International Journal of Artificial Intelligence in Medical Issues

Volume 3 Issue 2
ISSN 3025-4167
https://doi.org/10.56705/ijaimi.v3i2.363

Research Article

A Comparative Study of Machine Learning Models for Stress Level Classification Using Social Media and Lifestyle Data

Muh Ikbal Siami [1,*]; Aris Wahyu Murdiayanto [2]; Sumiyatun [3]
[1] Institut Teknologi dan Bisnis Stikom Ambon, Kota Ambon, Maluku, Indonesia, siami25@gmail.com
[2] Universitas Jenderal Achmad Yani Yogyakarta, Daerah Istimewa Yogyakarta 55293, Indonesia, ariswahyumurdiyanto@gmail.com
[3] Universitas Teknologi Digital Indonesia, aerah Istimewa Yogyakarta 55198, Indonesia, sumiyatun@utdi.ac.id

Correspondence should be addressed to Muh Ikbal Siami; siami25@gmail.com

Received 28 January 2026; Revised 06 February 2026; Accepted 12 May 2026; Published 30 May 2026

Abstract:
The increasing use of social media and digital platforms has raised concerns regarding its potential relationship with sleep patterns, lifestyle behaviors, productivity, and psychological well-being. Stress is a common health-related issue that may be influenced by daily behavioral patterns, including screen time, social media usage, sleep duration, physical activity, and work or study habits. This study aims to develop and evaluate machine learning models for predicting stress levels based on non-invasive digital behavior and lifestyle indicators. The dataset used in this study consisted of 11,000 records with three stress level categories: Low, Medium, and High. The predictor variables included age, daily screen time, social media usage duration, sleep hours, exercise duration, study or work hours, productivity score, and the most frequently used social media platform. Several machine learning algorithms were evaluated, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Gradient Boosting. Model performance was assessed using accuracy, precision, recall, F1-score, confusion matrix analysis, and 5-fold stratified cross-validation. The experimental results showed that the overall classification performance was modest. The Decision Tree model achieved the best testing performance with an accuracy and macro F1-score of 0.3400, while Gradient Boosting achieved the highest cross-validation performance with a mean accuracy of 0.3480 and a mean macro F1-score of 0.3467. Feature importance analysis using Random Forest indicated that productivity score, sleep hours, study or work hours, social media hours, and daily screen time were the most influential variables. These findings suggest that digital behavior and lifestyle indicators may provide useful exploratory insights for stress-related analysis, although their predictive power remains limited. Therefore, the proposed approach is more suitable as an exploratory digital well-being assessment framework rather than a clinical diagnostic tool.

Keywords: Stress Prediction, Machine Learning, Social Media Usage, Sleep Patterns, Digital Well-Being, Health Informatics, Feature Importance.

Dataset link: -

1. Introduction

The rapid growth of digital technology and social media has significantly transformed human behavior, communication patterns, work habits, and lifestyle routines [1]. Social media platforms are now widely used not only for communication and entertainment, but also for learning, professional networking, and information sharing [2]. However, excessive digital engagement may influence several aspects of well-being, including sleep quality, physical activity, perceived stress, and daily productivity. In the context of digital health, understanding the relationship between social media use and psychological well-being has become increasingly important, particularly as mental health conditions and stress-related problems continue to affect individuals across different age groups [3], [4].

Stress is a common psychological response that may occur when individuals experience pressure related to work, study, health, social relationships, or daily responsibilities. Although stress can be considered a normal response in certain situations, prolonged or unmanaged stress may contribute to negative health outcomes and reduced quality of life. Therefore, early identification of stress levels is important to support preventive health strategies and digital well-being interventions [5], [6]. In recent years, artificial intelligence and machine learning have shown potential in supporting health informatics by identifying patterns from behavioral and lifestyle data that may not be easily observed through conventional analysis [7].

Previous studies in stress detection have commonly utilized physiological signals, wearable sensors, mobile sensing data, or clinical questionnaires [8]. While these approaches are useful, they may require specific devices, controlled data collection procedures, or self-reported clinical instruments [9], [10]. In contrast, behavioral indicators such as daily screen time, social media duration, sleep hours, exercise habits, work or study duration, and productivity scores can provide a non-invasive and accessible alternative for estimating stress risk [11], [12]. These variables are particularly relevant in the modern digital environment, where individuals' daily routines are closely connected with technology use and online activities [13].

Despite the growing interest in digital mental health, there is still a need for machine learning studies that examine how social media usage, sleep patterns, physical activity, and productivity-related variables can be used to predict stress levels. Many existing studies focus primarily on the negative effects of social media or screen time, while fewer studies integrate multiple behavioral and lifestyle indicators into a predictive model. Furthermore, model interpretability remains an important issue in health-related artificial intelligence because prediction results should be understandable and useful for researchers, practitioners, and decision-makers.

This study aims to develop and evaluate machine learning models for predicting stress levels based on social media usage, sleep patterns, lifestyle habits, and productivity indicators. The dataset used in this study consists of 11,000 records containing demographic information, daily screen time, social media usage duration, sleep duration, exercise time, study or work hours, productivity scores, stress levels, and the most frequently used social media platform. Several machine learning algorithms are compared to identify the best-performing model for stress level classification. In addition, feature importance analysis is applied to provide interpretability regarding the factors that contribute most to stress level prediction.

The main contributions of this study are as follows. First, this study proposes a machine learning-based approach for stress level prediction using non-invasive digital behavior and lifestyle data. Second, it compares several classification algorithms to evaluate their effectiveness in predicting low, medium, and high stress levels. Third, it provides an interpretable analysis of important behavioral factors associated with stress prediction [14]. The findings of this study are expected to contribute to AI-driven health informatics, particularly in the development of early stress risk screening and digital well-being assessment systems.

2. Method

This study employed a quantitative machine learning approach to predict stress levels based on social media usage, sleep patterns, lifestyle behaviors, and productivity indicators. The research workflow consisted of several stages, including dataset preparation, exploratory data analysis, data preprocessing, model development, model evaluation, and feature importance analysis. The overall methodological process was designed to identify the most effective classification model and to interpret the contribution of each behavioral variable to stress level prediction [15].

Dataset Description:

The dataset used in this study contains 11,000 records representing individual behavioral patterns related to digital media usage, sleep habits, exercise routines, work or study duration, productivity, and stress levels. Each record consists of demographic and lifestyle-related variables that can be used to analyze digital well-being and behavioral health patterns [16].

The target variable in this study is stress_level, which consists of three categorical classes: Low, Medium, and High. These classes represent the level of stress experienced by each individual. The predictor variables include age, daily screen time, social media usage duration, sleep duration, exercise duration, study or work duration, productivity score, and the most frequently used social media platform [17].

The description of each variable is presented in **Table 1**.

Table 1. Dataset Variables

Variable	**Description**
user_id	Unique identifier for each individual
age	Age of the individual
daily_screen_time_hours	Total daily screen time in hours
social_media_hours	Time spent on social media platforms per day
sleep_hours	Average daily sleep duration
exercise_minutes	Daily exercise duration in minutes
study_work_hours	Hours spent on study or work activities
productivity_score	Productivity score on a scale of 1–100
stress_level	Stress level category: Low, Medium, or High
platform	Most frequently used social media platform

The `user_id` variable was removed from the modeling process because it only serves as an identifier and does not provide predictive information. The remaining variables were used as input features for the classification models.

Research Workflow:

The research procedure followed a structured machine learning pipeline. First, the dataset was loaded and examined to identify the number of records, variables, missing values, duplicate data, and class distribution. Second, exploratory data analysis was conducted to understand the distribution of numerical and categorical variables. Third, data preprocessing was performed to prepare the dataset for machine learning algorithms. Fourth, several classification models were trained and evaluated using the same data partition. Finally, feature importance analysis was performed to interpret the most influential variables in stress prediction [18].

Data Preprocessing:

Data preprocessing was conducted to transform the raw dataset into a suitable format for machine learning models. The preprocessing stage included data cleaning, feature selection, categorical encoding, target encoding, and feature scaling.

First, the dataset was inspected for missing values and duplicate records. The dataset did not contain missing values, allowing all available records to be used in the analysis. Duplicate records, if found, were removed to prevent redundancy in the training process. The `user_id` column was excluded because it is an identity variable and has no direct relationship with stress prediction.

Second, the categorical variable `platform` was transformed using one-hot encoding. This method converts categorical platform names into binary numerical features so that they can be processed by machine learning algorithms. The target variable `stress_level` was encoded into numerical labels representing the three stress categories.

Third, feature scaling was applied using standardization. Standardization transforms the feature values so that they have a mean of zero and a standard deviation of one. This step is particularly important for algorithms that are sensitive to feature scale, such as Logistic Regression, Support Vector Machine, and K-Nearest Neighbors.

Data Splitting:

The dataset was divided into training and testing sets using an 80:20 ratio. A total of 80% of the data was used for model training, while the remaining 20% was used for model testing. Stratified sampling was applied to preserve the proportion of each stress level class in both training and testing sets. This approach ensures that the distribution of Low, Medium, and High stress levels remains consistent across the training and testing subsets.

The data splitting process was performed using a fixed random state to ensure reproducibility. The training data were used to build the classification models, while the testing data were used to evaluate model performance on unseen data.

Machine Learning Models:

This study compared several machine learning algorithms to classify stress levels. The selected models represent different types of classification approaches, including linear models, tree-based models, distance-based models, margin-based models, and ensemble learning models [19].

The models used in this study were:

- Logistic Regression: Logistic Regression was used as a baseline classification model. Although it is commonly used for binary classification, it can also be extended to multi-class classification problems. This model provides a simple and interpretable benchmark for stress level prediction.
- Decision Tree: Decision Tree is a tree-based model that classifies data by splitting features based on decision rules. It is easy to interpret but may be prone to overfitting if not properly controlled.
- Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees to improve classification performance and reduce overfitting. It is also useful for identifying important features through feature importance scores.
- Support Vector Machine: Support Vector Machine is a classification algorithm that attempts to find an optimal decision boundary between classes. The radial basis function kernel was used to handle possible non-linear relationships among variables.
- K-Nearest Neighbors: K-Nearest Neighbors classifies data based on the majority class among the nearest neighboring samples. This model is sensitive to feature scaling, making standardization an important preprocessing step.
- Gradient Boosting: Gradient Boosting is an ensemble method that builds models sequentially, where each new model attempts to correct the errors of the previous model. It is often effective for structured tabular data.

The comparison of multiple models allows this study to identify which algorithm is most suitable for predicting stress levels using behavioral and lifestyle data.

Model Evaluation:

The performance of each model was evaluated using several classification metrics, including accuracy, precision, recall, and F1-score. These metrics were calculated using the testing dataset [20].

Accuracy measures the proportion of correctly classified samples among all samples. Precision measures how many predicted samples in a certain class are actually correct. Recall measures how many actual samples in a certain class are successfully identified by the model. F1-score is the harmonic mean of precision and recall and is useful for evaluating the balance between these two metrics.

Since the target variable consists of three classes, macro-average and weighted-average metrics were used. Macro-average treats all classes equally by calculating the average metric across classes, while weighted-average considers the number of samples in each class. In addition, a confusion matrix was generated for each model to provide a more detailed view of correct and incorrect predictions for each stress level category.

The evaluation metrics used in this study are defined as follows:

Accuracy	

Precision	
Recall	
F1-score	

where TP represents true positive, TN represents true negative, FP represents false positive, and FN represents false negative.

Cross-Validation:

In addition to the train-test evaluation, 5-fold stratified cross-validation was applied to assess model stability. In this procedure, the dataset was divided into five subsets while maintaining the proportion of each stress level class. Each model was trained and validated five times, with each subset used once as the validation set [21]. The average accuracy and macro F1-score were then calculated to evaluate the consistency of model performance.

The use of stratified cross-validation helps reduce bias caused by a single train-test split and provides a more reliable estimation of model generalization performance.

Feature Importance Analysis:

To improve model interpretability, feature importance analysis was conducted using the Random Forest model [22]. Feature importance provides information about the relative contribution of each predictor variable in the classification process. Variables with higher importance scores have greater influence on the model's prediction.

This analysis was used to identify which behavioral and lifestyle factors contributed most to stress level prediction. The examined features included daily screen time, social media usage duration, sleep duration, exercise duration, study or work hours, productivity score, age, and social media platform indicators.

The feature importance results were visualized using a bar chart to support interpretation and discussion. This step is important because health-related machine learning models should not only provide predictions but also offer understandable insights regarding the factors associated with stress levels.

Experimental Environment:

All experiments were conducted using the Python programming language. The main libraries used in this study included Pandas and NumPy for data processing, Matplotlib and Seaborn for data visualization, and Scikit-learn for machine learning model development and evaluation. The experiments were designed to be reproducible by using a fixed random state during data splitting and model training.

3. Result and Discussion

This section presents the experimental results of stress level prediction using several machine learning models. The evaluation includes dataset distribution, model performance comparison, cross-validation results, confusion matrix analysis, and feature importance interpretation. The objective of this analysis is not only to identify the best-performing model, but also to understand the extent to which social media usage, sleep patterns, lifestyle behaviors, and productivity indicators can be used to classify stress levels.

Dataset Characteristics:

The dataset consisted of 11,000 records with three stress level categories: Low, Medium, and High. The distribution of the target variable was relatively balanced, with 3,715 samples labeled as Low stress, 3,681 samples as Medium stress, and 3,604 samples as High stress. This balanced distribution indicates that the classification task was not strongly affected by class imbalance.

Table 2. Distribution of Stress Level Classes

Stress Level	Number of Samples
Low	3.715
Medium	3.681
High	3.604
Total	11

The dataset also included several numerical variables related to digital behavior and lifestyle patterns. The average age of individuals in the dataset was 32.56 years. The average daily screen time was 6.56 hours, while the average time spent on social media was 4.25 hours. The average sleep duration was 5.99 hours per day, and the average exercise duration was 59.47 minutes per day. The average productivity score was 50.62 on a scale of 1 to 100.

Table 3. Descriptive Statistics of Numerical Variables

Variable	Mean	Std. Dev.	Min	Max
Age	32.56	9.69	16.00	49.00
Daily screen time hours	6.56	3.18	1.00	12.00
Social media hours	4.25	2.16	0.50	8.00
Sleep hours	5.99	1.73	3.00	9.00
Exercise minutes	59.47	34.72	0.00	119.00
Study/work hours	5.54	2.59	1.00	10.00
Productivity score	50.62	28.84	1.00	99.99

The distribution of the most frequently used social media platforms was also relatively balanced. LinkedIn had the highest number of users, followed closely by Snapchat, YouTube, Twitter, and Instagram.

Table 4. Distribution of Social Media Platforms

Platform	Number of Samples

LinkedIn	2.239
Snapchat	2.235
YouTube	2.2
Twitter	2.183
Instagram	2.143

Overall, the dataset had a balanced target distribution and no missing values, making it suitable for direct machine learning experimentation. However, balanced class distribution does not necessarily guarantee high predictive performance, especially when the input variables do not have strong discriminative relationships with the target classes.

Model Performance on Testing Data:

Six machine learning models were evaluated using an 80:20 train-test split. The testing set consisted of 2,200 samples. The models were compared using accuracy, macro precision, macro recall, and macro F1-score. Macro-average metrics were emphasized because they treat each stress level class equally.

Table 5. Model Performance on Testing Data

Model	Accuracy	Precision Macro	Recall Macro	F1-Score Macro
Decision Tree	3,400	3,403	3,398	3,400
Logistic Regression	3,377	3,355	3,368	3,328
Gradient Boosting	3,286	3,277	3,282	3,271
K-Nearest Neighbors	3,309	3,326	3,317	3,259
Support Vector Machine	3,268	3,261	3,263	3,252
Random Forest	3,241	3,238	3,238	3,236

Based on the testing results, the Decision Tree model achieved the highest macro F1-score of 0.3400 and the highest accuracy of 0.3400. Logistic Regression ranked second, with an accuracy of 0.3377 and a macro F1-score of 0.3328. The remaining models produced macro F1-scores between 0.3236 and 0.3271.

Although Decision Tree obtained the best testing performance, the difference between models was very small. All models produced performance close to 0.33, which is approximately the expected baseline for a balanced three-class classification problem. This suggests that the available features may not contain sufficiently strong patterns to clearly distinguish Low, Medium, and High stress levels.

Figure 1 shows the comparison of machine learning models based on macro F1-score.

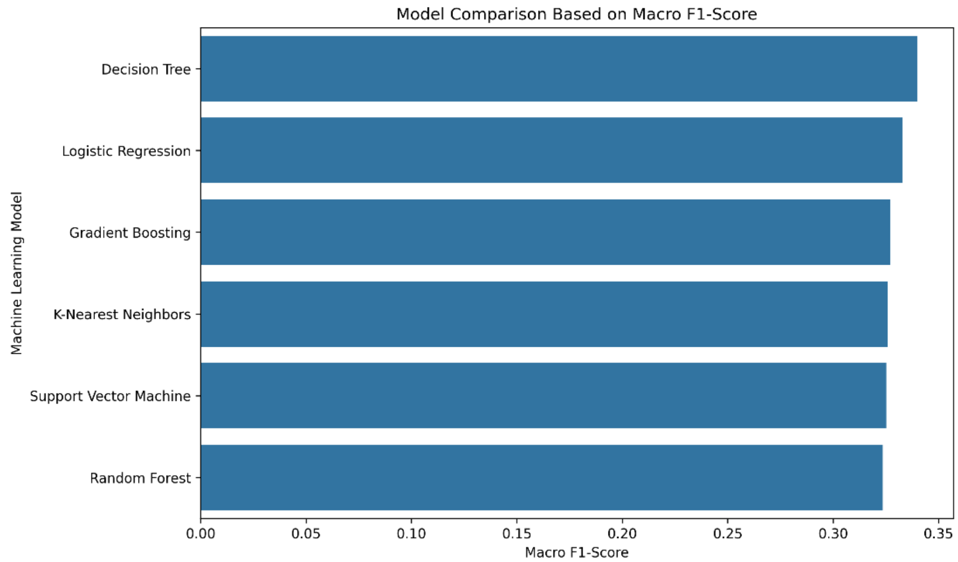

Figure 1. Model comparison based on macro F1-score

The figure confirms that Decision Tree produced the highest macro F1-score, followed by Logistic Regression and Gradient Boosting. However, the overall range of macro F1-scores was narrow, indicating that no model clearly outperformed the others.

Cross-Validation Results:

To evaluate the stability of the models, 5-fold stratified cross-validation was conducted. Unlike the single train-test split result, Gradient Boosting achieved the highest cross-validation performance, with a mean accuracy of 0.3480 and a mean macro F1-score of 0.3467.

Table 6. Five-Fold Cross-Validation Results

Model	CV Accuracy Mean	CV Accuracy Std	CV F1 Macro Mean	CV F1 Macro Std
Gradient Boosting	3,480	100	3,467	101
Random Forest	3,409	130	3,401	128
Logistic Regression	3,456	21	3,382	20
Support Vector Machine	3,372	93	3,356	98
K-Nearest Neighbors	3,356	78	3,298	76
Decision Tree	3,284	96	3,281	95

The cross-validation results show a slightly different ranking compared with the testing results. Gradient Boosting performed best during cross-validation, while Decision Tree performed best on the testing set. This difference indicates that the model ranking is not fully stable across different data partitions. However, the performance values remain close across all models, further supporting the finding that the dataset has limited class separability.

The low standard deviation values indicate that the models performed consistently across folds, but the average performance remained modest. Therefore, the issue is not mainly caused by unstable training, but more likely by the limited predictive signal contained in the available variables.

Confusion Matrix Analysis:

The confusion matrices provide a more detailed view of the classification behavior of the models. The confusion matrices for Gradient Boosting, Random Forest, and Support Vector Machine show that the models frequently misclassified samples across the three stress categories.

For Gradient Boosting, the model correctly classified 196 High stress samples, 254 Low stress samples, and 273 Medium stress samples. However, a large number of High stress samples were incorrectly predicted as Low or Medium. Similarly, Low and Medium stress samples were also distributed across incorrect predicted classes.

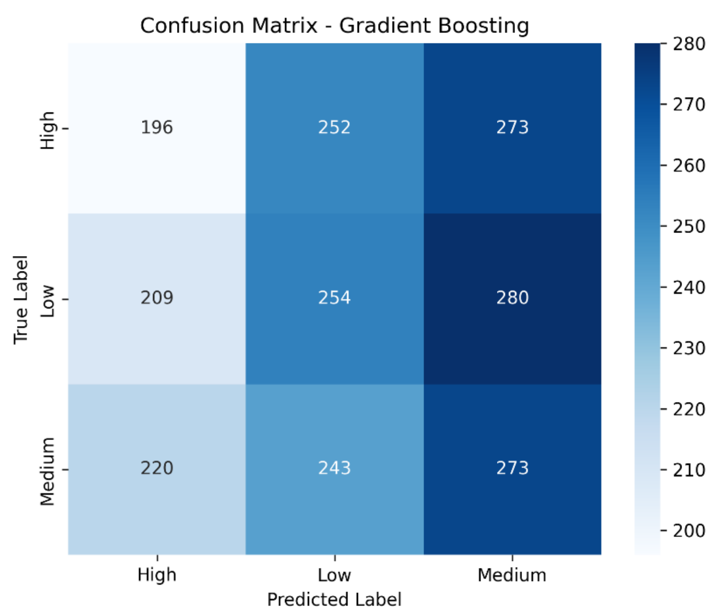

Figure 2. Confusion matrix of Gradient Boosting

For Random Forest, the model correctly classified 213 High stress samples, 255 Low stress samples, and 245 Medium stress samples. Similar to Gradient Boosting, Random Forest showed no strong diagonal dominance in the confusion matrix. This means that correct predictions were not substantially higher than incorrect predictions.

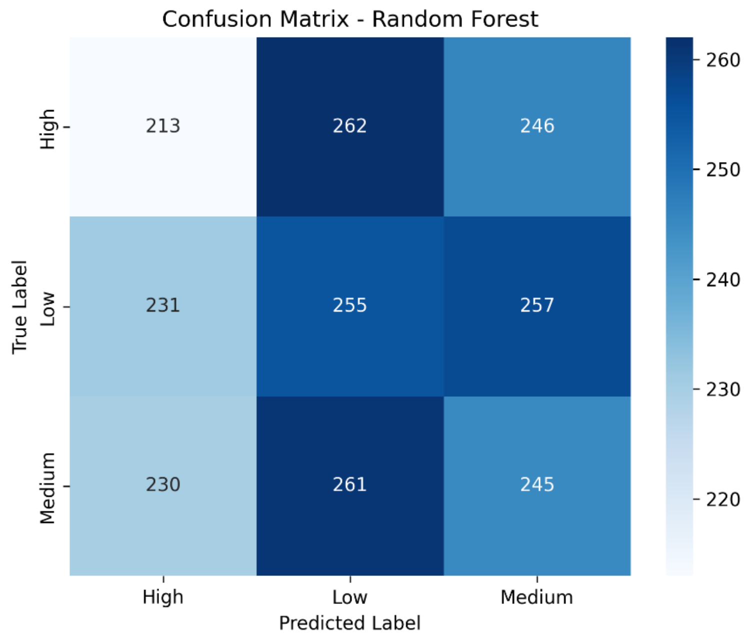

Figure 3. Confusion matrix of Random Forest

For Support Vector Machine, the model correctly classified 193 High stress samples, 263 Low stress samples, and 263 Medium stress samples. The model showed a tendency to predict samples as Low or Medium more frequently than High. This pattern suggests that High stress may be more difficult to distinguish using the current behavioral and lifestyle features.

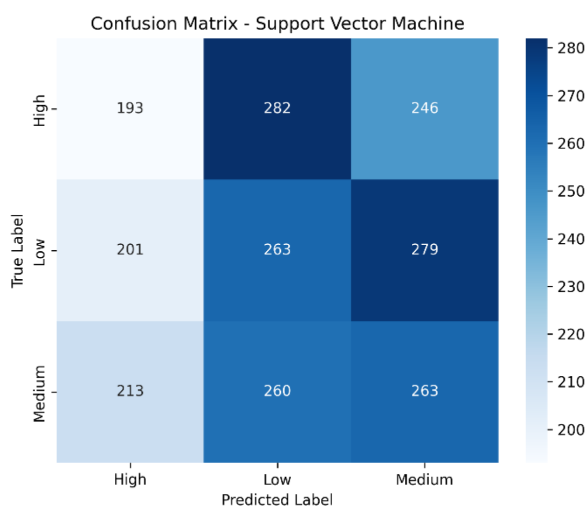

Figure 4. Confusion matrix of Support Vector Machine

Overall, the confusion matrix analysis indicates that the models struggled to separate the three stress categories. The values were distributed relatively evenly across predicted classes, which is consistent with the low macro F1-

scores. This result suggests that the current variables may represent general lifestyle patterns, but they may not be sufficient as standalone indicators for accurate stress level classification.

Feature Importance Analysis:

Feature importance analysis was conducted using the Random Forest model to identify which variables contributed most to the classification process. The results showed that productivity score, sleep hours, study/work hours, social media hours, and daily screen time were the most influential features.

Table 7. Feature Importance Based on Random Forest

Feature	Importance
Productivity score	1,425
Sleep hours	1,402
Study/work hours	1,400
Social media hours	1,400
Daily screen time hours	1,397
Exercise minutes	1,276
Age	1,082
Platform Twitter	159
Platform LinkedIn	157
Platform YouTube	157
Platform Snapchat	147

The five highest-ranked features had very similar importance values, ranging from 0.1397 to 0.1425. This indicates that no single feature dominated the prediction process. Productivity score had the highest importance score, followed by sleep hours, study/work hours, social media hours, and daily screen time [23].

The importance of productivity score suggests that perceived or measured productivity may be relevant to stress classification. Sleep hours also appeared as one of the most important variables, which is reasonable because sleep patterns are often associated with psychological well-being and daily functioning [24]. Social media hours and daily screen time also contributed to the model, supporting the idea that digital behavior may have some relationship with stress levels.

However, the platform-related variables had very low importance values. This indicates that the type of social media platform used was less informative than the amount of time spent on social media or overall screen exposure [25]. In other words, behavioral intensity appears to be more relevant than platform category in this dataset.

Discussion:

The experimental results show that machine learning models were able to produce measurable classifications of stress levels, but the overall predictive performance was low. The best model on the testing data was Decision Tree, with an accuracy and macro F1-score of 0.3400. Meanwhile, Gradient Boosting achieved the best cross-validation macro F1-score of 0.3467. These values are only slightly above the expected random baseline for a balanced three-class classification task.

This finding has an important implication. Although social media usage, sleep duration, exercise, productivity, and work or study hours are theoretically relevant to stress, their relationships with stress level in this dataset may not be sufficiently strong or linear for accurate classification. Stress is a complex psychological condition influenced by many factors, including emotional state, social support, work pressure, financial condition, personality, physical health, and environmental context. The current dataset only captures a limited subset of behavioral indicators, which may explain the modest model performance.

The confusion matrices further support this interpretation. The models did not show strong diagonal dominance, meaning that correct predictions were not clearly concentrated along the diagonal. Instead, many samples from each actual stress level were predicted as other stress categories. This indicates a high degree of overlap among Low, Medium, and High stress groups based on the available predictors.

From the feature importance analysis, productivity score, sleep hours, study/work hours, social media hours, and daily screen time were the most influential variables. This suggests that daily functioning and digital exposure may have a measurable role in stress prediction [26]. However, the relatively similar importance values across multiple features indicate that stress prediction in this dataset depends on a combination of weak signals rather than one dominant factor.

The results also highlight the importance of interpretability in health-related machine learning studies. If the study only focused on accuracy, the findings would appear limited. However, by including feature importance analysis, this study provides insight into which lifestyle and digital behavior variables contribute most to the prediction process. This is relevant for AI-driven health informatics because predictive systems in health-related contexts should be interpretable, transparent, and carefully evaluated.

These findings suggest that the proposed machine learning framework may be more suitable as an exploratory digital well-being assessment approach rather than a clinical diagnostic tool. The model should not be interpreted as a replacement for psychological assessment or clinical evaluation. Instead, it can be viewed as an initial computational approach for examining how behavioral data may relate to stress risk.

Future research should consider adding richer psychological, contextual, and physiological variables to improve predictive performance. For example, self-reported anxiety, depression indicators, perceived workload, quality of sleep, emotional well-being scores, social support, heart rate variability, and wearable sensor data may provide stronger signals for stress classification [27]. In addition, advanced explainable AI methods such as SHAP or LIME could be applied to provide a more detailed interpretation of individual-level predictions.

Overall, the results indicate that machine learning can be applied to digital behavior and lifestyle data for stress level prediction, but the current feature set has limited discriminative power. The main contribution of this study lies in demonstrating an interpretable machine learning workflow for stress prediction and identifying the relative influence of behavioral variables in a digital health informatics context.

4. Conclusion

This study developed and evaluated machine learning models for stress level prediction based on social media usage, sleep patterns, lifestyle behaviors, and productivity indicators. The dataset consisted of 11,000 records with three stress level categories: Low, Medium, and High. Several machine learning algorithms were compared, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Gradient Boosting.

The experimental results showed that the overall classification performance was relatively low. On the testing data, the Decision Tree model achieved the highest performance, with an accuracy and macro F1-score of 0.3400. Meanwhile, the 5-fold stratified cross-validation results showed that Gradient Boosting achieved the best average performance, with a mean accuracy of 0.3480 and a mean macro F1-score of 0.3467. Although these models produced measurable classification results, their performance was only slightly above the baseline for a balanced three-class classification problem.

The confusion matrix analysis indicated that the models had difficulty distinguishing between Low, Medium, and High stress levels. Misclassification occurred across all stress categories, suggesting that the available features did not provide strong separability among the target classes. This finding implies that stress is a complex psychological condition that may not be accurately predicted using only general digital behavior and lifestyle indicators.

Feature importance analysis using Random Forest showed that productivity score, sleep hours, study or work hours, social media hours, and daily screen time were the most influential variables in the prediction process. These results suggest that daily functioning, sleep duration, and digital exposure may contribute to stress level prediction. However, the importance values were relatively similar across several variables, indicating that no single feature strongly dominated the classification outcome.

Overall, this study demonstrates that machine learning can be applied to explore stress level prediction using non-invasive behavioral and lifestyle data [28]. However, the current dataset has limited predictive power for accurate stress classification. Therefore, the proposed approach is more appropriate as an exploratory digital well-being assessment framework rather than a clinical diagnostic tool. Future studies should incorporate richer psychological, contextual, and physiological variables, such as sleep quality, perceived workload, anxiety indicators, emotional well-being scores, social support, and wearable sensor data [29], [30]. In addition, advanced explainable AI techniques such as SHAP or LIME may be used to provide deeper interpretation of the factors influencing stress prediction.

References:

[1] O. Ahmed, E. I. Walsh, A. Dawel, K. Alateeq, D. A. Espinoza Oyarce, and N. Cherbuin, “Social media use, mental health and sleep: A systematic review with meta-analyses,” *J. Affect. Disord.*, vol. 367, pp. 701–712, Dec. 2024, doi: 10.1016/j.jad.2024.08.193.

[2] D. J. Yu, Y. K. Wing, T. M. H. Li, and N. Y. Chan, “The Impact of Social Media Use on Sleep and Mental Health in Youth: a Scoping Review,” *Curr. Psychiatry Rep.*, vol. 26, no. 3, pp. 104–119, Mar. 2024, doi: 10.1007/s11920-024-01481-9.

[3] X. Han, E. Zhou, and D. Liu, “Electronic Media Use and Sleep Quality: Updated Systematic Review and Meta-Analysis,” *J. Med. Internet Res.*, vol. 26, p. e48356, Apr. 2024, doi: 10.2196/48356.

[4] H. Shannon, K. Bush, P. J. Villeneuve, K. G. Hellemans, and S. Guimond, “Problematic Social Media Use in Adolescents and Young Adults: Systematic Review and Meta-analysis,” *JMIR Ment. Health*, vol. 9, no. 4, p. e33450, Apr. 2022, doi: 10.2196/33450.

[5] R. M. S. Santos *et al.*, “The associations between screen time and mental health in adolescents: a systematic review,” *BMC Psychol.*, vol. 11, no. 1, p. 127, Apr. 2023, doi: 10.1186/s40359-023-01166-7.

[6] M. Chen *et al.*, “Screen viewing patterns and their association with mental wellbeing and psychological distress: a cross-sectional study amongst university students,” *BMC Public Health*, vol. 24, no. 1, p. 2097, Aug. 2024, doi: 10.1186/s12889-024-19608-1.

[7] L. E. Hartstein *et al.*, “The impact of screen use on sleep health across the lifespan: A National Sleep Foundation consensus statement,” *Sleep Health*, vol. 10, no. 4, pp. 373–384, Aug. 2024, doi: 10.1016/j.sleh.2024.05.001.

[8] S. Gedam and S. Paul, “A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques,” *IEEE Access*, vol. 9, pp. 84045–84066, 2021, doi: 10.1109/ACCESS.2021.3085502.

[9] S. Bauducco, M. Pillion, K. Bartel, C. Reynolds, M. Kahn, and M. Gradisar, “A bidirectional model of sleep and technology use: A theoretical review of How much, for whom, and which mechanisms,” *Sleep Med. Rev.*, vol. 76, p. 101933, Aug. 2024, doi: 10.1016/j.smrv.2024.101933.

[10] T. Fischer, M. Reuter, and R. Riedl, “The Digital Stressors Scale: Development and Validation of a New Survey Instrument to Measure Digital Stress Perceptions in the Workplace Context,” *Front. Psychol.*, vol. 12, p. 607598, Mar. 2021, doi: 10.3389/fpsyg.2021.607598.

[11] A. J. Scott, T. L. Webb, M. Martyn-St James, G. Rowse, and S. Weich, “Improving sleep quality leads to better mental health: A meta-analysis of randomised controlled trials,” *Sleep Med. Rev.*, vol. 60, p. 101556, Dec. 2021, doi: 10.1016/j.smrv.2021.101556.

[12] K. Kaewpradit, P. Ngamchaliew, and N. Buathong, “Digital screen time usage, prevalence of excessive digital screen time, and its association with mental health, sleep quality, and academic performance among Southern University students,” *Front. Psychiatry*, vol. 16, p. 1535631, Mar. 2025, doi: 10.3389/fpsyt.2025.1535631.

[13] A. Sevic, N. Foldnes, and K. K. Brønnick, “Measuring digital stress in Norway: translation and validation of the Digital Stressors Scale,” *Front. Psychol.*, vol. 15, p. 1297194, Feb. 2024, doi: 10.3389/fpsyg.2024.1297194.

[14] B. M. Booth, H. Vrzakova, S. M. Mattingly, G. J. Martinez, L. Faust, and S. K. D’Mello, “Toward Robust Stress Prediction in the Age of Wearables: Modeling Perceived Stress in a Longitudinal Study With Information Workers,” *IEEE Trans. Affect. Comput.*, vol. 13, no. 4, pp. 2201–2217, Oct. 2022, doi: 10.1109/TAFFC.2022.3188006.

[15] B. Saylam and Ö. D. İncel, “Quantifying Digital Biomarkers for Well-Being: Stress, Anxiety, Positive and Negative Affect via Wearable Devices and Their Time-Based Predictions,” *Sensors*, vol. 23, no. 21, p. 8987, Nov. 2023, doi: 10.3390/s23218987.

[16] A. A. Al-Atawi *et al.*, “Stress Monitoring Using Machine Learning, IoT and Wearable Sensors,” *Sensors*, vol. 23, no. 21, p. 8875, Oct. 2023, doi: 10.3390/s23218875.

[17] V. Farrahi and M. Rostami, “Machine learning in physical activity, sedentary, and sleep behavior research,” *J. Act. Sedentary Sleep Behav.*, vol. 3, no. 1, p. 5, Jan. 2024, doi: 10.1186/s44167-024-00045-9.

[18] J. R. Beames *et al.*, “Use of smartphone sensor data in detecting and predicting depression and anxiety in young people (12–25 years): A scoping review,” *Heliyon*, vol. 10, no. 15, p. e35472, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35472.

[19] H. S. Saad, J. F. W. Zaki, and M. M. Abdelsalam, “Employing of machine learning and wearable devices in healthcare system: tasks and challenges,” *Neural Comput. Appl.*, vol. 36, no. 29, pp. 17829–17849, Oct. 2024, doi: 10.1007/s00521-024-10197-z.

[20] M. Ciharova *et al.*, “Machine-learning detection of stress severity expressed on a continuous scale using acoustic, verbal, visual, and physiological data: lessons learned,” *Front. Psychiatry*, vol. 16, p. 1548287, Jun. 2025, doi: 10.3389/fpsyt.2025.1548287.

[21] M. Razavi *et al.*, “Machine Learning, Deep Learning, and Data Preprocessing Techniques for Detecting, Predicting, and Monitoring Stress and Stress-Related Mental Disorders: Scoping Review,” *JMIR Ment. Health*, vol. 11, p. e53714, Aug. 2024, doi: 10.2196/53714.

[22] M. K. Moser, M. Ehrhart, and B. Resch, “An Explainable Deep Learning Approach for Stress Detection in Wearable Sensor Measurements,” *Sensors*, vol. 24, no. 16, p. 5085, Aug. 2024, doi: 10.3390/s24165085.

[23] L. As. Brautsch, L. Lund, M. M. Andersen, P. J. Jennum, A. P. Folker, and S. Andersen, “Digital media use and sleep in late adolescence and young adulthood: A systematic review,” *Sleep Med. Rev.*, vol. 68, p. 101742, Apr. 2023, doi: 10.1016/j.smrv.2022.101742.

[24] B. Brosnan, J. J. Haszard, K. A. Meredith-Jones, S.-R. Wickham, B. C. Galland, and R. W. Taylor, “Screen Use at Bedtime and Sleep Duration and Quality Among Youths,” *JAMA Pediatr.*, vol. 178, no. 11, p. 1147, Nov. 2024, doi: 10.1001/jamapediatrics.2024.2914.

[25] T. Sanders *et al.*, “An umbrella review of the benefits and risks associated with youths’ interactions with electronic screens,” *Nat. Hum. Behav.*, vol. 8, no. 1, pp. 82–99, Nov. 2023, doi: 10.1038/s41562-023-01712-8.

[26] M. Shanmugasundaram and A. Tamilarasu, “The impact of digital technology, social media, and artificial intelligence on cognitive functions: a review,” *Front. Cogn.*, vol. 2, p. 1203077, Nov. 2023, doi: 10.3389/fcogn.2023.1203077.

[27] A. Pinge, V. Gad, D. Jaisighani, S. Ghosh, and S. Sen, “Detection and monitoring of stress using wearables: a systematic review,” *Front. Comput. Sci.*, vol. 6, p. 1478851, Dec. 2024, doi: 10.3389/fcomp.2024.1478851.

[28] A. S. Hammad, A. Tajammul, I. Dergaa, and M. Al-Asmakh, “Machine learning applications in the analysis of sedentary behavior and associated health risks,” *Front. Artif. Intell.*, vol. 8, p. 1538807, Jun. 2025, doi: 10.3389/frai.2025.1538807.

[29] H. T. Pham, H.-L. Chuang, C.-P. Kuo, T.-P. Yeh, and W.-C. Liao, “Electronic Device Use before Bedtime and Sleep Quality among University Students,” *Healthcare*, vol. 9, no. 9, p. 1091, Aug. 2021, doi: 10.3390/healthcare9091091.

[30] J.-Z. Xiang, Q.-Y. Wang, Z.-B. Fang, J. A. Esquivel, and Z.-X. Su, “A multi-modal deep learning approach for stress detection using physiological signals: integrating time and frequency domain features,” *Front. Physiol.*, vol. 16, p. 1584299, Apr. 2025, doi: 10.3389/fphys.2025.1584299.