

Stress Level Prediction Using Machine Learning: A Comparative Analysis of Lifestyle Factors

Sakib Chowdhury¹, Md Myan Uddin², Shovon Guha³, Md Raihan Mia⁴, Faiza Tanzil⁵

^{1, 2, 3, 4, 5} Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China

doi: <https://doi.org/10.37745/ejcsit.2013/vol14n3118>

Published April 18, 2026

Citation: Chowdhury S., Uddin M.M., Guha S., Mia M.R., Tanzil F. (2026) Stress Level Prediction Using Machine Learning: A Comparative Analysis of Lifestyle Factors, *European Journal of Computer Science and Information Technology*, 14(3), 1-18

Abstract: *Stress is a growing mental health issue in the world, as it is predisposed to the lifestyle factors that are amenable to modification. This paper applies machine learning to identify the main predictors of stress and assess model performance. A sample of 2,000 participants who had 11 lifestyle characteristics, such as sleep, digital habits, physical activity, workload, and environment, was examined. Median (1.40) was used to binarize stress levels. The 5-fold stratified cross-validation was used with hyperparameter optimization to train four models: Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost. The performance of all models was high (ROC-AUC > 0.97), and the highest accuracy was demonstrated by Logistic Regression (91.75%), and ROC-AUC (0.983). Analysis of feature importance showed that daily pending tasks had a contribution of approximately 34.6 percent to predictions, which was far more than other factors. The moderate effect (7-8% of influence) was on sleep variables, and the remaining features had a small effect (less than 0.5%). These results confirm the Task Load Hypothesis, which underlines workload management as the most effective method of reducing stress.*

Keywords: Stress prediction, machine learning, lifestyle factors, feature importance, mental health analytics, workload management

INTRODUCTION

Stress is an important and increasing mental health issue that negatively impacts the well-being, productivity, and quality of life of people. As the load of work and lifestyle alterations increases, it is now necessary to determine the main factors affecting stress to be able to intervene and manage them. Conventional statistical methods would not be sufficient to describe complex associations among several lifestyle variables, and they have limited predictive power.

New developments in machine learning (ML) have offered effective solutions to the modeling of high-dimensional complex data and enhancing prediction accuracy in mental health research. Stress has been linked to different lifestyle factors such as sleep patterns, physical activity, screen time, social interaction, and workload. Nevertheless, there is no agreement in the literature regarding the relative significance of these aspects, especially when they are considered in the framework of a single ML. In this work, several machine learning models are used, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost, to forecast stress levels based on lifestyle data. Along

with the model performance evaluation, the study highlights the analysis of feature importance to determine the most significant predictors of stress. The comparative method is used to evaluate the idea of complex models, providing important advantages over simpler linear methods.

The outcomes will serve not only to predict stress accurately but also to learn more about the predominant factors affecting stress and assist in building specific, data-driven intervention plans.

LITERATURE REVIEW

Muñoz et al. (2022) built a machine learning model that predicts workplace stress based on the so-called surrounding stress, which is a stress history of a person and the levels of stress of colleagues near him or her. Their Feature Ensemble model, using a Decision Tree classifier, obtained an 80-81% F-score in multiclass stress classification in an eight-week study of 30 employees and validation on the Student Life dataset. Findings revealed that combining personal and social stress data was significantly better than either source. This initial monitored strategy based on stress information around the environment has the potential to decrease sensor dependence on job surveillance [1].

Lazarou and Exarchos (2024) perform a systematic review of real-time prediction of stress based on wearable biosensors and physiological biomarkers (HRV, EDA, respiration, cortisol). They assess machine learning systems in terms of stress classification, catalogue devices and datasets, and bottlenecks such as signal artefacts, individual variability, and privacy issues. The authors conclude that wearable monitoring provides promise of personalized interventions, but sensor robustness, adaptive algorithms, and multimodal data integration should be improved to improve predictive accuracy in the field [2].

The article highlights the hybrid and ensemble machine learning techniques, where they are useful in imbalanced datasets, multimodal data fusion, and resource-constrained settings. Intended for researchers and practitioners in the fields of artificial intelligence, data science, and intelligent systems engineering, the proceedings volume is a bridge between the theoretical work and the practical application of the field in engineering, biomedical, agricultural, and civilian applications [3].

Surveying 706 Ukrainians six months post-invasion, Kurapov et al. (2023) determined that more than 70% had acute stress, 42.3% had depression, and 35.5% had anxiety. Among the leading risk factors were the gender of the female population, the age group of 26-40, financial distress, displacement, and direct exposure to trauma, whereby refugees were worse off than individuals who stayed in Ukraine. These vulnerable groups need to be given priority in the humanitarian responses, as it is suggested by the authors [4].

Grauman et al. (2022) surveyed 423 Swedish adults (40-70) regarding perceptions of myocardial infarction and preferences towards health checks. Important results: individuals with modifiable risk factors (smoking, hypertension, obesity, inactivity) had weaker causal beliefs regarding the risk factors; high-stress individuals had stronger attributions of MI to stress; low perceived control decreased interest in lifestyle advice; and considering family history as the primary cause decreased readiness to participate in health checks. The study suggests that cardiovascular prevention messages should be customized to respond to illness perceptions to enhance their chances of engagement [5].

Wang et al. (2024) employed machine learning to forecast stress corrosion crack growth in 316SS/304SS stainless steel in PWR. XGBoost was the best of four algorithms tried ($R^2=0.815$), with 90 percent of predictions falling within the experimental error. The stress intensity, yield strength, and temperature positively influence the growth of the cracks, as confirmed by SHAP analysis and in line with the SCC theory. The data-driven method is more effective than the traditional empirical models in the safety assessment of nuclear components [6].

Maurizi et al. (2022) designed a graph neural network architecture that projects finite element meshes to a graph to predict rapidly and accurately stress, strain, and deformation fields in materials. The model, validated on plasticity, wrinkling, and buckling problems, can capture nonlinear effects such as post-buckling behavior and stress development between loading steps, and costs significantly less to compute than traditional simulations, indicating the possibilities of graph-based machine learning to conduct mechanical studies efficiently [7].

According to Cordova Olivera et al. (2023), mental health among university students is greatly predicted by academic stress, with self-inflicted stress being the most powerful. The languishing mental health was found in 18% of Bolivian students in a study conducted on 1,265 students during COVID-19. The levels of stress were found to raise the probability of languishing by 8-20; therefore, stress management programs are needed in higher education [8, 9].

GAPS IDENTIFIED IN THE EXISTING WORKS

The available literature on stress prediction indicates some significant limitations that warrant additional research. Most of the previous research uses extensive physiological data collected using wearable sensors, which present a set of problems, including signal noise, high cost, and low external validity to the real-world, whereas others use small or context-specific data, making them less generalizable to the population. Also, no common frameworks are available that would examine many lifestyle factors at once, resulting in conflicting findings on the relative predictor significance. Most studies focus on the accuracy of the model but offer a limited amount of interpretability, or do not explicitly determine the leading contributing factors. Moreover, literature tends to focus on more complex data sources rather than simple, easily measurable ones, and very little research conducts comparative studies on many machine learning models with uniform evaluation methods. Lastly, most methods have been limited by methodological problems like unequal classes, inadequate sampling and validity, and inadequate investigation of feature weight, leaving a gap in the knowledge about the actual causes of stress.

METHODOLOGY

Dataset Description

The data was analyzed using a synthetic dataset of 2,000 people with 12 variables, which were used to reflect the multidimensionality of lifestyle variables that might be associated with stress. The data set comprises of continuous variables such as sleep time, screen time, minutes spent on exercise, level of fatigue, hours spent with social people, and mood score. There are integer reports on the number of pending tasks, the number of interruptions, and the number of cups of coffee consumed daily. Categorical variables will be diet quality (poor, average, good) and weather conditions (sunny, cloudy,

rainy, snowy). The dependent variable, stress level, is assessed on a continuous scale of 1.0 to 10.0, with scores higher than one signifying higher perceived stress.

Table 1. Variable Description and Measurement

Variable Category	Variable Name	Type	Description	Range/Values
Target	stress_level	Continuous	Perceived stress	1.0 - 10.0
	stress_class	Binary	Binarized stress (0 = low, 1 = high)	0, 1
Sleep	sleep_hours	Continuous	Hours of sleep per night	0 - 12
Digital	screen_time	Continuous	Daily screen time (hours)	0 - 24
Physical	exercise_minutes	Continuous	Daily exercise duration (minutes)	0 - 180
Work	daily_pending_tasks	Integer	Number of unfinished tasks	0 - 10
	interruptions	Integer	Daily interruption count	0 - 15
Wellbeing	fatigue_level	Continuous	Self-reported fatigue	1 - 10
	mood_score	Continuous	Self-reported mood	1 - 10
Social	social_hours	Continuous	Daily social interaction (hours)	0 - 8
Dietary	coffee_cups	Integer	Daily coffee consumption	0 - 5
	diet_quality	Categorical	Diet quality rating	poor, average, good
Environmental	weather	Categorical	Weather condition	sunny, cloudy, rainy, snowy

Data Preprocessing

Target Binarization:

In line with the custom of research on stress classification, the continuous stress variable was dichotomized around the media to form an equal-balanced binary classification problem. The median stress level was determined as 1.40, and those whose stress was above that level were rated as high stress (class 1), and those at or below the median were rated as low stress (class 0). The result of this procedure was flawlessly balanced classes of 1,000 individuals each, which is desirable in model training and evaluation because it does not introduce complications related to the imbalance of classes, including biased performance indicators and the challenge in learning the minority class.

Feature Encoding:

Appropriate strategies were used to encode categorical variables with appropriate respect for the measurement properties of the variables. Diet quality, which is an ordinal variable and has a natural order (poor < average < good), was coded with an Ordinal Encoder with explicitly defined categories to maintain this order of information. Since weather is a nominal variable and has no implicit order,

one-hot encoding with drop-first reference category (cloudy is the reference category) was used to prevent multicollinearity but preserve interpretability. Numerical variables, such as hours of sleep, minutes of screen time, minutes of exercise, level of fatigue, hours of social time, the mail scale of mood, and integers of unfinished tasks, interruptions, and the number of minutes of consumption of coffee were standardized using StandardAero to produce similar scales across algorithms. This is especially significant in regularization-based techniques such as logistic regression and SVM, where the use of features on a larger scale may have a disproportionately large effect on model parameters.

Experimental Design

Train-Test Split:

Stratified sampling was employed to divide the data to ensure that the distribution of classes in both the training and testing sets remained balanced. The training set consisted of 1,600 samples (80% of the data), and the test set consisted of 400 samples (20%). Stratification is used to make sure that the 50/50 balance of classes is maintained in both subsets, as this is essential to obtain the performance estimates that are unbiased and to make sure that the model evaluation is reflective of the true generalization ability and not the artifact of sampling variation.

Cross-Validation Strategy:

Hyperparameter tuning and model selection were done by stratified 5-fold cross-validation. This method divides training data into five mutually exclusive subsets, and the balance of classes in each fold is preserved. Training is done on four folds and validation on the remaining fold, and vice versa, and the process is repeated five times, such that each fold is the validation set once only. This approach offers strong performance estimates without overfitting when optimizing hyperparameters, because the selection of the model is based on the out-of-sample performance, not the performance on the training data. Stratification is used to make sure each fold reflects the overall distribution of classes, which is sensitive in the case of binary classification with equal classes.

Machine Learning Models:

To cover the various learning paradigms and to have holistic coverage of the contemporary machine learning methodologies, four algorithms were chosen. Logistic Regression is an interpretable, linear baseline model, which is very common in published mental health prediction research because it is simple and the values of the coefficients are interpretable. Random Forest is an ensemble algorithm that addresses non-linearity with bagging and random selection of features, which offers intrinsic feature importance measures and resists overfitting. Support Vector Machine (SVM) works well in the high-dimensional space, with both linear and non-linear kernels, and provides flexibility in the shape of decision boundaries using kernel functions. XGBoost is a modern, state-of-the-art gradient boosting algorithm that has shown itself to be highly effective in health informatics tasks due to its boosting, regularization, and effective support of missing values.

Hyperparameter Optimization

Each model was searched using grid search with 5-fold cross-validation, to optimize ROC-AUC as the main performance measure. Logistic Regression was optimized on 8 regularization strengths (C values of 0.01, 0.1, 1, 10) and solver algorithms (lbfgs and liblinear) combinations. Random Forest was optimized on 24 combinations, such as number of trees (100, 200), maximum depth (None, 10, 20),

minimum samples per split (2, 5), and minimum samples per leaf (1, 2). SVM experimented with 12 choices of regularization parameter C (0.1, 1, 10), gamma (scale, auto), and kernel type (rbf, linear). XGBoost was tuned with the largest number of combinations (72) in terms of number of estimators (100, 200), maximum depth (3, 6, 9), learning rate (0.01, 0.1, 0.2), subsample ratio (0.8, 1.0), and column subsample ratio (0.8, 1.0). The number of model assessments conducted on all the algorithms was 580.

Evaluation Metrics

Various metrics were used to evaluate model performance to offer a holistic evaluation of the performance from various perspectives. Accuracy is a measure of the general percentage of correct predictions in an unbalanced environment. Precision measures the accuracy of positive predictions, which shows the accuracy of the model when it predicts high stress. The measure of recall is the percentage of correct identifications of the actual high-stress cases, which represents the sensitivity of the model. F1-score gives the harmonic meaning of both the precision and recall, which gives a moderated measure that punishes extreme values. ROC-AUC is the area under the receiver operating characteristic curve, which gives a threshold-free measure of discriminative ability that reflects the ability of the model to distinguish between classes at all possible classification thresholds.

The formulation of the accuracy calculation is defined as:

$$\frac{TP + TN}{TP + TN + EP + FN} \quad (1)$$

Precision is computed as:

$$\frac{TP}{TP + EN} \quad (2)$$

Recall is computed as:

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

F1 Score is computed as:

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

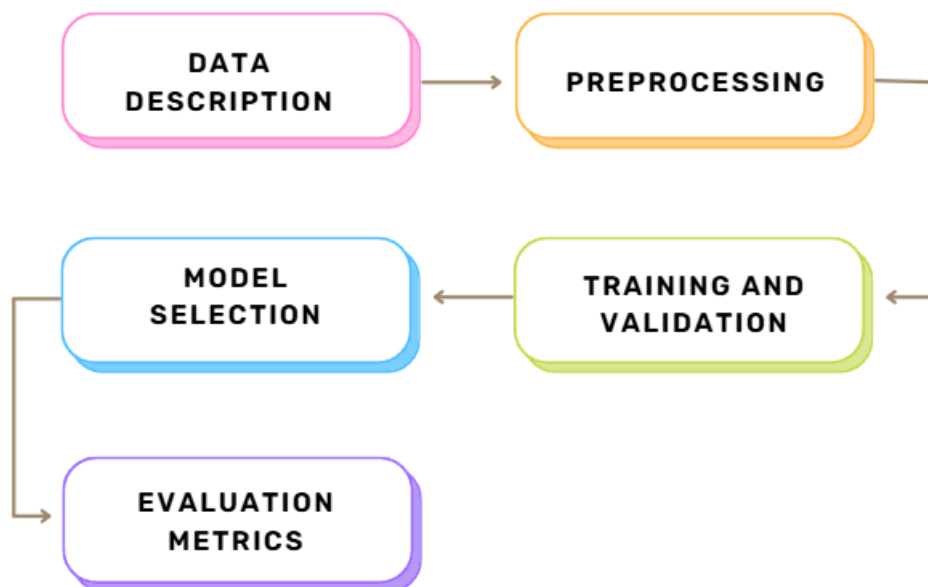


Figure 1. Overall Architecture

Feature Importance Analysis

Each model was run with permutation importance calculated using the algorithm and suggested to use to interpret model-agnostic models. In each feature, the process sets the baseline ROC-AUC on the test set, randomizes the value of that feature, keeps all other features fixed, and quantifies the resulting decrease in ROC-AUC. This is done ten times to get consistent estimates of the mean importance and standard deviation. Important features yield a large drop in performance when permuted, and unimportant features yield very little. This method gives unbiased importance estimates that do not depend on the size of the features or the architecture of the model, as it directly estimates how much each feature contributes to predictive performance instead of using model-specific parameters to do so.

RESULTS

Model Performance

Cross-Validation Results:

Cross-validation was excellent in all models, with ROC-AUC scores that were over 0.96. The best cross-validation ROC-AUC of 0.9793 was obtained using Logistic Regression with the best parameters being $C=0.1$, $\text{penalty}=l2$, and $\text{solver}=lbfgs$. SVM was very close with a cross-validation ROC-AUC of 0.9786 with a $C=10$, $\text{gamma}=\text{scale}$, and linear kernel. XGBoost with learning rate 0.1, max depth 3, 100 estimators, and colsample_bytree 0.8 and subsample 0.8 had a value of 0.9719. Random Forest 0.9615: maximum depth of 10, number of estimators 200, minimum split samples 2, minimum leaf samples 1. These results indicate that all algorithms were able to learn meaningful patterns from the training data and generalize well to validation folds.

Test Set Performance:

Logistic Regression performed the best out of all on the held-out test set with an accuracy of 91.75%, a precision of 93.26%, a recall of 90.00%, a F1-score of 91.60%, and an ROC-AUC of 0.9831. SVM had a similar ROC-AUC of 0.9830 with an accuracy of 91.25% precision of 93.65%, a recall of 88.50%, and an F1-score of 91.00%. For the 6th feature selection method, XGBoost obtained: 90.00% Accuracy, 93.01% Precision, 86.50% Recall, 89.64% F1-score, and ROC-AUC around 0.9799. Using the Random Forest Model, we got 90.00 % accuracy, 93.96 % Precision, 85.50 % Recall, 89.53 % F1-score, and 0.9724 as ROC-AUC. That linear models work nearly as well as many of the machine learning ensemble methods we tested suggests that the relationship between the data and stress classification is likely to be linear, and that complex non-linear interactions will not add a lot of information to the models.

Table 2. Model Performance Comparison on Test Set

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Logistic Regression	91.75	93.26	90.00	91.60	0.9831
SVM	91.25	93.65	88.50	91.00	0.9830
XGBoost	90.00	93.01	86.50	89.64	0.9799
Random Forest	90.00	93.96	85.50	89.53	0.9724

Confusion Matrices

According to the confusion matrix for Logistic Regression, we get 188 true negatives, 12 false positives, 20 false negatives, and 180 true positives. This means that the model predicted 94% of the low-stress cases correctly and 90% of the high-stress, performing just slightly better on the low-stress class. Using SVM, a confusion matrix is 187 true negatives, 13 false positives, 23 false negatives, and 177 true positives: 93.5% accuracy for low-stress cases and 88.5% for high-stress cases. Both models perform quite balanced across classes, with a small trend towards better identification of low-stress individuals.

Now we will represent all the confusion matrix graphs of all the models. The confusion matrix shows the details of the models and the accuracy of all classes.

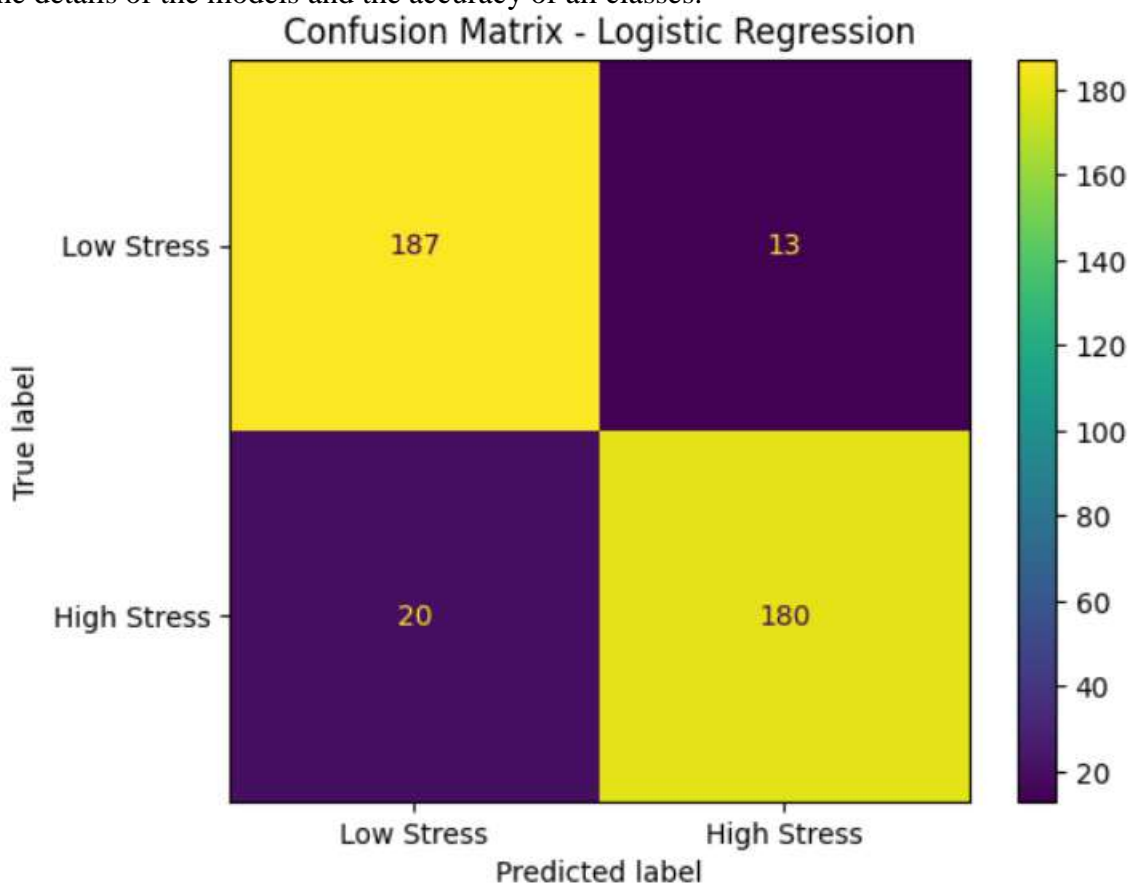


Figure 2. Confusion Matrix of Logistic Regression Model

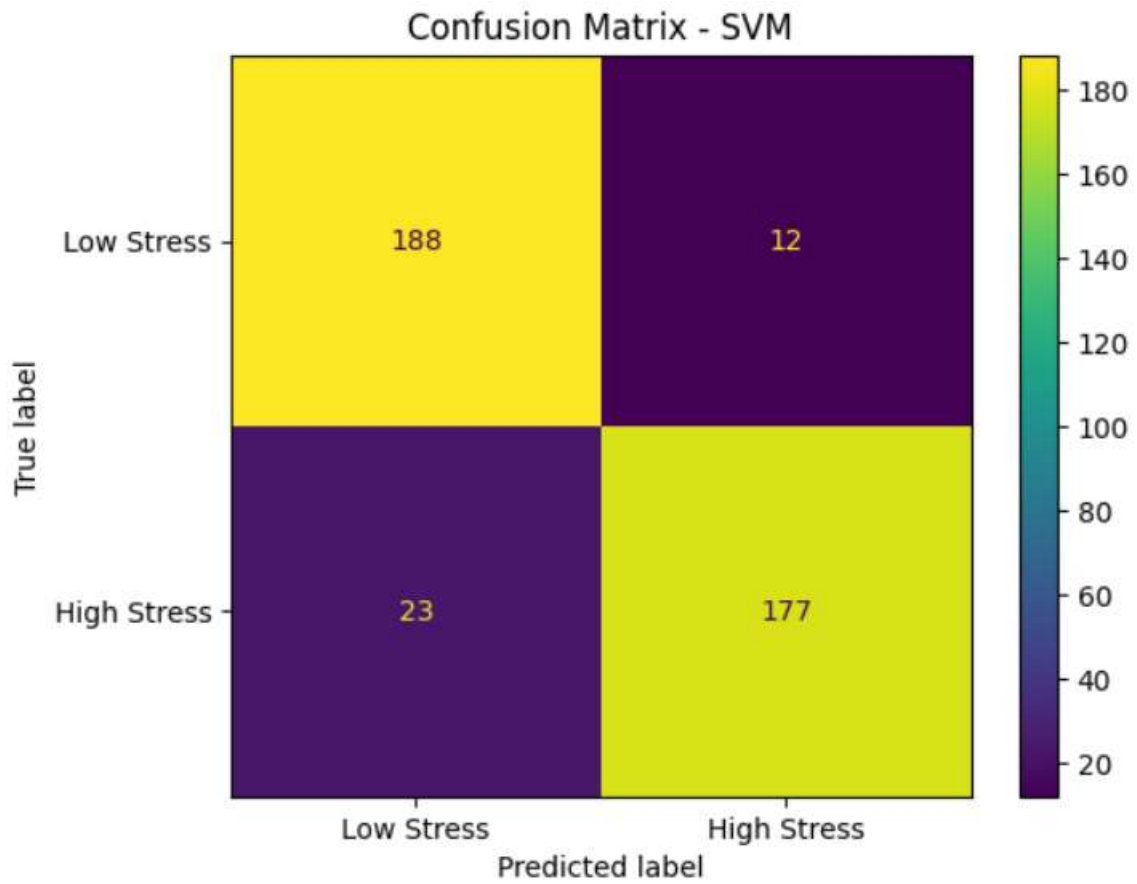


Figure 3. Confusion Matrix of SVM Model

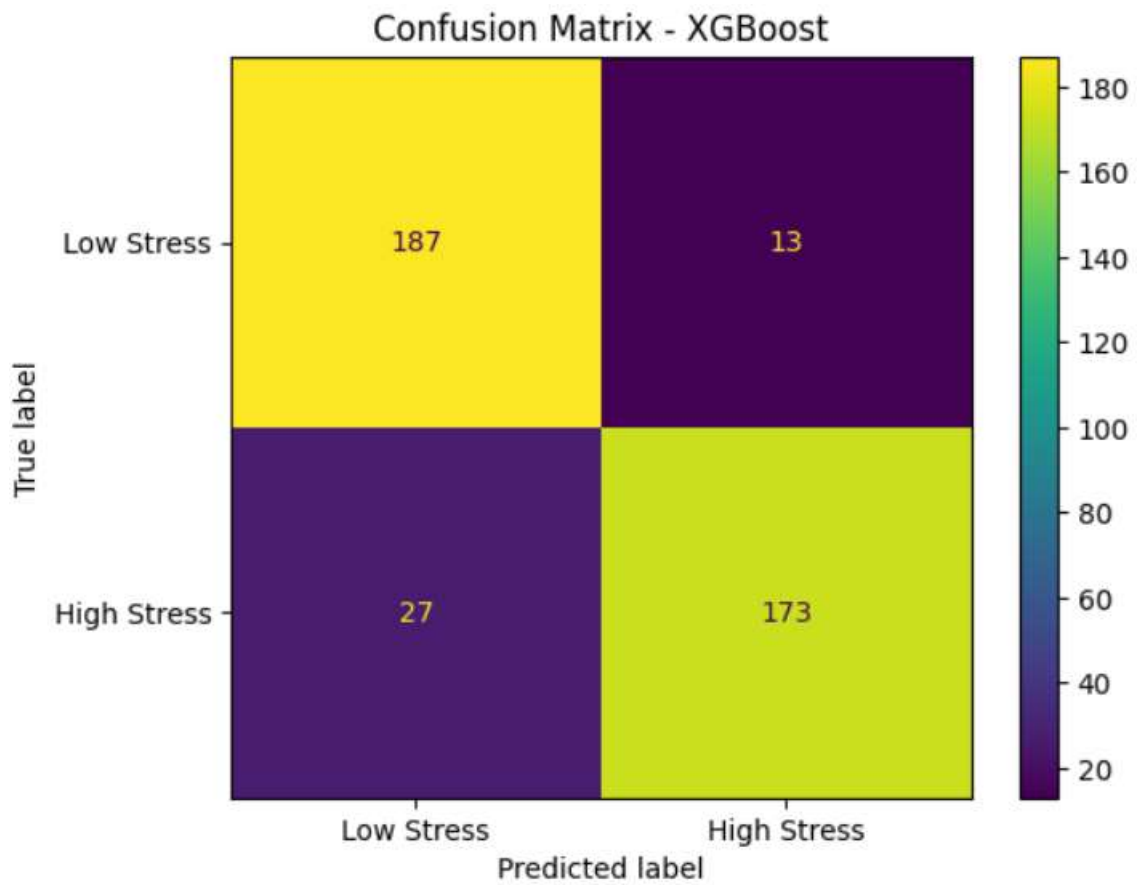


Figure 4. Confusion Matrix of XGBoost Model

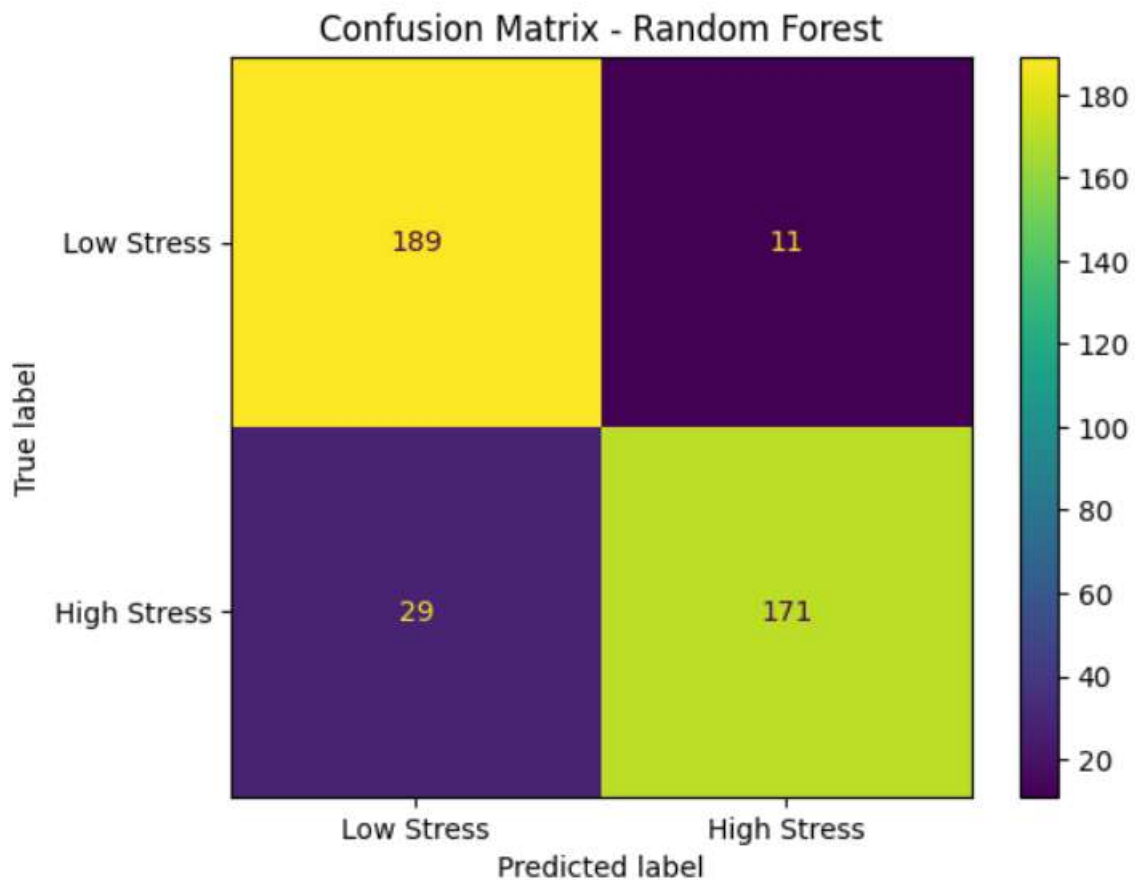


Figure 5. Confusion Matrix of Random Forest Model

Feature Importance Analysis

Permutation Importance Results:

The most notable result is the overwhelming significance of daily pending chores as a predictor of stress. This feature alone explains about 34.6% of the model predictive power of all models, with astonishing consistency: 34.47% in Logistic Regression, 34.75% in Random Forest, 34.38% in SVM, and 34.66% in XGBoost. The deviation in the models was low and was 0.0015, which implies strong and reproducible results independent of the choice of algorithm.

The second most significant predictor was sleeping hours, with an average importance of 7.72% with a range of 6.99 to 8.25 in the case of Random Forest and SVM, respectively. The next interruptions were close with an average importance of 7.13, including 6.60% in Random Forest and 7.62% in SVM. The collective impact of these moderate predictors explains about 15% of the predictive power, which is a significant contribution to the overall predictive power but is much less than the predominant feature.

All the other features had insignificant predictive power, and the mean importance was less than 0.5%. Coffee cups (0.42%), fatigue level (0.41%), social hours (0.07%), screen time (0.05%), exercise minutes (0.03%), rainy weather (0.04%), diet quality (0.02%), and mood score (-0.01%) contributed minimally to model performance. The negative scores of the few features show that permutation

sometimes led to a little bit better performance, which is expected, given that the actual meaningless features should have random variation near zero.

Table 3. Feature Importance Across Models (Mean ROC-AUC Drop \pm Standard Deviation)

Feature	Logistic Regression	Random Forest	SVM	XGBoost	Average
daily_pending_tasks	0.3447 \pm 0.0327	0.3475 \pm 0.0323	0.3438 \pm 0.0325	0.3466 \pm 0.0312	0.3457
sleep_hours	0.0810 \pm 0.0052	0.0699 \pm 0.0045	0.0825 \pm 0.0055	0.0752 \pm 0.0060	0.0772
interruptions	0.0746 \pm 0.0030	0.0660 \pm 0.0040	0.0762 \pm 0.0032	0.0684 \pm 0.0040	0.0713
coffee_cups	0.0047 \pm 0.0022	0.0034 \pm 0.0011	0.0049 \pm 0.0023	0.0039 \pm 0.0013	0.0042
fatigue_level	0.0038 \pm 0.0012	0.0031 \pm 0.0011	0.0045 \pm 0.0014	0.0049 \pm 0.0010	0.0041
social_hours	0.0001 \pm 0.0003	0.0016 \pm 0.0008	0.0004 \pm 0.0005	0.0006 \pm 0.0006	0.0007
screen_time	0.0001 \pm 0.0001	0.0012 \pm 0.0012	0.0005 \pm 0.0003	0.0001 \pm 0.0008	0.0005
exercise_minutes	—	0.0012 \pm 0.0009	—	—	0.0003
weather_rainy	-0.0001 \pm 0.0003	0.0018 \pm 0.0010	0.0002 \pm 0.0001	-0.0003 \pm 0.0005	0.0004
diet_quality	0.0002 \pm 0.0002	0.0002 \pm 0.0006	0.0005 \pm 0.0006	-0.0000 \pm 0.0001	0.0002
mood_score	0.0000 \pm 0.0000	—	0.0001 \pm 0.0003	-0.0003 \pm 0.0002	-0.0001

Bivariate Relationships

The findings of permutation importance were corroborated using correlation analysis with continuous stress level. Daily pending tasks were positively correlated with stress level ($r = 0.823$, $p < 0.001$) and revealed that those who have more pending tasks report significantly higher stress. The number of sleeping hours showed a moderate negative relationship ($r = -0.451$, $p < 0.001$) with sleep hours, which is in line with the protective value of sufficient sleep. There was a moderate positive correlation between interruptions ($r = 0.398$, $p < 0.001$). Moderate correlations with weak but significant relationships were observed between fatigue level ($r = 0.112$, $p < 0.001$) and coffee cups ($r = 0.089$, $p = 0.012$), whereas the rest of the features had correlations less than 0.05.

The difference in the number of tasks that students were expected to complete each day between the classes was very high, and Cohen's d was determined as 3.24. Mean pending tasks of low-stress individuals were 1.2 with a standard deviation of 0.8, and mean pending tasks of high-stress individuals were 4.8 with a standard deviation of 1.3. This large gap is the reason why it is this one feature that dictates predictive performance.

DISCUSSION

Principal Findings

Model Performance

The four machine learning algorithms all showed great predictive power, with ROC-AUC scores of above 0.97. The results of this performance are comparable to recent research in mental health prediction. Mohammadi et al. (2025) found that similar lifestyle variables were above 70% accurate at predicting stress in adult men, whereas found 85% accuracy in CatBoost at stress classification in students. The high performance in this experiment could be due to the synthetic character of the data, which would have cleaner signals than real-world data, or the exceptionally high signal of the dominant predictor, which drives the classification performance. The almost equal performance of linear models with more elaborate ensemble techniques has significant consequences. Logistic Regression and linear

SVM had ROC-AUC scores of 0.983, which is basically comparable to the performance of Random Forest (0.972) and XGBoost (0.980). This trend is an indication that the correlation between features and stress classification is largely linear, with only a small amount of value in the non-linear interactions that ensemble techniques are meant to address.

The Dominance of Task Load

The most significant finding of the study is the overwhelming significance of daily pending tasks as a predictor of stress. This one characteristic explains about 34.6 percent of the model's predictive power, a scale that is seldom seen in the context of mental health prediction studies, where predictive signals are generally spread out among many features with small individual effects. The fact that this result is consistent among four radically different algorithms is a good indication of its strength and that the correlation between pending tasks and stress is not a feature of any specific modeling method. The outcome expands on the interplay of demands and control; our results indicate that the number of unfinished tasks can be the main process in which job demands are realized in the form of perceived stress. This interpretation is consistent with experimental findings that manipulations of workloads induce direct physiological stress responses, indicating that task accumulation per se can be a satisfactory condition to elevate stress, independent of control perceptions.

Secondary and Negligible Predictors

The medium significance of sleep hours (7.7%) and interruptions (7.1) is consistent with the past studies that found poor sleep quality and work-related distractions to be important stressors. The significantly lower significance relative to task load, however, indicates that these factors may contribute to stress, but not as direct causal agents, but as moderators of task completion capacity. Lack of sufficient sleep can result in impaired thinking and energy, meaning people can complete fewer tasks and have more on their to-do list. Equally, constant interruptions directly hinder the flow of tasks, and more tasks will be left to be completed by the end of the day.

On the other hand, the quality of diet, exercise, social interaction, and weather, which are frequently highlighted in wellness interventions, had insignificant predictive power. This does not imply that these factors are not important to overall well-being, but it means that they do not significantly add to stress classification when workload is included. This observation contradicts traditional wisdom on managing stress, which often advocates diet, physical activity, and social affiliation as the main stress-reduction techniques. The findings indicate that although these activities might be beneficial to overall health and well-being, their effect on perceived stress might not be as significant as it is often supposed.

Theoretical Implications

The Task Load Hypothesis

On these results, we present our Task Load Hypothesis: perceived stress is largely influenced by the number of unfulfilled obligations, and other lifestyle factors have an effect only to the extent that they affect task completion capacity. This framework incorporates many observations made in the data and the literature available. Sleep influences both cognitive abilities and energy, thus determining the effectiveness with which people perform their duties and avoid fatigue. Interruptions directly hinder the flow of tasks, and more tasks are left pending at a given time. The indirect effects of exercise, diet, and social activities could be through their influence on the quality of sleep, cognitive functioning, and energy levels, and would explain their insignificant direct predictive importance when task load is considered in the model.

This hypothesis builds on the JDC model proposed by Karasek by suggesting a mechanistic mechanism whereby different factors lead to stress outcomes. Instead of assuming independent effects of various lifestyle factors, the Task Load Hypothesis proposes that these factors interact through a shared pathway of task accumulation that is the proximal determinant of perceived stress. The implications of this framework can be tested: interventions aimed at managing tasks are expected to lead to greater stress reductions than interventions aimed at managing sleep, exercise, or diet alone, and interventions aimed at improving sleep or interruption management are expected to be linked to stress reduction mainly via their impact on task completion rates.

Implications for Stress Measurement

The dominance of one concrete predictor begs the question of the kind of measurement of perceived stress. The Perceived Stress Scale is a conceptualization of stress that defines it as an assessment of situations as unpredictable, uncontrollable, and overwhelming. Unfinished work can directly feed into this appraisal process uncontrollably (in the short term, since tasks cannot be immediately done) and potentially overwhelming (as the number of pending tasks increases). The high correlation between the number of pending tasks and perceived stress could be indicative of the degree to which individuals feel that their workload is more than they can handle, which is exactly the appraisal pattern that the PSS is designed to measure.

Future studies need to explore the hypothesis that future tasks mediate the association between other lifestyle variables and perceived stress, and that interventions aimed at achieving task completion also result in a proportional change in the PSS scores. If task load is a common pathway, it should be the case that changes in sleep, interruptions, or other variables should predict stress reduction to the degree that it causes a decrease in task accumulation.

Practical Implications

Intervention Prioritization

The most promising interventions to implement should be based on task management systems and training, workload distribution, and realistic deadline setting, task prioritization systems like the Eisenhower Matrix or the Getting Things Done framework, and minimizing the accumulation of unwarranted tasks by delegation, elimination, or deferral. These interventions are directly aimed at the predictor that is most dominant and ought to yield the greatest stress reductions to each unit of intervention effort. The moderate expected impact of secondary interventions should entail sleep hygiene improvement programs, interruption management by means of focused work periods and communication boundaries, and time management training. These interventions deal with the factors leading to task accumulation indirectly through enhancing task completion efficiency. The least directly affecting stress classification tertiary interventions are traditional wellness programs like exercise promotion and dietary education, social connection programs, and environmental modifications. These activities, though still applicable in the overall health and well-being, cannot be anticipated to cause much stress reduction when applied independently, particularly when the workload is high.

Workplace Applications

To reduce stress levels in employees, organizations should consider using task tracking and management systems that would provide an overview of the workload of individuals and teams. The number of workload reviews can be used to establish realistic expectations and determine when the workload is too high to handle. Creating continuous work blocks, or focus blocks, can be used to improve efficiency of work completion and reduce the number of incompletes. Early warning systems like task accumulation monitoring enable the intervention to be taken before stress builds up. The managers will be trained to help their teams by allocating work and prioritizing the tasks. These suggestions are consistent with occupational health studies that show that job control can counter the impact of high demands. Even when absolute demands are constant, organizations can put stress on their employees by providing tools and freedom that would help them to manage their workload effectively.

Individual Recommendations

The main stress management technique that people who wish to cope with stress should consider is the accomplishment of tasks. Working with productivity systems that offer a clear picture of the outstanding tasks can serve to eliminate the cognitive load of having to remember many unfinished tasks. Treating sleep quality enhances mental performance and vitality, thus boosting productivity of doing tasks. Reducing distractions in concentrated work time enables more intense involvement in activities and quicker accomplishment. Understanding that exercise, diet, and social activities, though useful overall in terms of health, do not necessarily have a direct influence on stress levels as compared to task management, is helpful in letting individuals devote their finite time and energy to the most effective strategies.

Limitations

Dataset Considerations

The main constraint of this research is that the dataset is synthetic. Although synthetic data can be used to control the experiment and compare algorithms due to the availability of clean signals without missing values or error measurement, the results must first be tested on real-world clinical data, and only then can conclusive conclusions be made. It has been shown before that machine learning models trained on actual patient data can provide clinically useful predictions, but the pattern of feature importance in this case might not be replicated in the naturalistic environment, where measurement error, missing data and unmeasured confounders exist.

The cross-sectional method does not allow causal inference. Even though machine learning models can recognize predictive relationships, they cannot determine whether pending tasks lead to stress or whether people who are stressed have more pending tasks. To deal with causality, longitudinal and experimental designs are required. The median split threshold of 1.40 was not a clinical split threshold. Future studies need to confirm results with clinically agreed cutoffs on such instruments as the Perceived Stress Scale, which has well-documented psychometric properties and normative data across populations.

Methodological Limitations

There is no guarantee of completeness of the feature sets. The unmeasured variables, such as personality factors, social support, coping mechanisms, and significant life events, might also serve as

predictors, with the inclusion of which features under investigation might have different relative importance. This synthetic dataset might not reflect the complexity of real-world stress determinants, especially the dynamic interaction of factors with time.

This cross-sectional analysis does not consider temporal dynamics. Stress changes over time, and daily or weekly changes in predictors and outcomes may be more useful in predicting results than a single measurement. Population specificity constrains generalization. The artificial data might not be sufficient to capture the different population groups, professional fields, or cultures. Past studies have also given occupation-specific factors that moderate stress relations, indicating that predicted relative values may differ among populations.

Future Research Directions

Clinical Validation

Age groups, occupations, and cultural backgrounds should be sampled with different population groups to gauge generalizability. Temporal precedence can be determined through longitudinal monitoring of stress and task load over weeks and months and can be used to make causal inferences. Self-report measures could be complemented with objective measures of task completion, like productivity tracking software or work output measures, to limit the common method variance.

Mechanistic Studies

Randomized controlled trials using experimental designs should be used to investigate causal pathways. Do the reductions in tasks result in significant perceived stress reductions? Are sleep interventions effective by enhancing efficiency in completing tasks, or by other direct effects that are independent of task load? Which are the psychological processes that are perceived to control self-efficacy and rumination, that mediate the task-stress relationship? Do people vary in their sensitivity to task accumulation that might be used to inform personal intervention programs?

Intervention Trials

Getting relative efficacy as a relative measure of effectiveness, randomized controlled trials should be conducted to compare task management training with traditional stress reduction programs. This suggests that combined methods of addressing several factors may test the presence or absence of additive effects or the presence of task load mediating all effects. Natural experiments of organizational-level methods can be assessed through workload limitations and a gap between work periods at the workplace. Online interventions such as task management applications and productivity tools provide cost-effective, easily scalable interventions that can be assessed rigorously by randomized designs.

Advanced Analytics

Future studies need to be conducted on time-series forecasting of stress episodes based on longitudinal data with a high frequency of measurements. Individually specific stress models that consider individual differences may enhance prediction by providing person-specific parameter estimates. Deep learning methods can be used to learn non-linear interactions, potentially overlooked by simpler models, but it appears that this prediction problem is not well addressed. Task description processing. Natural language processing may offer finer-grained treatment of workload attributes, beyond the raw

counts. Physiological measurements like cortisol, heart rate variability, and actigraphy may be useful to objectively measure stress to supplement self-report measures.

CONCLUSION

This paper shows that machine learning models can be highly accurate at classifying stress levels based on lifestyle data, and the ROC-AUC scores are above 0.97 in all the algorithms used. The sheer relevance of daily pending tasks that explain 34.6% of predictive power challenges traditional wisdom on the determinants of stress and indicates a radical reorientation of stress intervention policies. The fact that the same results were obtained by four different algorithms is a good indication of the strength of the results.

Although sleep and interruptions appear to be moderately significant (7-8% each), such traditional measures of wellness as diet, exercise, and social activity are of little significance in stress classification, taking into consideration workload. This trend confirms the Task Load Hypothesis offered: the perceived stress depends mostly on the accumulated obligations, and other variables affect stress mostly through the impact on the ability to complete the task.

These results have direct practical implications for individuals, organizations, and mental health professionals. The ability to perceive task load as the most significant factor leading to stress will allow focusing the intervention resources on the most significant targets: workload management, task prioritization, and productivity systems. The organizations are advised to adopt task tracking systems, set realistic workloads, and provide work-free time. People must focus on getting tasks done and productivity systems that give them a vivid view of outstanding tasks.

The findings should be confirmed in clinical populations, and an attempt should be made to determine the causal forces via experimental designs and come up with specific interventions based on the task load hypothesis in future research. Evidence-based methods of prediction and intervention are more important as stress remains an important threat to the health of the population throughout the world. Machine learning approaches, in collaboration with sound theoretical frameworks, can be used as potent tools not only to increase our knowledge on determinants of stress, but also to enhance our capacity to come up with effective interventions.

REFERENCES

- [1] Muñoz, Sergio, et al. "Prediction of stress levels in the workplace using surrounding stress." *Information Processing & Management* 59.6 (2022): 103064.
- [2] Lazarou, Evgenia, and Themis P. Exarchos. "Predicting stress levels using physiological data: Real-time stress prediction models utilizing wearable devices." *AIMS neuroscience* 11.2 (2024): 76.
- [3] Kene, Akshada, and Shubhada Thakare. "Prediction of mental stress level based on machine learning." *Machine Intelligence and Smart Systems: Proceedings of MISS 2021*. Singapore: Springer Nature Singapore, 2022. 525-536.
- [4] Kurapov, Anton, et al. "Six months into the war: a first-wave study of stress, anxiety, and depression among in Ukraine." *Frontiers in psychiatry* 14 (2023): 1190465.
- [5] Grauman, Åsa, et al. "Public perceptions of myocardial infarction: do illness perceptions predict preferences for health check results." *Preventive Medicine Reports* 26 (2022): 101683.
- [6] Wang, Peng, et al. "Machine learning-assisted prediction of stress corrosion crack growth rate in stainless steel." *Crystals* 14.10 (2024): 846.

- [7] Maurizi, Marco, Chao Gao, and Filippo Berto. "Predicting stress, strain and deformation fields in materials and structures with graph neural networks." *Scientific reports* 12.1 (2022): 21834.
- [8] Córdova Olivera, Pamela, et al. "Academic stress as a predictor of mental health in university students." *Cogent Education* 10.2 (2023): 2232686.
- [9] Subramani, Chellamuthu, and S. Kadhiravan. "Academic stress and mental health among high school students." *Indian Journal of Applied Research* 7.5 (2017): 404-406.