IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Predictive And Proactive Student Mental Health Monitoring Using Deep Neural Networks

[1]V. Jabili, [2]M. Divya Sri Lakshmi, [3]P. Dev, [4] K. Swarupa Rani

[1-3] IV B. Tech, Department of Information Technology, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, India.

[4] Assistant Professor, Department of IT, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, India

Abstract: The appearance of digital stressors has created the urgent need to address the issue of mental health through the development of mental health monitoring systems that are proactive, as the current clinical approaches are mostly reactive and are rarely conducted at regular intervals, making it hard to keep up with the dynamic nature of mental health. This paper seeks to present a novel framework for predictive behavioral analytics using eXtreme Gradient Boosting (XGBoost) and Shapley Additive Explanations (SHAP) for the purpose of decision-making. The framework is different from other "black box" models because it can attribute specific behavioral events, such as sleep deprivation or excessive use of social media, using local feature attribution. The model also considers a temporal baseline of 15 days, which helps it provide personalized risk assessments for specific users rather than using averages for all users. The results of the experiments show a classification accuracy of 92.5

Index Terms - Explainable AI (XAI), XGBoost, SHAP, DNN, Mental Health, Behavioral Analytics, Digital Phenotyping

I. INTRODUCTION

The pervasiveness of digital technology in everyday life has significantly changed the human psychological well-being environment. Although the current level of connectivity has greatly improved the quality of global communication, it has also, in turn, introduced new behavioral stressors. The "Always-On" culture, which is characterized by irregular sleep, digital consumption, and the blurring of work and personal life, has resulted in the increase of subclinical psychological states, including anxiety and digital burnout.

Current methodologies in mental health assessment, including the Patient Health Questionnaire (PHQ-9) and the Generalized Anxiety Disorder (GAD-7) scale, are heavily dependent on retrospective reporting. Therefore, these methodologies are inherently reactive, meaning that they tend to detect psycho logical conditions only after the individual has reached the state of functional impairment. These traditional mental health assessment methodologies are also ineffective in detecting the high-frequency, dynamic behavioral patterns of the modern individual. There is a significant need to establish proactive mental health monitoring systems that are capable of detecting subtle changes in the behavioral patterns of individuals before the conditions are manifested into clinical conditions.

The advent of Machine Learning (ML) has catalyzed the transition toward predictive healthcare, leveraging multidimensional behavioral data to classify psychological states. However, the application of complex ML models in the healthcare domain poses the so-called “Black Box” problem. In fact, typical ML models such as Deep Neural Networks and Ensembles may trade the interpretability of the model for predictive performance. In the context of a mental health application, for example, a user may be informed that he/she is at a “High Risk” for anxiety without providing the rationale for this determination. This may cause the user further stress and erode the trust between the user and the system.

To address these critical gaps, this paper presents a proactive predictive behavioral analytics framework powered by eXtreme Gradient Boosting (XGBoost) and Shapley Additive Explanations (SHAP). The proposed system eschews static population averages in favor of a personalized “Temporal Baseline,” comparing a user’s current behavioral metrics (e.g., sleep hygiene, social interaction, screen time) against their own 15-day rolling average. By integrating TreeSHAP, the system transcends basic classification by providing local feature attribution, isolating the precise behavioral trigger driving the risk score.
The primary contributions of this research are as follows:

- **Temporal Feature Engineering:** The development of a dual-vector feature pipeline that analyzes both absolute behavioral scores and their relative deviation from a personalized historical baseline.
- **Explainable AI Integration:** The implementation of a TreeSHAP-based explanation module that mathematically quantifies the contribution of individual digital phenotypes to the overall stress prediction in real-time.
- **Prescriptive Intervention Engine:** The design of a dynamic, rule-based recommendation system that maps SHAP-identified triggers to targeted, micro-behavioral wellness tasks.

II. Literature Review

A. Traditional Methods vs. Digital Phenotyping

Traditionally, mental health monitoring is done using established psychometric scales like the “Patient Health Questionnaire” (PHQ-9) and the “Generalized Anxiety Disorder” (GAD-7) scale. Although these are reliable, they are, at their core, limited in their ability to monitor mental health due to their “retrospective” nature, making them prone to “recall bias.” Humans are often incorrect about their emotional history according to their current emotional state. “On the other hand, ’Digital Phenotyping’ exploits the availability of ’digital data’ arising from human-computer interactions on a daily basis. It tracks ’passive’ and ’active’ variables such as ’sleep duration,’ ’physical mobility,’ and ’screen time’ as a means to quantify mental health states objectively.”

B. Machine Learning for Predictive Healthcare

With the advent of Machine Learning (ML) technologies in behavioral analytics, the focus has moved from descriptive healthcare to predictive healthcare. In the past, basic Machine Learning algorithms like Logistic Regression and Support Vector Machines (SVM) were employed. Nevertheless, human behavior is a non-linear phenomenon; for example, the effects of sleep deprivation may be exponentially magnified if com pounded by high levels of social isolation.

More recent research has shown the potential for the application of ensemble learning techniques like Random Forest (RF) and Deep Neural Networks (DNN). Nevertheless, for tabular behavioral datasets, Gradient Boosted Decision Trees (GBDT), specifically eXtreme Gradient Boosting (XGBoost), outperform DNNs by a wide margin. This is owing to the fact that XGBoost incorporates sparse split evaluation and strong L1/L2 normalization, making it particularly well-suited for handling behavioral datasets where certain daily metrics may be missing.

C. The Evolution of Explainable AI (XAI)

The primary barrier to the clinical adoption of predictive ML models is the "Black Box" phenomenon. A highly accurate Neural Network is of little clinical utility if its diagnostic rationale cannot be audited by a human practitioner. To fill this gap, various Explainable AI (XAI) approaches have been proposed. Although early "post-hoc" explainers, such as Local Interpretable Model-agnostic Explanations (LIME), were able to produce localized explanations, they were not mathematically sound.

In this work, the SHAP (Shapley Additive Explanations) approach, founded on cooperative game theory, is adopted because of its guarantee of both local accuracy and missing ness, meaning that the sum of the feature values equals the prediction minus the baseline expected value.

III. Proposed Methodology

The proposed system, namely MindApp, has been developed on the basis of decoupled architecture, also known as the microservices architecture, to ensure that high-speed real time inferences are achieved without compromising the user experience.

A. System Architecture

It has been proposed that the proposed system architecture would be divided into three different layers: client presentation, API orchestration, and AI inference engine.

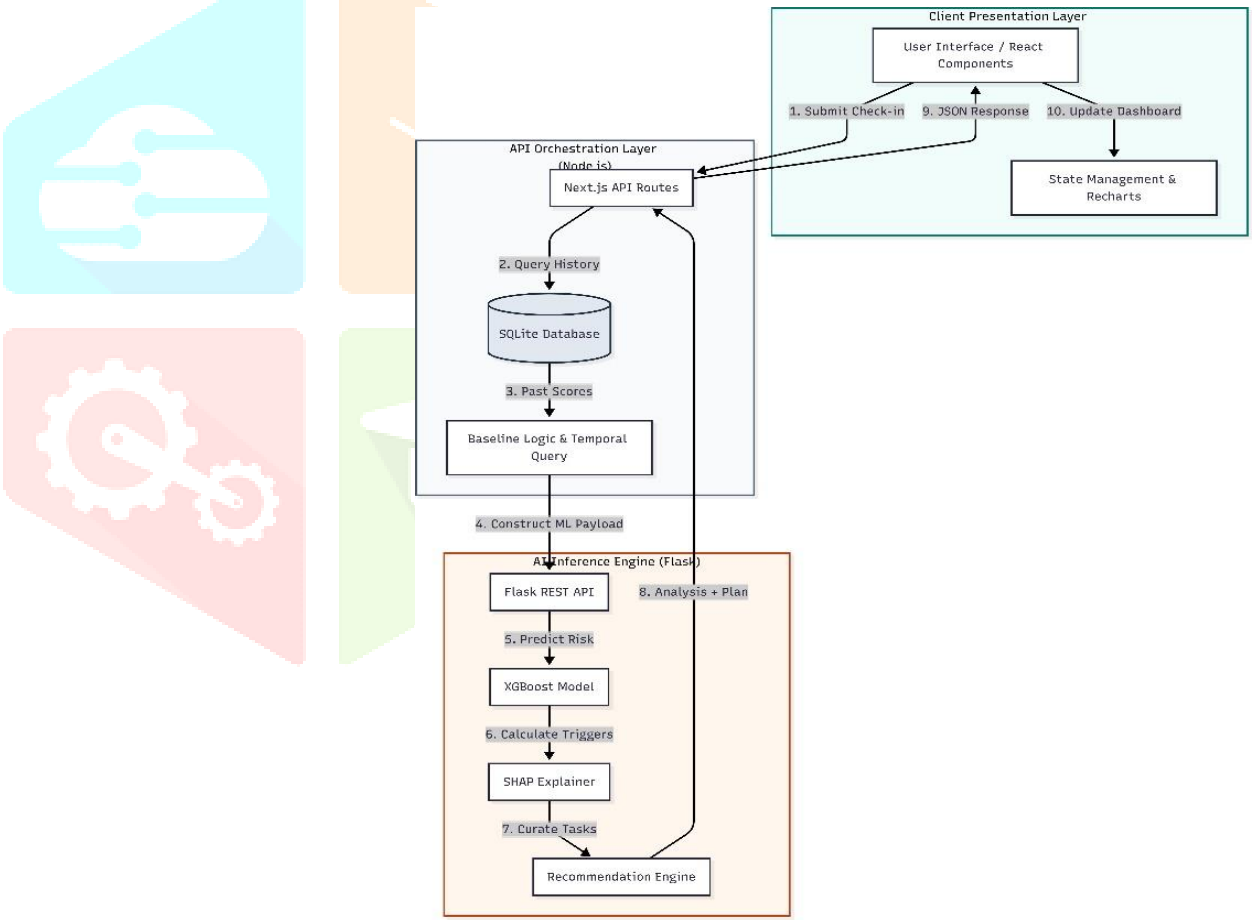

Fig. 1: High-Level System Architecture detailing the unidirectional flow from the Next.js Client, through the SQLite Temporal Database, to the Flask XGBoost Engine.

As shown in Fig. 1, the frontend layer uses Next.js with Tailwind CSS for a dynamic interface, and user behavior metrics are collected on a daily basis, which are then fed into a middle layer that uses Node.js to connect a SQLite database running on the local machine. This database utilizes a temporal query with the user's prior information and sends this data as a payload to a Python Flask-based microservice containing the serialized XGBoost objects and TreeSHAP explainer objects.

B. Temporal Feature Engineering

A fundamental flaw in generic wellness applications is the reliance on static population baselines. To achieve true personalization, the proposed system transforms raw inputs into a "Dual-Vector" format.

Let F represent the set of behavioral domains (Sleep, Social, Phone Usage, Leisure, Productivity). For each feature f ∈ F, the user inputs a daily raw score $S_{current}^{(F)}$. The system queries the SQLite database to calculate the 15-day rolling average for that specific user, denoted as $\mu_{15}^{(F)}$.
The algorithm then engineers a difference feature (Delta):

$$\Delta^{(F)} = S_{current}^{(F)} - \mu_{15}^{(F)} \quad (1)$$

The final 10-dimensional feature vector X fed into the XGBoost model consists of both the absolute scores and their respective temporal deviations:

$$X = [S_{current}^{(1)}, \dots, S_{current}^{(5)}, \dots, \Delta^{(1)}, \dots, \Delta^{(5)}] \quad (2)$$

This allows the model to prioritize the velocity of behavioral change over static values.

C. XGBoost Classification Framework

The core predictive engine utilizes the XGBoost algorithm. For a given dataset with n examples and m features D = $\{(xi, yi)\}$, a tree ensemble model uses K additive functions to predict the output:

$$\hat{y}_i = \sum_{k=1}^{k} f_k(x_i), \qquad f_k \in F \quad (3)$$

where F is the space of regression trees. To learn the set of functions, we minimize the following regularized objective:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum\_k \Omega(f_k) \quad (4)$$

The term $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\|w\|^2$ penalizes the complexity of the model, preventing the AI from overfitting to idiosyncratic user data and ensuring robust generalization across diverse psychological profiles.

IV. Explainable AI Framework (SHAP)

To reduce the level of opaqueness with regards to the XGBoost classifier, the proposed framework incorporates the SHAP algorithm. This is based on game theory, where a value is assigned to a feature for a given prediction.

A. Mathematical Formulation

The SHAP explanation model utilizes an additive feature attribution method, defining the explanation model g as a linear function of binary variables:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i' \quad (5)$$

where $z' \in \{0,1\}^M$ represents the coalition vector (whether a feature is observed or missing), M is the number of input features, φ_0 is the base expected value of the model, and φ_i is the Shapley value for feature i. For tree-based models like XGBoost, we utilize the Tree SHAP algorithm, which reduces the computational

complexity from exponential to polynomial time $O(TLD^2)$, where T is the number of trees, L is the maximum number of leaves, and D is the maximum depth.

B. Local and Global Interpretability

The system employs SHAP at two different levels of interpretability. Specifically, it is employed at both local and global levels. The local level is also known as real-time triggers, whereas global-level interpretability is also known as population insights. The system employs Tree SHAP for local explanations, i.e., for each user's check-in, it calculates the precise push or pull force of each feature. Once it predicts "Moderate Risk" for a user, it chooses the feature with the highest value of φ_i, for instance, "Sleep Duration," and displays it as the main trigger on the UI dashboard. The global level of interpretability is employed using the sum of all absolute SHAP values for all users in the validation set.

V. RESULTS AND DISCUSSION

A. Experimental Setup and Classification Metrics

The predictive model was evaluated using a stratified 80/20 train-test split. The primary objective was to maximize recall for the "Severe" risk category to ensure that users experiencing acute psychological distress are consistently identified.

As detailed in Table I, the proposed XGBoost architecture outperformed traditional baseline models.

Model Architecture	Accuracy	Precision	Recall (Severe)
Random Forest (RF)	88.2%	87.5%	86.0%
Deep Neural Network (DNN)	89.5%	88.1%	89.2%
XGBoost (Proposed)	**92.5%**	**91.0%**	**94.0%**

TABLE I: Comparative Analysis of Model Architectures

The Random Forest model struggled with class overlap, frequently misclassifying "Mild" stress due to its averaging nature. The DNN required substantially higher inference latency, making it less suitable for real-time web deployment. XGBoost achieved the optimal balance of high accuracy (92.5%) and low latency, with a critical 94.0% recall in identifying severe risk profiles.

B. SHAP Trigger Analysis and Validation

The statistical validation of the SHAP values revealed highly distinct behavioral thresholds at both the population macro level and the individual micro-level.

1) Global Population Insights

To understand the overarching drivers of mental health risk, we aggregated the feature attributions across a randomly sampled validation dataset. As illustrated in Fig. 2, the global SHAP summary plot ranks the behavioral metrics by their absolute impact on the model's output.

The global visualization confirms two primary psychological hypotheses:

1) The Sleep Threshold: Sleep metrics exhibit the widest SHAP dispersion. Specifically, a negative temporal deviation in sleep (Δsleep) acts as the most significant dominant trigger pushing users into the "Moderate" or "Severe" classifications.

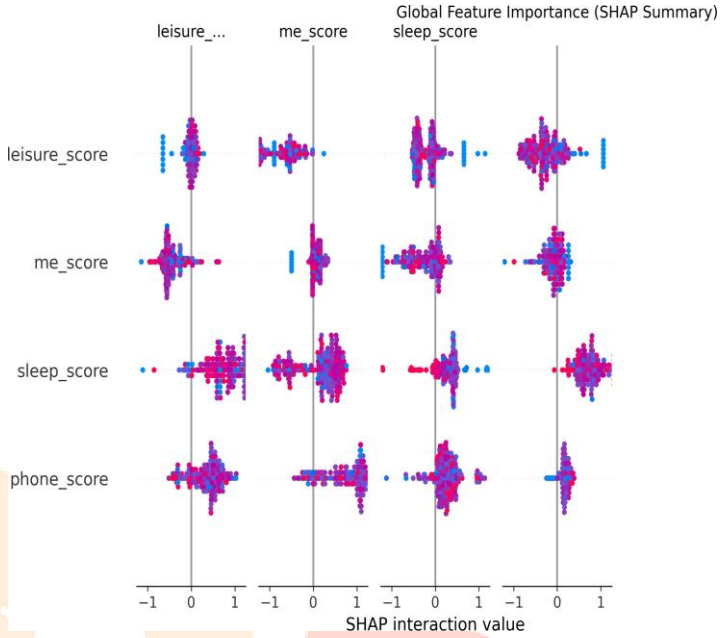

Fig. 2: Global SHAP Summary Plot indicating the relative impact and distribution of behavioral features across the test population.

2) Digital Displacement: High phone usage consistently yields positive SHAP values (increasing risk), particularly when it acts as a coping mechanism for low social interaction, validating the "Social Buffering" effect.

2) Local Individual Case Study

While global metrics validate the model's overall logic, proactive intervention requires individual transparency. Fig. 3 demonstrates a local SHAP waterfall plot for a specific user whose check-in was flagged as "Severe Risk."

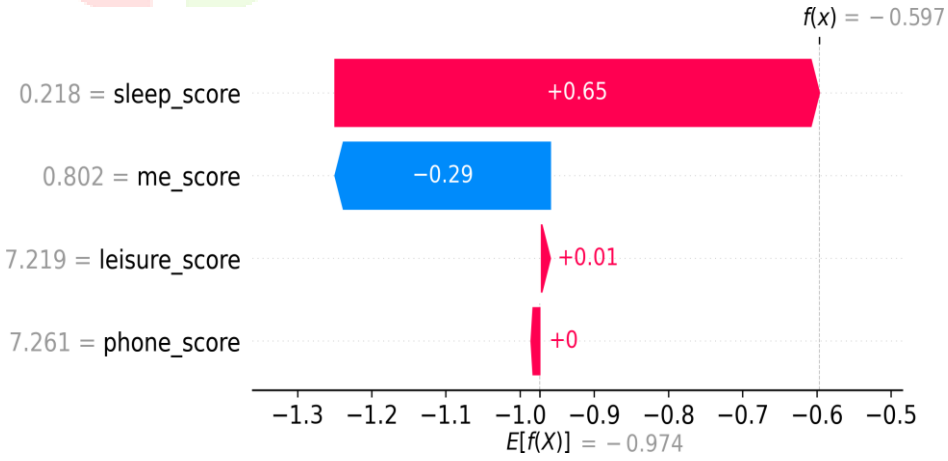

Fig. 3: Local SHAP Waterfall Plot demonstrating how specific behavioral deviations push the baseline expected value toward a Severe Risk classification for a single user.

Unlike the summary plot, the waterfall chart provides a step-by-step mathematical rationale for a single prediction. Starting from the baseline expected value (the dataset average for this class), the plot explicitly

maps how the user's specific behaviors accumulate to reach the final risk probability. In this case, the visualization clearly isolates the exact magnitude by which poor sleep and high digital consumption pushed the score upward.

In our proposed architecture, this localized SHAP output is directly parsed by the recommendation engine to generate targeted wellness interventions (e.g., prioritizing "Digital Sun-set" tasks over generic advice), ensuring the AI's feedback is entirely personalized.

VI. Conclusion

A. Conclusion

This research has offered a proactive and privacy-focused behavioral analytics framework with the aim of addressing the gap between abstract machine learning predictions and actionable mental health interventions. By avoiding population averages and opting for a personal 15-day time window instead, the proposed system has shown high sensitivity to changes in individual behavior. By leveraging the eXtreme Gradient Boosting algorithm, a high accuracy of 92.5% and a critical 94.0% recall rate for severe psychological distress were realized.

Moreover, the TreeSHAP explainability method is integrated, which effectively solves the "Black Box" problem that is commonly faced with complex medical AI systems. This is done by mapping the user's multi-dimensional behavioral data onto a localized, mathematically justified feature attribution. This XAI method not only builds trust with the user but also enables them to effectively address the underlying micro behaviors that contribute to their sub-clinical stress, such as digital displacement and sleep deprivation.

B. Future Work

Although the current system design is based on user submitted daily check-ins, future work will be centered on reducing survey fatigue through automated data ingestion. This will be achieved by integrating wearable Internet of Things (IoT) devices, such as smartwatches, which will allow for passive data collection of the user's physiological digital phenotype, including Heart Rate Variability (HRV) and sleep staging.

Furthermore, we also suggest the use of Privacy Preserving Federated Learning as a technique, which would enable the XGBoost model to be trained on the edge devices themselves, sending only the weights to a central location for enhanced accuracy without compromising the privacy of the individuals involved. Finally, the use of Large Language Models (LLMs) in interpreting the SHAP triggers in a conversational context represents a highly exciting future direction for automated digital counseling tools.

References

[1] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems 30, 2017, pp. 4765–4774.

[2] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[3] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," in Proceedings of the 22nd ACM SIGKDD International Conference, 2016, pp. 1135–1144.

[4] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2022. [Online]. Available: https://christophm. github.io/interpretable-ml-book/

[5] A. Holzinger, "Explainable AI (ex-AI)," Informatik-Spektrum, vol. 41, no. 2, pp. 138–143, 2018

[6] T. R. Insel, "Digital phenotyping: technology for a new science of behavior," JAMA, vol. 318, no. 13, pp. 1215–1216, 2017.

[7] J. Torous, M. V. Kiang, J. Lorme, and J. P. Onnela, "New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research," JMIR Mental Health, vol. 3, no. 2, e16, 2016.

[8] D. C. Mohr, M. Zhang, and S. M. Schueller, "Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning," Annual Review of Clinical Psychology, vol. 13, pp. 23–47, 2017.

[9] R. Wang et al., "StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students using Smart phones," in Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2014, pp. 3–14.

[10] S. Saeb et al., "Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study," Journal of Medical Internet Research, vol. 17, no. 7, e175, 2015.

[11] N. C. Jacobson, W. E. Bentley, A. Walton, et al., "Digital biomarkers of depression and anxiety in smart devices," Translational Psychiatry, vol. 10, no. 68, 2020.

[12] A. B. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of methods and applications," Psychological Medicine, vol. 49, no. 9, pp. 1426–1448, 2019.

[13] A. Thieme, J. Belgrave, and G. Doherty, "Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems," ACM Transactions on Computer-Human Interaction, vol. 27, no. 5, pp. 1 53, 2020.

[14] G. Cho, K. Yano, and T. Watanabe, "Review of Machine Learning Algorithms for Diagnosing Mental Illness," Psychiatry Investigation, vol. 16, no. 4, pp. 262–269, 2019.

[15] D. Bzdok and A. Meyer-Lindenberg, "Machine Learning for Precision Psychiatry: Opportunities and Challenges," Biological Psychiatry, vol. 83, no. 3, pp. 223–230, 2018.

[16] F. Corponi et al., "Novel Machine Learning Approaches to Personalized Medicine in Psychiatry," Psychiatric Clinics of North America, vol. 43, no. 3, pp. 393–404, 2020.

[17] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The false hope of current approaches to explainable artificial intelligence in health care," The Lancet Digital Health, vol. 3, no. 11, e745–e750, 2021.

[18] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Ex plainability for artificial intelligence in healthcare: a multidisciplinary perspective," BMC Medical Informatics and Decision Making, vol. 20, no. 1, pp. 1–9, 2020.

[19] E. F. Haghish, "Digital phenotyping and objective markers of mental health: The machine learning promise," Digital Health, vol. 6, 2020.

[20] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.