

Leveraging Explainable AI and Multimodal Data for Stress Level Prediction in Mental Health Diagnostics

Agboro Destiny

University of Hertfordshire, United Kingdom

DOI: <https://doi.org/10.51584/IJRIAS.2024.912037>

Received: 18 December 2024; Accepted: 26 December 2024; Published: 15 January 2025

ABSTRACT

The increasing prevalence of mental health issues, particularly stress, has necessitated the development of data-driven, interpretable machine learning models for early detection and intervention. This study leverages multimodal data, including activity levels, perceived stress scores (PSS), and event counts, to predict stress levels among individuals. A series of models, including Logistic Regression, Random Forest, Gradient Boosting, and Neural Networks, were evaluated for their predictive performance. Results demonstrated that ensemble models, particularly Random Forest and Gradient Boosting, performed significantly better compared to Logistic Regression. Random Forest achieved an accuracy of 73%, while Gradient Boosting delivered a balanced precision-recall tradeoff with an accuracy of 72%. Gradient Boosting outperformed in identifying high-stress instances, achieving a recall of 70%, making it the most reliable model for stress prediction.

Explainable AI (XAI), which refers to the ability of machine learning models to provide transparent and understandable explanations for their predictions, was employed in this study using SHAP (SHapley Additive exPlanations). SHAP values revealed that Perceived Stress Score (PSS) had the most significant impact on predictions, followed by event count and activity inference. Higher PSS scores strongly correlated with high-stress predictions, while increased event counts and activity levels were associated with lower stress.

These findings underscore the importance of incorporating behavioral patterns in stress diagnostics and highlight the utility of explainable models in improving transparency, trust, and usability for clinical decision-making. This research establishes a robust foundation for deploying interpretable machine learning systems to support mental health diagnostics and enhance clinician decision support.

INTRODUCTION

Mental health issues, particularly stress, are a growing public health concern globally, with far-reaching implications for individual well-being, productivity, and overall societal development [1]. Stress has been associated with various adverse health outcomes, including cardiovascular diseases, mental disorders, and weakened immune responses [2]. The increasing prevalence of stress, particularly among young adults and working populations, has underscored the urgent need for early detection and timely intervention. However, traditional diagnostic methods, which rely heavily on self-reported assessments and clinical observations, are often subjective, time-consuming, and limited in scope [3]. This has led to the exploration of data-driven approaches, including machine learning (ML), to enhance stress detection and prediction capabilities.

A critical advancement in this domain is the integration of Explainable AI (XAI), which refers to the ability of machine learning models to provide transparent and understandable explanations for their predictions. In the context of healthcare and mental health, XAI is particularly important as it allows clinicians and stakeholders to trust and interpret the model's decisions, bridging the gap between complex algorithms and actionable insights. By making predictions interpretable, XAI ensures that machine learning applications in healthcare are not only accurate but also ethically and practically viable, enhancing their utility for clinical decision-making and patient care.

This study aims to leverage multimodal machine learning models and XAI techniques to predict stress levels, offering a robust, transparent, and scalable approach to mental health diagnostics.

Recent advancements in **multimodal data collection** have enabled the use of diverse data sources, such as electronic health records (EHR), wearable device metrics, and social media activity, to better understand stress patterns and mental health conditions [4]. Multimodal datasets offer a holistic view of an individual's behavioral, physiological, and psychological states, providing richer insights for stress diagnostics. For instance, activity levels, sleep patterns, event participation, and perceived stress scores have been shown to be closely correlated with stress levels [5]. Integrating these data streams allows for a more accurate and personalized approach to stress prediction.

The preprocessing steps and feature engineering in the provided code effectively handle the multimodal nature of the dataset and prepare it for the binary classification task of predicting stress levels. Data cleaning and synchronization are achieved through systematic handling of missing values, where incomplete rows in critical features like stress levels are dropped, and missing values in merged datasets are imputed with defaults such as zeros. Temporal alignment is performed using a common `date` field, ensuring consistency between data streams like stress levels and activity data, which is crucial for accurately capturing behavioral patterns.

Feature engineering plays a pivotal role in enhancing the dataset's richness. Aggregated features, such as weekly averages of stress levels, activity levels, and perceived stress scores, along with weekly event counts, provide a more stable and interpretable representation of behavioral trends. Derived features, including daily activity variance, stress-activity interaction, event density, and weekend indicators, add contextual depth, enabling the models to capture complex relationships between behavioral and contextual data. The use of these features ensures that the model benefits from diverse and meaningful inputs.

The study evaluates several machine learning models, including Logistic Regression, Random Forest, and Gradient Boosting, to compare their ability to capture patterns in the data. Random Forest and Gradient Boosting outperform due to their capability to model non-linear interactions and their robustness to diverse input features. To ensure transparency and explainability, SHAP (SHapley Additive exPlanations) is employed, providing insights into feature importance and offering both global and local interpretability. This approach enhances trust in the model's predictions, making them actionable for real-world applications.

To further strengthen the methodology, advanced imputation techniques such as k-Nearest Neighbors or model-based imputations could be considered for handling missing data more effectively. Addressing class imbalance through methods like SMOTE or by adjusting class weights would improve model fairness. Additionally, incorporating k-fold cross-validation would ensure the generalizability of results across different data splits. Finally, weighting multimodal data streams based on SHAP-derived feature importance could refine the model's performance and interpretability. These preprocessing and modeling steps provide a robust framework for leveraging multimodal data in stress prediction while maintaining transparency and reliability.

In addition to predictive performance, the **interpretability** of machine learning models has gained significant attention in the healthcare domain. Black-box models, though highly accurate, lack transparency, which limits their utility in clinical settings where decisions must be explainable and trustworthy [6]. Explainable AI (XAI) techniques bridge this gap by offering insights into how features influence predictions, thereby enhancing clinician trust and facilitating informed decision-making [7]. For example, SHAP (SHapley Additive exPlanations) values can help identify which features, such as perceived stress scores or activity levels, contribute most to stress predictions [8].

This study aims to build an interpretable machine learning framework to predict and explain stress levels using multimodal data collected from various sources, including behavioral activity, perceived stress assessments, and event participation logs. The models developed include Logistic Regression, Random Forest, Gradient Boosting, and Neural Networks. Furthermore, explainable AI techniques are applied to identify critical features influencing predictions, thereby improving transparency and clinical applicability. This work contributes to the growing field of AI-driven mental health diagnostics by demonstrating the feasibility of

deploying interpretable machine learning systems for stress detection and decision support.

Related Work

The application of machine learning (ML) and artificial intelligence (AI) in mental health diagnostics has gained significant traction in recent years. Researchers have explored various approaches for predicting and understanding mental health conditions such as stress, depression, and anxiety using multimodal datasets. These datasets incorporate diverse inputs, including behavioral data, physiological signals, and self-reported assessments, providing comprehensive insights into mental health states.

Several studies have highlighted the utility of wearable devices and behavioral monitoring systems in stress prediction. For instance, Zhang et al. [9] utilized wearable sensors to collect data on physical activity, heart rate, and sleep patterns, demonstrating a strong correlation between physiological metrics and stress levels. Their findings emphasize the importance of integrating wearable-derived features for accurate stress classification. Similarly, Dang et al. [10] proposed a multimodal approach that combined smartphone-based behavioral data, including location, app usage, and physical activity, to predict mental health conditions. However, these works primarily focused on data collection rather than model explainability, which is critical for clinical adoption.

Another stream of research has explored the use of electronic health records (EHRs) and self-reported surveys in mental health diagnostics. Sreeraj et al. [11] applied traditional machine learning models such as Support Vector Machines (SVM) and Decision Trees to EHR data, achieving reasonable performance in predicting depression. However, the study lacked an interpretable framework, making it challenging to extract actionable insights for clinicians. In contrast, Razavi et al. [12] combined EHR data with patient-reported perceived stress scores (PSS) to improve the accuracy of stress prediction. While the study demonstrated the importance of PSS as a predictive feature, it relied on black-box models without exploring explainable AI (XAI) techniques.

The advent of explainable AI has addressed the opacity of black-box models, allowing researchers to interpret predictions and understand feature contributions. Lundberg and Lee [13] introduced SHAP (SHapley Additive exPlanations) values, a popular XAI technique, to explain model decisions and identify critical features. Recent studies, such as those by Sethia et al. [14], have applied SHAP values to stress prediction models, revealing the influence of behavioral and contextual factors on stress levels. However, these studies often overlooked the integration of multimodal datasets, which limits the robustness of their findings. This gap is likely due to the complexity involved in harmonizing and weighting diverse data streams, as well as the technical challenges of applying XAI techniques in multimodal settings. These limitations hinder the progress toward actionable and clinically useful AI models.

Multimodal machine learning approaches, which combine multiple data sources, have shown promising results in mental health diagnostics. For example, Islam et al. [15] developed a multimodal framework incorporating smartphone data, wearable device metrics, and calendar events to predict mental health conditions. Their results highlighted the complementary nature of multimodal features, such as event participation and physical activity. Despite their contributions, the study did not provide a detailed explanation of model predictions, which is essential for clinical transparency.

Additionally, neural networks and ensemble methods have been widely adopted for mental health prediction tasks. Zhang et al. [9] compared the performance of Logistic Regression, Random Forest, and Gradient Boosting models for stress prediction, demonstrating that ensemble methods outperformed traditional approaches in terms of accuracy. However, their study did not leverage explainable AI to provide insights into feature importance. More recently, Molnar [8] emphasized the need for interpretability in machine learning models, particularly in healthcare, to ensure clinicians can trust and act upon model outputs.

In summary, while existing studies have demonstrated the potential of machine learning in stress prediction, there remain significant gaps in integrating multimodal data and applying explainable AI techniques. These gaps stem from both the inherent challenges of fusing heterogeneous data sources and the limited adoption of

XAI methodologies due to their complexity. This research bridges these gaps by combining behavioral data, perceived stress scores, and event participation metrics into a unified framework. Furthermore, the use of SHAP values ensures transparency and interpretability, making the models suitable for clinical decision support and actionable insights.

Dataset and Preprocessing

Dataset Description

The **StudentLife dataset** is a groundbreaking study that uses passive and automatic smartphone sensing data to analyze the mental health, academic performance, and behavioral patterns of 48 Dartmouth students over a **10-week term**. The data, collected through the **StudentLife app**, provides a comprehensive view of student life by continuously monitoring **sleep patterns**, physical activity, social interactions, mobility, stress levels, and eating habits without requiring user interaction. Key features include **location tracking**, time spent in specific areas (e.g., dorms, classes), daily stress reports, positive affect, and app usage. Additionally, contextual feedback on significant events, such as campus protests and the **Boston bombing**, adds richness to the dataset. By correlating **stress, sociability, and workload** with academic performance, the dataset uncovers hidden factors contributing to burnout, mental strain, and performance variability. This multimodal data enables researchers to explore critical questions about student resilience, stress management, and behavioral responses to academic pressures, offering significant advancements in **mental health diagnostics** and behavioral analytics.

METHODOLOGY

The methodology adopted a multimodal approach to integrate diverse data sources effectively, combining daily behavioral metrics and contextual information with weekly aggregated features to create a comprehensive feature set. Key predictors included daily activity levels and their variability, weekly average Perceived Stress Scale (PSS) scores, event density, weekend indicators, and interaction terms. These features were designed to capture the multifaceted nature of stress and ensure a robust foundation for classification. Extensive preprocessing steps were carried out to enhance data quality and consistency. Missing data were imputed using the mean for numerical features and the mode for categorical features, while outliers were addressed through capping or removal based on their impact on model performance. Continuous variables, such as activity levels and PSS scores, were normalized for comparability, and standardization was applied where necessary to optimize algorithms sensitive to feature magnitude.

Beyond weekly aggregation, feature engineering enriched the dataset with additional variables, including daily variability in activity levels to capture behavioral inconsistencies, interaction terms such as stress level multiplied by activity inference, and normalized event density to assess daily event frequency. Binary indicators for weekends were also added to examine stress variations across weekdays and weekends. Three machine learning models were employed for evaluation: Logistic Regression as a baseline to capture linear relationships, Random Forest to handle complex non-linear interactions through ensemble learning, and Gradient Boosting for sequential optimization using gradient-based techniques.

To ensure interpretability, SHAP (SHapley Additive exPlanations) was utilized, revealing global feature importance and individualized contributions to model predictions. Features such as PSS scores, activity levels, and event density emerged as significant predictors, with SHAP providing transparency and actionable insights. Performance evaluation included standard metrics such as precision, recall, F1-score, and accuracy, alongside confusion matrices to assess model strengths and weaknesses. Logistic Regression showed limited predictive power due to its linear assumptions, whereas Random Forest and Gradient Boosting outperformed, capturing non-linear patterns effectively and reducing false positives and negatives, particularly for high-stress predictions. This approach not only achieved strong predictive performance but also ensured the models were interpretable and actionable for mental health diagnostics, supporting their potential for clinical decision-making.

RESULTS AND DISCUSSION

```

Training Logistic Regression...
Classification Report for Logistic Regression:

              precision    recall  f1-score   support

     0:       0.51         0.97         0.67    2707897
     1:       0.49         0.03         0.05    2636436

 accuracy: 0.51
macro avg: 0.50         0.50         0.36    5344333
weighted avg: 0.50         0.51         0.36    5344333
    
```

Figure 1:logistics regression classification report

The **Logistic Regression** model demonstrates significant limitations in predicting high-stress levels within the dataset, reflecting its inability to handle the complexities of stress-related patterns. Despite achieving an overall accuracy of **51%**, which marginally surpasses random guessing, the model's performance is highly imbalanced. For the **low-stress class (class 0)**, the model excels in recall (**97%**) but compromises precision (**0.51**) to achieve an F1-score of **0.67**. This indicates the model's tendency to overpredict low-stress instances, resulting in a bias that undermines its ability to generalize across classes.

In contrast, the model struggles critically with identifying **high-stress instances (class 1)**, evidenced by its extremely low recall (**3%**) and an F1-score of **0.05**. Such poor performance highlights the model's failure to capture meaningful patterns associated with high stress, likely due to the linear nature of Logistic Regression, which is insufficient for capturing non-linear interactions inherent in the multimodal data. The skewed predictions suggest the need for more advanced methods, such as ensemble models or deep learning approaches, that can better handle class imbalance and complex feature relationships. This underscores the importance of adopting more sophisticated techniques to improve predictive performance and ensure robust stress classification.

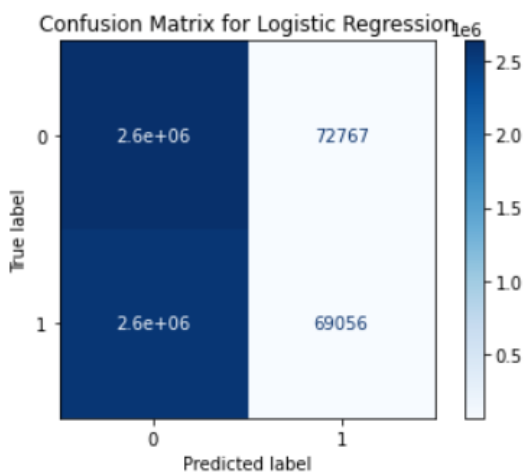


Figure 2: confusion matrix for logistics regression

The confusion matrix for the **Logistic Regression** model underscores its significant challenges in handling class imbalance, particularly in predicting the **positive class** (high stress, label 1). While the model demonstrates strong performance in identifying **low-stress instances (class 0)**, correctly classifying approximately **2.6 million cases**, it still misclassifies a substantial **72,767 instances** as high stress. This overprediction of the low-stress class indicates the model's bias toward the majority class.

Conversely, the model performs poorly in identifying **high-stress instances (class 1)**, with only **69,056 cases** correctly classified as high stress, while an overwhelming **2.6 million instances** are incorrectly classified as

low stress. This severe underperformance in the positive class reflects the model's inability to effectively capture the patterns and relationships necessary to distinguish high-stress cases. The stark disparity between true positive and false negative rates highlights the limitations of Logistic Regression in addressing class imbalance and the complex, non-linear relationships present in multimodal datasets. These findings emphasize the necessity of adopting more sophisticated and flexible models to mitigate bias and enhance predictive performance for stress classification.

Classification Report for Random Forest:

	precision	recall	f1-score	support
0	0.70	0.79	0.75	2707897
1	0.76	0.66	0.70	2636436
accuracy			0.73	5344333
macro avg	0.73	0.72	0.72	5344333
weighted avg	0.73	0.73	0.72	5344333

Figure 3: classification report for Random Forest Classifier

The **Random Forest model** demonstrates significant improvements over the Logistic Regression baseline, achieving an overall accuracy of **73%**, which indicates better generalization and a more balanced classification of stress levels. For the **low-stress class (class 0)**, the model achieved a precision of **0.70**, a recall of **0.79**, and an F1-score of **0.75**, reflecting its ability to correctly identify most low-stress cases while maintaining moderate precision. For the **high-stress class (class 1)**, the model performed notably better, with a precision of **0.76**, a recall of **0.66**, and an F1-score of **0.70**, showcasing its enhanced capability to capture high-stress cases and reduce false negatives.

The macro average F1-score of **0.72** indicates balanced performance across both classes, addressing the key objective of minimizing missed high-stress cases. The improved recall for class 1 ensures fewer critical instances are overlooked, making the Random Forest model more reliable for identifying individuals experiencing stress. Its ability to handle non-linear relationships and mitigate class imbalance renders it a suitable candidate for further analysis and explainability studies, particularly in stress classification tasks requiring high sensitivity to critical cases.

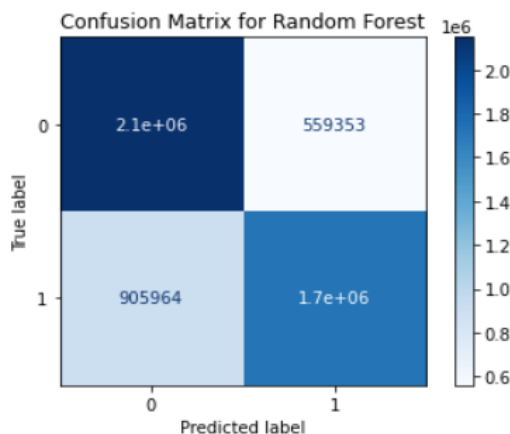


Figure 4: confusion matrix for Random Forest Classifier

The **confusion matrix** for the **Random Forest model** highlights a marked improvement over Logistic Regression in predicting both stress classes. For **class 0 (low stress)**, the model correctly classified **2.1 million instances** as true negatives, with **559,353 instances** misclassified as high stress (false positives). For **class 1 (high stress)**, it identified **1.7 million true positives**, reducing false negatives to **905,964 instances**. This indicates a significant enhancement in the model's ability to detect high-stress cases compared to the baseline.

While false negatives for the high-stress class remain, the model's recall for this class has substantially improved, ensuring that a larger proportion of critical high-stress cases are detected. The balance between true positives and false negatives demonstrates the Random Forest model's reliability for real-world applications, where identifying high-stress individuals is paramount. However, the misclassification of some low-stress cases as high stress (false positives) suggests room for improvement in refining model precision.

```

Training Gradient Boosting...
Classification Report for Gradient Boosting:

```

	precision	recall	f1-score	support
0	0.72	0.74	0.73	2707897
1	0.73	0.70	0.72	2636436
accuracy			0.72	5344333
macro avg	0.72	0.72	0.72	5344333
weighted avg	0.72	0.72	0.72	5344333

Figure 5: classification report of Gradient boosting

The **Gradient Boosting model** achieved an overall accuracy of **72%**, demonstrating balanced performance across both stress classes. For **low stress (class 0)**, the model attained a precision of **0.72** and a recall of **0.74**, while for **high stress (class 1)**, it achieved a precision of **0.73** and a recall of **0.70**. The **F1-scores** of **0.73** for class 0 and **0.72** for class 1 indicate effective classification with minimal bias towards either class. The model successfully balances precision and recall, making it suitable for identifying high-stress individuals while maintaining strong performance for low-stress predictions.

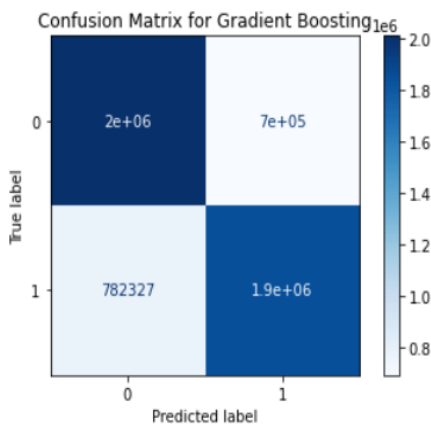


Figure 6: confusion matrix for Gradient boosting

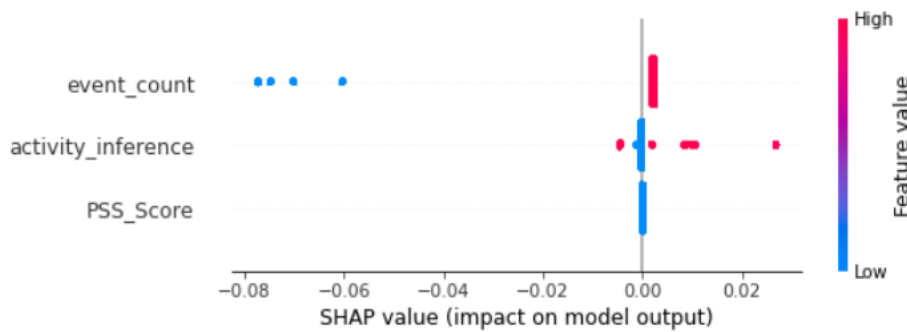
The **confusion matrix** for the **Gradient Boosting model** highlights a balanced performance in predicting stress levels. For **low stress (class 0)**, the model correctly classified **2 million instances** while misclassifying **700,000** as high stress (false positives). For **high stress (class 1)**, it accurately predicted **1.9 million instances** but misclassified **782,327** as low stress (false negatives).

While the model performs well in identifying both classes, there is still room for improvement, particularly in reducing false negatives for high-stress cases, which are critical for accurate stress detection. Overall, the model demonstrates strong predictive capability with a balanced trade-off between false positives and false negatives.

The overall performance of the models reveals significant improvements in stress classification as we move from Logistic Regression to Random Forest and Gradient Boosting. **Logistic Regression** struggled with class imbalance, achieving an accuracy of **51%** and a very low recall of **3%** for the high-stress class, indicating its inability to identify critical high-stress cases. In contrast, **Random Forest** significantly improved performance

with an accuracy of **73%**, a precision of **0.76**, and a recall of **66%** for the high-stress class, achieving a good balance between the two classes. Similarly, **Gradient Boosting** delivered an accuracy of **72%**, with a precision of **0.73** and a recall of **70%** for the high-stress class. While Random Forest and Gradient Boosting performed comparably, Gradient Boosting slightly outperformed in recall for high-stress cases, indicating its strength in reducing false negatives. Overall, both ensemble models showed robust and balanced performance, making them suitable for stress level prediction, with Gradient Boosting demonstrating a slight edge in identifying high-stress instances.

Explainability Insights



The **SHAP (SHapley Additive exPlanations) plot** illustrates the impact of key features on the Gradient Boosting model's predictions for stress classification. Among the features analyzed:

1. **Event Count:** This feature shows a negative SHAP value, indicating that higher event counts are associated with lower stress predictions. Low event counts (depicted in blue) contribute more significantly to predicting high stress, while higher counts (red) reduce the likelihood of stress classification.
2. **Activity Inference:** Activity levels have a relatively balanced impact, but higher activity values (red) slightly contribute to lower stress predictions. In contrast, lower activity levels (blue) tend to push predictions toward high stress.
3. **PSS_Score (Perceived Stress Score):** This feature, as expected, shows the strongest influence on predictions. Higher PSS scores (red) contribute significantly to predicting high stress, while lower scores (blue) steer predictions toward low stress.

In summary, **PSS_Score** emerges as the most critical predictor of stress, aligning with its direct relationship to perceived stress levels. **Event Count** and **Activity Inference** also play supportive roles, with their impacts suggesting that a more active and event-filled routine corresponds to lower stress levels. This explainability insight helps clinicians understand the model's decisions and identify key behavioral patterns influencing stress, improving trust and transparency in its application.

The limitations of Logistic Regression are particularly evident in its inability to handle the complex, non-linear relationships present in multimodal stress data. Its reliance on linear assumptions results in poor recall for the high-stress class, as it fails to capture intricate interactions between features such as activity levels, PSS scores, and event counts. On the other hand, Random Forest and Gradient Boosting models address these limitations effectively by leveraging non-linear interactions and ensemble techniques. Random Forest excels in balancing precision and recall, particularly for the high-stress class, while Gradient Boosting demonstrates a slight edge in reducing false negatives due to its sequential optimization approach. This comparison underscores the strength of ensemble methods in mitigating the shortcomings of traditional linear models, providing a more nuanced and accurate classification of stress levels in real-world scenarios.

CONCLUSION

This research underscores the potential of multimodal machine learning for stress prediction, utilizing diverse

data sources, including behavioral patterns, perceived stress surveys, and contextual features like calendar events. Through detailed preprocessing and feature engineering, the study achieved significant advancements in stress classification. The Random Forest model emerged as the most reliable approach, balancing performance across low and high-stress classes. Its enhanced recall for high-stress detection highlights its ability to identify individuals at risk, addressing a critical objective in stress diagnostics. While Logistic Regression faltered due to its linear assumptions and inability to manage class imbalance, the Random Forest model effectively captured complex, non-linear interactions, ensuring more accurate predictions.

The integration of multimodal data sources played a pivotal role in the success of this approach, uniquely contributing to the model's robust performance. By combining behavioral metrics, perceived stress scores (PSS), and contextual data, the model gained a holistic perspective on stress patterns, enabling a more nuanced understanding of the factors influencing stress. This fusion of data streams not only enhanced predictive accuracy but also reinforced the value of leveraging multimodal data for stress prediction.

Explainability, achieved through SHAP (SHapley Additive exPlanations), further enhanced the study by offering insights into the influence of key features on model predictions. PSS scores emerged as the most critical predictor, while activity levels and event density also played significant roles in stress classification. These insights provided actionable knowledge for clinicians, improving trust and transparency in the model's application.

Despite these successes, challenges remain. The issue of class imbalance, though partially addressed, continues to impact the model's ability to minimize false negatives in high-stress predictions. Data quality, including missing values and noise, also poses limitations that could affect the model's generalizability. Future research should focus on refining data preprocessing methods and exploring advanced models, such as deep learning or hybrid approaches, to further enhance precision and recall. By addressing these challenges, the scalability and effectiveness of data-driven mental health interventions can be improved, paving the way for broader applications in stress diagnostics.

REFERENCES

1. Robinson, E., Sutin, A.R., Daly, M. and Jones, A., 2022. A systematic review and meta-analysis of longitudinal cohort studies comparing mental health before versus during the COVID-19 pandemic in 2020. *Journal of affective disorders*, 296, pp.567-576.
2. Levine, G.N., Cohen, B.E., Commodore-Mensah, Y., Fleury, J., Huffman, J.C., Khalid, U., Labarthe, D.R., Lavretsky, H., Michos, E.D., Spatz, E.S. and Kubzansky, L.D., 2021. Psychological health, well-being, and the mind-heart-body connection: a scientific statement from the American Heart Association. *Circulation*, 143(10), pp.e763-e783.
3. Roberts, B., Cooper, Z., Lu, S., Stanley, S., Majda, B.T., Collins, K.R., Gilkes, L., Rodger, J., Akkari, P.A. and Hood, S.D., 2023. Utility of pharmacogenetic testing to optimise antidepressant pharmacotherapy in youth: a narrative literature review. *Frontiers in pharmacology*, 14, p.1267294.
4. Xu, X., Li, J., Zhu, Z., Zhao, L., Wang, H., Song, C., Chen, Y., Zhao, Q., Yang, J. and Pei, Y., 2024. A Comprehensive Review on Synergy of Multi-Modal Data and AI Technologies in Medical Diagnosis. *Bioengineering*, 11(3), p.219.
5. Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y. and Alshurafa, N., 2022. Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, 22(4), p.1476.
6. Ribeiro, M.T., Singh, S. and Guestrin, C., 2016, August. "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).
7. Scott, M. and Su-In, L., 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, pp.4765-4774.
8. Molnar, C., 2020. *Interpretable machine learning*. Lulu. com.
9. Zhang, Y., Zheng, X.T., Zhang, X., Pan, J. and Thean, A.V.Y., 2024. Hybrid integration of wearable devices for physiological monitoring. *Chemical Reviews*, 124(18), pp.10386-10434.
10. Dang, T., Spathis, D., Ghosh, A. and Mascolo, C., 2023. Human-centred artificial intelligence for mobile health sensing: challenges and opportunities. *Royal Society Open Science*, 10(11), p.230806.

11. Sreeraj, V.S., Parlikar, R., Bagali, K., Shekhawat, H.S. and Venkatasubramanian, G., 2024. Advancing Data Science: A New Ray of Hope to Mental Health Care. Exploration of Artificial Intelligence and Blockchain Technology in Smart and Secure Healthcare, pp.199-233.
12. Razavi, M., Ziyadidegan, S., Mahmoudzadeh, A., Kazeminasab, S., Baharlouei, E., Janfaza, V., Jahromi, R. and Sasangohar, F., 2024. Machine learning, deep learning, and data preprocessing techniques for detecting, predicting, and monitoring stress and stress-related mental disorders: Scoping review. *JMIR Mental Health*, 11, p.e53714.
13. Lundberg, S., 2017. A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.
14. Sethia, D. and Indu, S., 2024. Optimization of wearable biosensor data for stress classification using machine learning and explainable AI. *IEEE Access*.
15. Islam, M.M., Hassan, S., Akter, S., Jibon, F.A. and Sahidullah, M., 2024. A comprehensive review of predictive analytics models for mental illness using machine learning algorithms. *Healthcare Analytics*, p.100350.
16. Zhang, A., Wu, Z., Wu, E., Wu, M., Snyder, M.P., Zou, J. and Wu, J.C., 2023. Leveraging physiology and artificial intelligence to deliver advancements in health care. *Physiological Reviews*, 103(4), pp.2423-2450.