressed as Eq. (1).

$$Importance_j = \frac{1}{K} \sum_{i=1}^{K} \left|\varphi_{i,j}\right| \tag{1}$$

Where:

$\varphi_{i,j}$: SHAP value of sample *i* for feature *j*.

K: Total number of samples used in the interpretability analysis.

2.4.2 Feature-interaction interpretation (Stage II)

Feature-interaction effects are examined using dependence plots, which reveal the relationships between features and their contributions to predictions for each class. These plots provide a deeper understanding of how features interact and collectively influence the model's predictions.

2.4.3 Local-level interpretation (Stage III)

At the local level, the contribution of individual features varies across specific data instances. For example, the mode of travel might be the dominant factor contributing to stress for one individual, whereas travel distance could play a more significant role for another. SHAP enables a detailed instance-level analysis to explain how different features influence predictions in a multi-class classifier.

For this study, a randomly selected instance is analyzed to demonstrate SHAP's local-level interpretability. The expected value for each class is given in Eq. (2):

$$EV_i = \mu_i + \sum_{j=1}^{M} \emptyset_{ij} \tag{2}$$

Where:

EV_i: Expected value for class *i*.

μ_i: Base value for class *i*.

$\emptyset_{ij}$: SHAP value for feature *j* and class *i*.

M: Total number of features.

3. Results and Discussions

3.1 DHOT Analysis

The implementation of the Double Hyperparameter Optimization Technique (DHOT), as outlined in Fig. 2, significantly enhances the accuracy and efficiency of the five ML classifiers, particularly for small datasets. Fig. 3 highlights the improvements achieved through DHOT, with notable accuracy gains across all models. Among the classifiers, XGBoost demonstrates the highest accuracy, while k-NN exhibits the largest percentage increase, achieving a 19.99% improvement. The Random Forest (RF) classifier follows with a 16.36% increment, whereas XGBoost, Extra Trees (ET), and Decision Tree (DT) classifiers show enhancements of 14.77%, 12.73%, and 9.09%, respectively.

It is observed that ET and XGBoost achieved optimal performance at *n_estimators* = 10, whereas RF performs best at *n_estimators* = 50 (Table 4). Each model also attains peak performance at its respective optimal maximum depth, underscoring the importance of targeted hyperparameter tuning in boosting model accuracy and reliability.

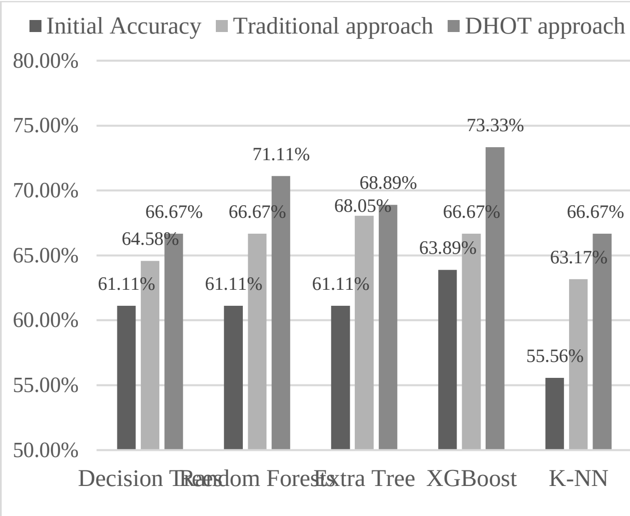

Fig. 3. Accuracies of ML Models 'Before Optimization', 'After Traditional Approach, and 'After DHOT Approach,

3.2 Comparison and Evaluation of Performance Metrics

The classification performance of all five ML classifiers is summarized in Table 5, which includes precision, recall, and F1-score metrics. Among the classifiers, XGBoost achieves the highest performance with precision at 75.6%, recall at 73.3%, and an F1 score of 69.8%. The Random Forest (RF) classifier follows closely, with precision, recall, and F1-score values of 74.6%, 71.1%, and 69%, respectively. For the Extra Trees (ET) classifier, these metrics are 70.3%, 66.7%, and 65.2%, respectively. The Decision Tree (DT) and k-NN classifiers show similar performance, with precision scores of 69.4% and 68.2% and identical recall scores of 66.7%. The corresponding F1 scores are 63.4% for DT and 61.6% for k-NN.

To assess performance variability, standard deviation (SD) is incorporated into the analysis. A low SD reflects robustness and minimal sensitivity to random variations in the data. Among the classifiers, k-NN and DT exhibit lower SD, indicating consistent performance across multiple runs or cross-validation folds.

Following Li's approach [39], the classification accuracy is adjusted by ± SD to evaluate statistical significance. The revised accuracy range for XGBoost, spanning from 67.43% to 79.23%, confirms it as the most reliable model based on both minimum and maximum accuracy brackets. In terms of overall performance, the classifiers rank as follows: XGBoost > RF > ET > k-NN > DT.

Table 5

Comparison of classification report of experimented models

	DT	RF	ET	XGB	k-NN
Accuracy (%)	66.67	71.11	68.89	73.33	66.67
Precision (%)	69.4	74.6	70.3	75.6	68.2
Recall (%)	66.7	71.1	66.7	73.3	66.7
F1-score (%)	63.4	69.0	65.2	69.8	61.6
SD (%)	4.6	6.5	5.7	5.9	4.5
Accuracy ± SD (%)	62.07 ↔ 71.27	64.61 ↔ 77.61	63.19 ↔ 74.59	67.43 ↔ 79.23	62.17 ↔ 71.17
Rank	5	2	3	1	4

The study identifies the XGBoost model as the most effective for the stress prediction problem, outperforming other tree-based models in all evaluated metrics. Consequently, the XGBoost classifier is selected for detailed analysis and interpretation using SHAP in the subsequent sections.

3.3 SHAP Global Interpretation

The SHAP summary plot provides a global interpretation of the XGBoost model's output, with features arranged in descending order of importance (Fig. 4). The top-ranked feature has the greatest impact on stress prediction. Positive SHAP values on the x-axis indicate a higher likelihood of increased stress, while negative values suggest a mitigating effect. The multi-colored vertical bars reflect class intensity: blue denotes lower values, and red signifies higher values. The three subplots represent stress levels –low, medium, and high.

For low stress (Fig. 4a), DistancePerDay (DPD) and ModeOfTravel (MoT) are the most influential factors. Shorter distances and private travel modes positively correlate with reduced stress, aligning with findings by [12], which highlighted the stress-reducing effects of shorter commutes and personal travel modes.

In medium stress levels (Fig. 4b), DPD remains the dominant factor, with a lesser contribution from MoT. Stress levels increase with travel distance, consistent with studies [12] and [40] showing that commute duration adversely affects satisfaction irrespective of travel mode. Additionally, the model suggests males are more prone to medium stress on longer routes than females.

High-stress levels (Fig. 4c) are predominantly influenced by Occupation and AgeClass, unlike other stress levels. Officers and technical staff experience greater stress than students, and older individuals are more affected than younger ones. This is supported by [17], which states that older age, long distances, and higher income contribute to high stress levels. These factors are typical of top-position employees who are likely to suffer from job stress, which, in turn, contributes to travel-related stress. Interestingly, the type of commuter (driver or passenger) has minimal influence on predicting road stress.

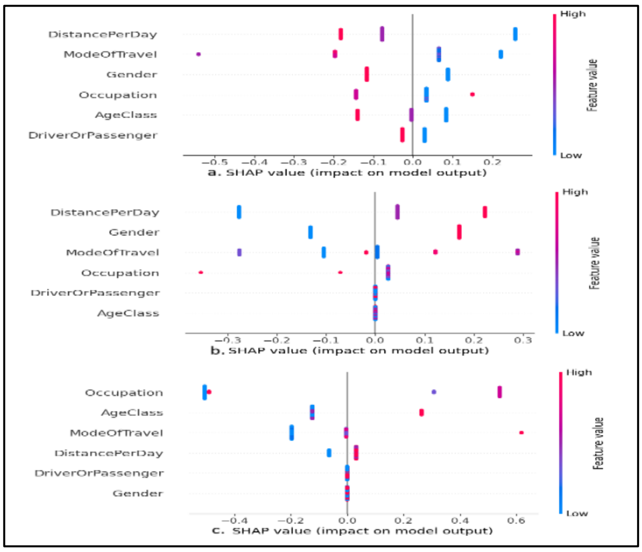

Fig. 4. Global Level Interpretation of Stress (a) Low, (b) Medium, (c) High

3.4 SHAP Feature Interaction Interpretation

'SHAP Dependence Plots' illustrate feature interactions for the XGBoost model. Fig. 5 visually represents the relationship between DistancePerDay (DPD) and ModeOfTravel (MoT) through SHAP values. The x-axis represents seven MoT subclasses, ranging from private modes (e.g., cars, bikes) to public transport (e.g., buses). The y-axis displays SHAP values while the vertical color intensity reflects DPD, with red indicating larger distances and blue indicating shorter distances.

Key observations from the SHAP dependence plot include:

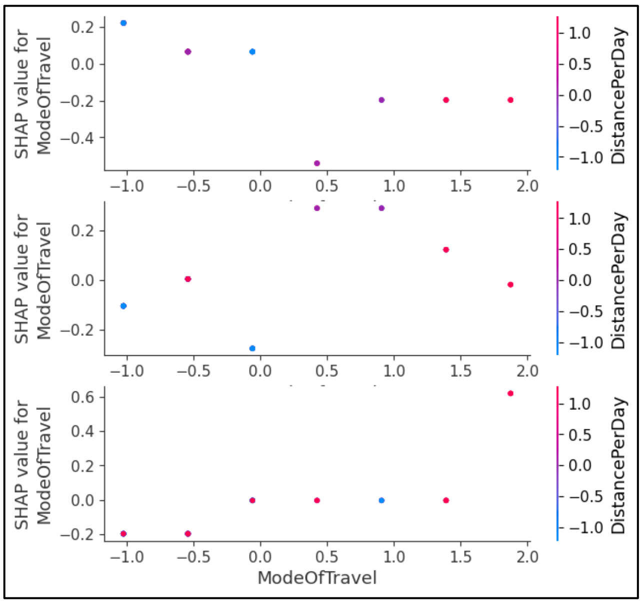

Fig. 5. Feature Interaction Interpretation using Dependence Plot for XGBoost Classifier

(i) Short-distance travelers experience lower stress

Individuals traveling shorter distances using private or semi-private modes of transport experience low stress levels. This aligns with [6], which found that shorter walking or commuting distances improve travel satisfaction.

(ii) Stress levels among bike riders and chingchi travelers

Both bike riders and Chingchi travelers exhibit low stress levels, especially for medium distances. This is supported by [41], which associates paratransit modes like Chingchi with efficient navigation through congested areas, cost-effectiveness, time savings, and limited passenger occupancy, all contributing to reduced stress levels. Additionally, bike riders can navigate traffic more easily, often using shortcuts or less congested routes, further reducing stress levels.

(iii) Public transport use for long distances

Long-distance commuters predominantly prefer buses, while short-distance travelers tend to avoid them. This highlights the practical use of public transport for extended commutes.

(iv) Van Services as Stress-Free Travel

Van services consistently predict the lowest stress levels across all distances. This can be attributed to features like dedicated seating, fixed routes, and limited passenger occupancy, offering a comfortable travel experience.

(v) Increased Trust in Model Predictions

The model does not predict Chingchi users for long distances, which is practically accurate since Chingchi services are typically offered for short distances or nearby places. This observation increases trust in the model's prediction.

3.5 SHAP Local Interpretation

SHAP facilitates effective local interpretation at the individual level, enabling a detailed analysis of specific predictions. This is demonstrated using a Decision Plot, which visualizes the step-by-step contribution of SHAP values for individual features towards the model's final prediction [42].

A few random instances are selected to interpret the model's decision-making process. Fig. 6 displays three colored lines representing the predicted stress classes: "High" (blue), "Medium" (purple), and "Low" (red-dotted). The rightmost line indicates the final prediction. The central grey line is the baseline, which is 0.346 for our model. The baseline is the average of the expected values for all output classes in the model. Deviations of the stress-class lines from the baseline reflect the positive or negative contributions of features toward the final prediction. The dotted line indicates the stress level predicted by the model.

For a random instance, e.g., sample_index =0, is interpreted for the local level interpretation as shown in Fig. 6(a). The expected values for Low, Medium, and high stress levels are [1.096, 0.803, -0.860], respectively. The XGBoost model identifies *Occupation*, *DPD*, and *AgeClass* as the top contributing features for this prediction. The participant is assigned a "low" stress level, which aligns with her profile retrieved from the dataset (Table 6). She is a student (occupation), travels a short distance (DPD), and belongs to a younger age group (AgeClass). These factors collectively influence the prediction of a low-stress level.

For sample_index = 7 (Fig. 6b), the XGBoost model identifies the features such as ModeOfTravel, Occupation, and Gender as key contributors towards stress level, whereas for sample_index = 12 (Fig. 6c) the XGBoost model identifies *Occupation, AgeClass* and *DistancePerDay* as the most contributing features towards stress level. The participant in Fig. 6(b) is classified with a 'medium' stress level, while the one in Fig. 6(c) is assigned a "high" stress level. These classifications align with the actual stress levels reported in the survey, as confirmed by their respective profiles in the dataset.

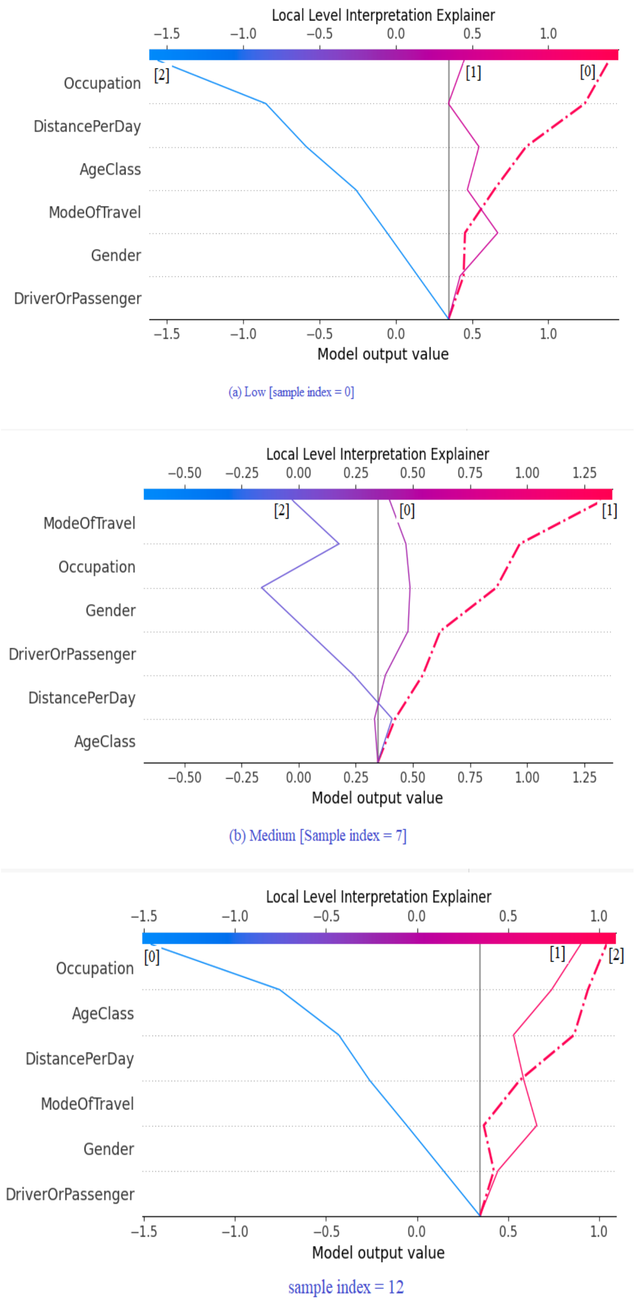

Fig. 6. A Scenario for Randomly Chosen Instance (a) Low, (b) Medium (c) High

Table 6

Details retrieved from the dataset of sample index =0

Features	Report
Gender	Female
Age class	Young
Distance per Day	Short
Mode of Travel	Rickshaw
Occupation	Student
Driver or passenger	Passenger
Stress level	Low

4. Conclusion

The study demonstrates a robust understanding of the stress prediction model and its interpretability at both global and local levels. Among the classifiers evaluated, the XGBoost model stands out as the most suitable due to its superior accuracy, precision, and low standard deviation, making it highly stable for small and imbalanced datasets that are prone to overfitting.

The reported XGBoost accuracy (73.33%) reflects the complexity of our multiclass stress classification task, dataset characteristics, and feature space. Unlike previous studies that may have used binary classification, different datasets, or feature engineering techniques, our approach priorities real-world generalizability over solely optimizing accuracy. Moreover, for smaller datasets ($n < 200$) and fewer input variables ($x \leq 6$), model accuracy can be significantly enhanced through a double hyperparameter optimization technique (DHOT). Traditional optimization methods improve accuracy moderately, DHOT achieves substantial gains – enhancing the XGBoost model compared to 4.35% improvement from conventional methods. Similarly, DHOT outperforms standard tuning in k-NN, increasing accuracy by 19.99% versus 13.69%. While traditional optimization methods remain effective for larger datasets, DHOT demonstrates particular efficacy in maximizing accuracy for smaller datasets.

The study also underscores the variability in stress levels among individuals, even when exposed to similar factors. Utilizing SHAP as an explainable AI tool, the model provides insights through global and local interpretations. The summary plot ranks features by their global influence, while the decision plot reveals the most critical factors for individual predictions, enhancing the model's transparency and trustworthiness.

Furthermore, this work offers a comprehensive analysis of road travel stress by considering diverse factors such as vehicle types, age groups, socio-economic backgrounds, continuous travel histories, varied route experiences, and the effects of road infrastructure and traffic congestion. These findings can guide policymakers in designing targeted interventions to mitigate travel stress, such as improving road conditions, optimizing traffic flow, and promoting alternative travel modes.

In conclusion, this study underscores the importance of adopting inclusive, data-driven approaches to enhance commuter well-being and transportation system efficiency, offering valuable insights for advancing smart and sustainable urban mobility solutions.

5. Acknowledgment

The authors acknowledge the efforts of the students who helped gather the data.

6. References

[1] Britannica, "Karachi, Pakistan", Accessed: Dec. 03, 2024. Available at: https://www.britannica.com/place/Karachi.

[2] Worldometer, "Largest cities in the world", Accessed: Dec. 03, 2024. Available at: https://www.worldometers.info/population/largest-cities-in-the-world/.

[3] M. Welde and E. Tveter, "The wider local impacts of new roads: A case study of 10 projects", Transport Policy, vol. 115, pp. 164–180, 2022, doi: 10.1016/j.tranpol.2021.11.012.

[4] R. Mushtaq and O. Hashmi, "Traffic congestion and prevalence of mental and physical health issues in Karachi, Pakistan", Journal of Psychology, vol. 53, no. 02, pp. 15–24, 2022.

[5] V. C. Magaña, M. M. Organero, J. A. Fisteus, and L. S. Fernández, "Estimating the stress for drivers and passengers using deep learning", Proceedings of "JARCA 2016", Almería, Spain, 2016, pp. 1–6.

[6] M. A. Conceição, M. M. Monteiro, D. Kasraian, P. van den Berg, S. Haustein, I. Alves, C. L. Azevedo, and B. Miranda, "The effect of transport infrastructure, congestion, and reliability on mental wellbeing: a systematic review of empirical studies", Transport Reviews, vol. 43, no. 2, pp. 264–302, 2023, doi: 10.1080/01441647.2022.2100943.

[7] A. Ahmed and B. Aijaz, "A Case Study on the Potential Applications of V2V Communication for Improving Road Safety in Pakistan", Engineering Proceedings, vol. 32, no. 1, p. 17, Karachi, Pakistan, 2023, doi: 10.3390/engproc2023032017.

[8] S. Mirza, M. Pervaiz, and N. Khatoon, "Stress-inducing factors among occupational drivers in Karachi, Pakistan", Eastern Mediterranean Health Journal, vol. 26, no. 10, pp. 1233–1241, 2020, doi: 10.26719/emhj.20.059.

[9] S. Zhong, X. Fu, W. Lu, F. Tang, and Y. Lu, "An expressway driving stress prediction model based on vehicle, road, and environment features", IEEE Access, vol. 10, pp. 57212–57226, 2022, doi: 10.1109/ACCESS.2022.3165570.

[10] G. Gottholmseder, K. Nowotny, G. J. Pruckner, and E. Theurl, "Stress perception and commuting", Health Economics, vol. 18, no. 5, pp. 559–576, 2009, doi: 10.1002/hec.1389.

[11] A. Legrain, N. Eluru, and A. M. El-Geneidy, "Am stressed, must travel: The relationship between mode choice and commuting stress", Transportation Research Part F: Traffic Psychology and Behaviour, vol. 34, pp. 141–151, 2015, doi: 10.1016/j.trf.2015.08.001.

[12] S. M. A. Jahangeer, N. Hasnain, M. T. Tariq, A. Jamil, S. Y. Zia, and W. Amir, "Frequency and association of stress levels with modes of commuting among medical students of a developing country", Malaysian Journal of Medical Sciences, vol. 28, pp. 113–122, 2021, doi: 10.21315/mjms2021.28.4.12.

[13] D. Ettema, M. Friman, T. Garling, and L. E. Olsson, "Travel mode use, travel mode shift, and subjective well-being: Overview of theories, empirical findings and policy implications", Mobility, Sociability and Wellbeing of Urban Living, D. Wang and S. He, Eds. Berlin, Heidelberg: Springer, 2016, pp. 129–150. ISBN: 978-3-662-48184-4.

[14] V. Singh, K. Gupta, A. Agarwal, and N. Chakrabarty, "Psychological impacts on the travel behaviour post COVID-19", Asian Transport Studies, vol. 8, p. 100087, 2022, doi: 10.1016/j.eastsj.2022.100087.

[15] M. C. I. van Schalkwyk and J. S. Mindell, "Current issues in the impacts of transport on health", British Medical Bulletin, vol. 125, no. 1, pp. 67–77, 2018, doi: 10.1093/bmb/ldx048.

[16] L. Montoro, S. Useche, F. Alonso, and B. Cendales, "Work environment, stress, and driving anger: A structural equation model for predicting traffic sanctions of public transport drivers", International Journal of Environmental Research and Public Health, vol. 15, no. 3, p. 497, 2018, doi: 10.3390/ijerph15030497.

[17] N. A. Mohammad, "Stress and anxiety on the road: The silent victims and their sufferings", Journal of Advanced Vehicle System, vol. 13, no. 1, pp. 57–69, 2022.

[18] W. Wei, X. Fu, S. Zhong, and H. Ge, "Driver's mental workload classification using physiological, traffic flow, and environmental factors", Transportation Research Part F: Traffic Psychology and Behaviour, vol. 94, pp. 151–169, 2023, doi: 10.1016/j.trf.2023.02.004.

[19] M. Kavitha, M. Pingili, M. Spurthi, S. K. Fayaz, K. Kirthan, and S. Santhosh, "Classification algorithms based mental health prediction using data mining", Turkish Journal of Computer and

Mathematics Education, vol. 13, no. 2, pp. 1168–1175, 2020, doi: 10.17762/turcomat.v13i2.13708.

[20] V. Dham, K. Rai, and U. Soni, “Mental stress detection using artificial intelligence models”, Journal of Physics: Conference Series, vol. 1950, p. 012047, 2021, doi: 10.1088/1742-6596/1950/1/012047.

[21] C. Ding, Y. Zhang, and T. Ding, “A systematic hybrid machine learning approach for stress prediction”, PeerJ Computer Science, vol. 9, p. e1154, 2023, doi: 10.7717/peerj-cs.1154.

[22] Y. Meng, N. Yang, Z. Qian, and G. Zhang, “What makes an online review more helpful: An interpretation framework using XGBoost and SHAP values”, Journal of Theoretical and Applied Electronic Commerce Research, vol. 16, no. 3, pp. 466–490, 2020, doi: 10.3390/jtaer16030029.

[23] Y. Xue, “An overview of overfitting and its solutions”, Journal of Physics: Conference Series, vol. 1168, no. 2, p. 022022, 2019, doi: 10.1088/1742-6596/1168/2/022022.

[24] M. U. Abdulazeez, W. Khan, and K. A. Abdullah, “Predicting child occupant crash injury severity in the United Arab Emirates using machine learning models for the imbalanced dataset”, IATSS Research, vol. 47, no. 2, pp. 134–159, 2023, doi: 10.1016/j.iatssr.2023.05.003.

[25] S. Devlin, C. Singh, W. J. Murdoch, and B. Yu, “Disentangled attribution curves for interpreting random forests and boosted trees”, Machine Learning, 2019, doi: 10.48550/arXiv.1905.07631.

[26] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions”, Advances in Neural Information Processing Systems, vol. 30, 2017.

[27] M. T. Kashifi, “Investigating two-wheeler risk factors for severe crashes using an interpretable machine learning approach and SHAP analysis”, IATSS Research, vol. 47, no. 3, pp. 357–371, 2023, doi: 10.1016/j.iatssr.2023.07.005.

[28] S. M. Lundberg, G. G. Erion, and S.-I. Lee, “Consistent individualized feature attribution for tree ensembles”, ArXiv, vol. abs/1802.03888, 2018.

[29] G. M. Sullivan and A. R. Artino Jr., “Analyzing and interpreting data from Likert-type scales”, Journal of Graduate Medical Education, vol. 5, pp. 541–542, 2013, doi: 10.4300/JGME-5-4-18.

[30] T. Shimazaki and M. Rahman, “Physical characteristics of paratransit in developing countries of Asia”, Journal of Advanced Transportation, vol. 30, no. 2, pp. 5–24, 1996, doi: 10.1002/atr.5670300203.

[31] M. McCaffery and A. Beebe, Pain: Clinical Manual for Nursing Practice. Mosby, 1989.

[32] S. M. Saqib, T. Mazhar, M. Iqbal, T. Shahazad, A. Almogren, A. U. Rehman, et al., “Enhancing electricity theft detection with ADASYN-enhanced machine learning models”, Electrical Engineering, 2025/03/23 2025. doi: 10.1007/s00202-025-03044-4.

[33] E. Bartz, T. Bartz-Beielstein, M. Zaefferer, and O. Mersmann, Hyperparameter Tuning for Machine and Deep Learning with R., 2023.

[34] L. Breiman, “Random forests”, Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[35] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees”, Machine Learning, vol. 63, pp. 3–42, 2006, doi: 10.1007/s10994-006-6226-1.

[36] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system”, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, USA, 2016, ISBN 9781450342322.

[37] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithm”, Machine Learning, vol. 6, pp. 37–66, 1991, doi: 10.1007/BF00153759.

[38] L. S. Shapley, “A value for n-person games”, Contributions to the Theory of Games II, H. W. Kuhn and A. Tucker, Eds. Princeton: Princeton University Press, 1953, pp. 69–79.

[39] L. Li, X. Pan, H. Yang, T. Zhang, and Z. Liu, “Supervised dictionary learning with regularization for near-infrared spectroscopy classification”, IEEE Access, vol. 7, pp. 100923–100932, 2019, doi: 10.1109/ACCESS.2019.2930288.

[40] K. Chatterjee, S. Chng, B. Clark, A. Davis, J. De Vos, D. Ettema, S. Handy, A. Martin, and L. Reardon, “Commuting and wellbeing: a critical overview of the literature with implications for policy and future research”, Transport Reviews,

vol. 40, no. 1, pp. 5–34, 2020, doi: 10.1080/01441647.2019.1649317.

[41] F. Rahman, Md. A. Islam, and Md. Hadiuzzaman, “Paratransit service quality modeling reflecting users' perception: A case study in Dhaka, Bangladesh”, IATSS Research, vol. 47, no. 3, pp. 335–348, 2023, doi: 10.1016/j.iatssr.2023.07.001.

[42] M. Harikha, J. C. Kimmel, M. Abdelsalam, and M. Gupta, “Analyzing and explaining black-box models for online malware detection”, IEEE Access, vol. 11, pp. 25237–25252, 2023, doi: 10.1109/ACCESS.2023.3255176.

