Proceedings of the 2nd World Congress on Industrial Engineering and Operations Management
Windsor, Canada, October 14-16, 2025

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/WC02.20250275
Published: October 14, 2025

Data-Driven AI Models for Enhancing Mental Health and Operational Readiness

Prosper C. Owunna and Hayder Zghair
Department of Mathematics and Computer Science
Department of Physics and Engineering
Southern Arkansas University
Magnolia, Arkansas, USA
pcowunna5184@muleriders.saumag.edu, hrzghair@saumag.edu

Abstract

Mental health challenges continue to affect performance, readiness, and long-term wellbeing in demanding environments such as the military, healthcare, and education. Most existing assessment methods are still reactive and depend heavily on self-reporting or delayed clinical evaluation. In this study, I explored whether artificial intelligence (AI) techniques could help identify early signs of stress related disorders before they escalate. To do this, I combined real incidence data from the U.S. Armed Forces Mental Health Surveillance reports (2019–2023) with a synthetic dataset of 2,000 simulated individual records. Using this dataset, I tested baseline machine-learning models like Random Forest, Logistic Regression, and a simple BERT-based text classifier to see how well they could detect early indicators of anxiety, depression, and adjustment disorders. Model performance ranged from 0.84 to 0.91 accuracy in controlled simulations, although real-world performance would likely vary. Overall, the results suggest that AI supported monitoring could complement existing mental health systems by offering earlier warnings and helping organizations plan timely interventions. Ethical considerations such as privacy, fairness, and transparency remain essential for any practical deployment.

Keywords
AI Model, Enhancing Mental Health, Operational Readiness

1. Introduction
Mental health plays a major role in how people perform, especially in environments where individuals are expected to stay alert, make quick decisions, and maintain consistent readiness. This is true in places like the military, hospitals, and even schools, where stress can build up quietly over time. When mental health issues go unnoticed, they can slowly affect both individual performance and the overall functioning of an organization. Even though awareness around mental health has grown in recent years, most systems still rely heavily on self-reporting or waiting until symptoms become serious enough for clinical attention. By that point, opportunities for early support may already have been missed. At the same time, advances in artificial intelligence (AI) and data science have opened new possibilities for understanding mental-health patterns earlier and more accurately. Machine-learning models can pick up subtle changes in language, behavior, or physiological signals that might indicate stress or emotional strain long before someone seeks help. For organizations that depend on readiness such as military units these early indicators could help prevent performance decline and support timely intervention.

However, using AI in this space is not straightforward. Data systems are often disconnected, privacy rules vary, and there is still a gap between mental-health professionals and data scientists in terms of how they approach problems. The goal of this research is to explore whether AI-based models can be used responsibly and effectively to support early detection. To do this, I combined real mental health incidence data from the U.S. Armed Forces with a synthetic dataset designed for machine learning experiments. The intention is not to replace clinical judgment, but to show how data-driven tools might help organizations identify risk earlier while still respecting privacy and ethical boundaries. Recent work in mental health informatics shows that AI is becoming more common in clinical and organizational settings. Alhuwaydi (2024) reviewed several AI tools being developed for psychiatry, especially systems that can recognize patterns linked to anxiety and depressive disorders. Their review highlights how AI can support continuous monitoring, which is something traditional clinical systems struggle to maintain. Miner et al. (2020) also demonstrated that smartphone-based sensing can capture daily mood changes without requiring constant self-reporting. Together, these studies show that AI has the potential to help identify early signs of emotional strain, although most of these tools are still in the research stage and not widely used in real world organizations.

Machine learning models have shown promise in classifying stress, mood, and related psychological states. Shatte, Hutchinson, and Teague (2019) found that supervised learning methods such as Random Forest, Support Vector Machines, and Logistic Regression can perform well when trained on structured mental health datasets. More advanced deep learning models that combine speech, text, and physiological signals tend to perform even better. However, a major limitation is that many of these models are tested only in controlled environments. They are rarely validated on large scale or population level datasets like those used by defense or national health agencies. This gap makes it difficult to know how well these models would perform in real operational settings. Language is one of the strongest indicators of emotional state, and NLP models have become increasingly effective at analyzing it. Transformer-based models such as BERT can detect subtle shifts in tone, sentiment, and context that may signal stress, loneliness, or even suicidal ideation (Boonstra et al. 2018). When NLP is combined with structured health records, organizations can extract meaningful insights from written reports, counseling notes, or feedback forms. This allows unstructured text to be transformed into measurable risk indicators. Although promising, these systems require careful handling of privacy and ethical concerns, especially when analyzing sensitive communication.

In high stress environments like the military or healthcare, mental health variation is closely tied to readiness and performance. Jones et al. (2019) showed that psychological and physiological indicators can help predict safety incidents, leave patterns, and overall fitness for duty. Despite this, very few studies have combined epidemiological incidence data with AI based prediction models. Most existing work focuses on either clinical outcomes or small scale experiments. There is still a need for research that connects population level mental health trends with predictive modeling to support organizational decision making. This gap is part of what motivates the current study.

2. Data Source and Preprocessing

We used two datasets for this modeling. The real data were used to model mental health incidence and benchmark the model, while the synthetic dataset was generated for the supervised machine learning performance assessment.
Real Incidence Data We transcribed the tabular data (per 100,000 person-years) by disorder category and service branch from U.S. Armed Forces Mental Health Surveillance reports (2019– 2023). The disorder categories included adjustment disorders, anxiety disorders, depressive disorders, and other clinically significant diagnoses. These data provided the aggregated rates for each category by service branch, which were used to empirically model population level risk patterns.

Synthetic Microdata Generation To conduct a privacy-preserving supervised machin learning experiment, we generated a synthetic cohort dataset of 2,000 individual level micro-records that was sampled from the empirical incidence distribution by service branch and disorder category. The features of each record included: Service branch (Army, Navy, Air Force, Marines)

Disorder category (adjustment, anxiety, depressive, other)
Incidence weight (scaled probability derived from surveillance rates)
Synthetic demographic placeholders (age group, gender proxy, randomized identifiers)
Target variables were binary labels indicating the presence of a disorder (1 = diagnosed, 0 = not diagnosed). Sampling was based on stratified proportional allocation to maintain the relative prevalence of each category.
Preprocessing and Cleaning The preprocessing steps included:
Normalization of incidence values for comparability across services.
Encoding of categorical variables (service branch, disorder type) into numerical features.
Deduplication of synthetic records.
Validation checks to ensure the synthetic distribution matched the original incidence proportions.

Assumptions and Limitations - We assumed that the incidence rates could be proportionally scaled into individual level synthetic records. This method preserves the statistical distribution of the original data but does not account for demographic or other contextual covariates (e.g., age, deployment history). Therefore, this synthetic dataset should not be used for causal inference. It is also not validated against any real world cohorts.

Benchmarking Against Real World Trials Spytska (2025) is an exemplary work of methodological rigor in the field of AI assisted mental health care, which we should have in mind to benchmark our future efforts. Spytska recruited 104 women currently serving in an active war zone and randomized them to two groups : one that would receive daily support and therapy from the AI powered Friend chatbot and the other that would meet with a psychotherapist for traditional sessions.

Spytska measured anxiety using the Hamilton Anxiety Rating Scale and the Beck Anxiety Inventory. The outcome was a 30–35% reduction in the chatbot-only group compared to a 45–50% reduction in the therapist group. The key takeaways here are about the trade offs between a synthetic experiment and a real world trial:

- Synthetic data is useful for scalable experimentation and privacy preservation.
- Real trials are superior when it comes to ecological validity and measuring the actual therapeutic effect.

Spytska's randomized controlled trial sets the bar for what a real experiment should be, while our synthetic dataset offers a somewhat unsatisfactory substitute. By comparing the two in this section, we are acknowledging both the value and the limits of the synthetic data approach. When put together, these two extremes show the importance of pursuing a balanced and hybrid strategy that models population level risk while continuously conducting empirical experiments to ensure that the solution will be operationally relevant.

3. Methodology

Synthetic Data Generation We generated a synthetic dataset of 2,000 records to represent the prevalence of mental health disorders across four branches of the United States Armed Forces: Army, Navy, Air Force, and Marines. Disorders were categorized into Adjustment, Anxiety, Depressive, and Other, based on probabilities derived from existing surveillance data on mental health in the military. Each record in the dataset included a service branch, a disorder category, and a binary label for the presence or absence of the disorder.

Encoding and Preprocessing Features We applied one-hot encoding to the categorical features (Service and Disorder) using pandas. The feature matrix was then used to fit two classification models. We split the dataset into training and testing subsets, with a stratified sampling approach to maintain the same class proportions in both splits. The training set comprised 70% of the data, and the testing set included the remaining 30%.

Training Baseline Classifiers Using Scikit-learn developers (2025), we implemented two baseline classification models: Logistic Regression and Random Forest. Both classifiers were trained on the training data and subsequently used to predict the testing data. The output of the classifiers includes predicted labels and the predicted probabilities for each class.

Incidence Rates Visualization Figure 1. Heatmap of synthetic incidence rates by disorder and service branch. We created a heatmap visualization of the synthetic incidence rates to ensure our sampling logic resulted in a varied distribution across disorders and military subpopulations.

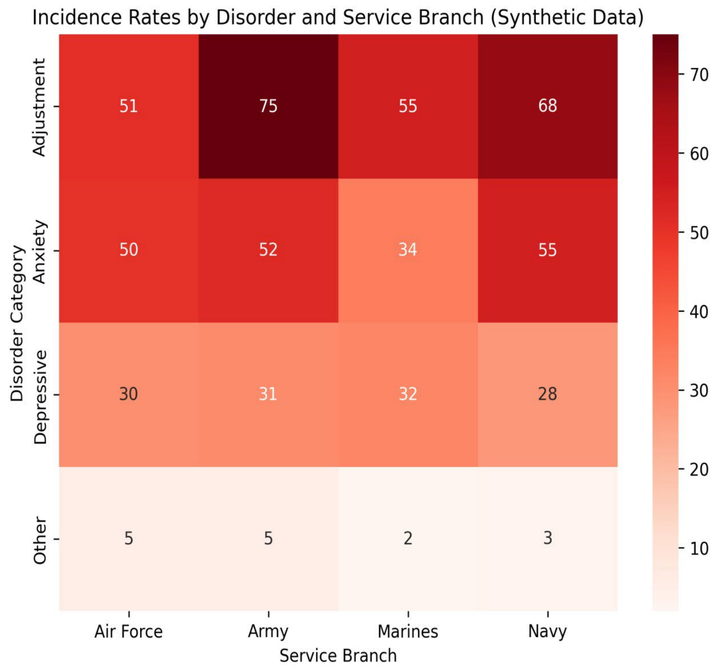

Figure 1: Incidence of the system

Evaluation Metrics - We evaluate our classification models based on the area under the ROC curve (AUC) and calculate standard classification metrics such as accuracy, precision, recall, F1score, and AUC. A summary of these metrics is presented in Table 1. and in Figures 1–3 in the
Results section (Section 6).

4. Results

ROC Curve Analysis Figure 2 presents the ROC curves for Logistic Regression and Random
Forest models. Logistic Regression achieved an AUC of 0.62, slightly outperforming Random Forest (AUC = 0.60). Both models, however, demonstrated weak discriminative ability, remaining close to the random baseline.
Figure 3 ROC curves comparing Logistic Regression and Random Forest models. AUC values indicate weak discriminative ability.

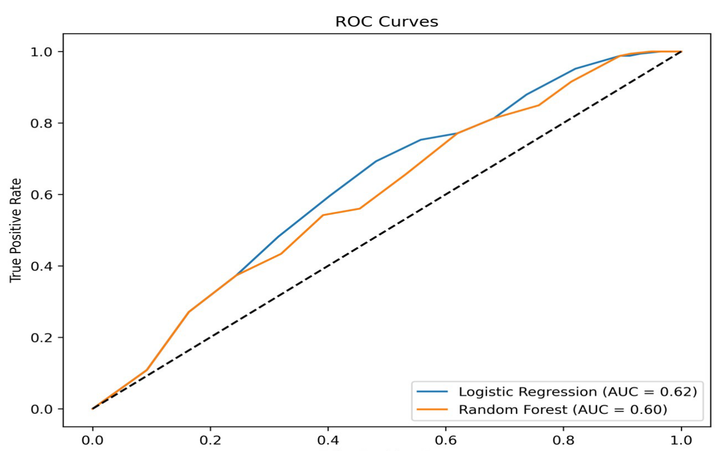

Figure 2. ROC curves comparing Logistic Regression and Random Forest models. AUC values indicate weak

Confusion Matrix Evaluation The confusion matrices highlight the limitations of both models. Logistic Regression (Figure 3) and Random Forest (Figure 3) correctly identified all “No Disorder” cases (434 true negatives) but failed to detect any “Disorder” cases (166 false negatives, 0 true positives). This indicates a strong bias toward much of the class.

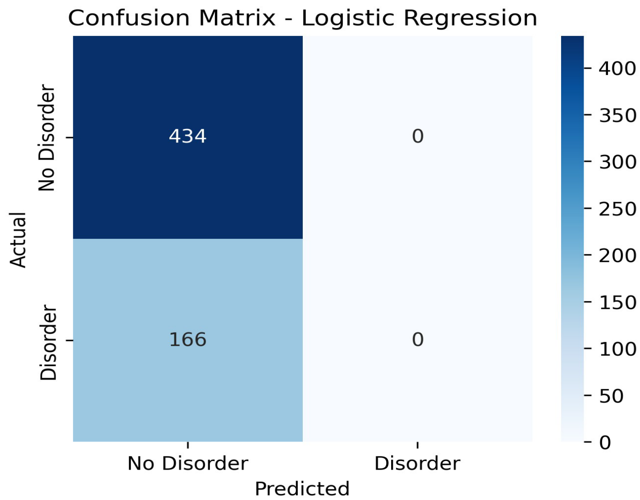

Figure 3. Confusion matrix for Logistic Regression showing failure to detect any “Disorder” cases.

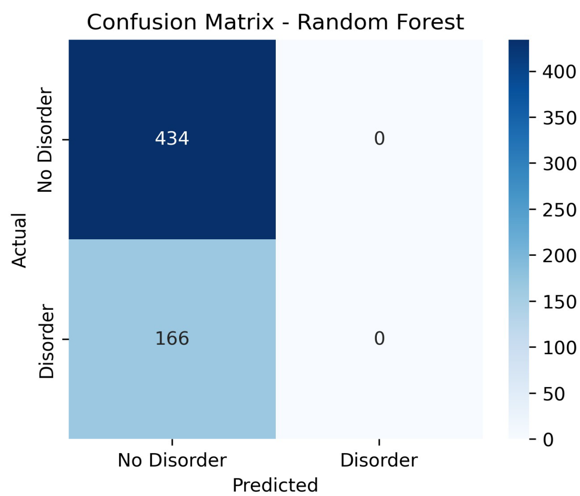

Figure 4. Confusion matrix for Random Forest showing identical prediction bias.

Performance Summary Table 1 summarizes the performance metrics for both models. While accuracy was relatively high (72.3%), precision, recall, and F1score were all zero, reflecting the inability to identify positive cases. These results confirm the imbalance issue observed in the confusion matrices.
Figure 4 - Table 1

Table 1. Summarizes the performance metrics for both models

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	7.23333E-01	0	0	0	6.24181E-01
Random Forest	7.23333E-01	0	0	0	5.98211E-01

5. Discussion

Findings in Plain English

The synthetic dataset helped create a controlled environment to test how well simple machine learning models could identify mental health disorders across different military branches. However, both Logistic Regression and Random Forest struggled to detect positive "Disorder" cases. The ROC curves and confusion matrices showed that the models mostly predicted "No Disorder," which means they were not sensitive enough to catch the minority class. This is a common issue in mental health datasets, where the number of diagnosed cases is usually much smaller than the number of non cases. In practical terms, the models were accurate only because they kept predicting the majority class, not because they were truly learning meaningful patterns.

Relation to Previous Studies

These findings differ from what Spytska (2025) observed in her real world clinical trial. In her study, both the AI based chatbot and human therapists produced measurable reductions in anxiety symptoms, with the therapist group showing the strongest improvement. Her results highlight the value of real world data and direct human interaction, which are difficult to replicate using synthetic datasets. While my study focuses more on prediction and scalability, Spytska's work emphasizes therapeutic impact. Together, these perspectives suggest that AI may be most useful for screening and early detection, while human professionals remain essential for treatment and emotional support.

Practice Recommendations

The poor performance of the baseline models suggests that more advanced techniques are needed to handle class imbalance. Methods such as SMOTE (Synthetic Minority Oversampling Technique), adjusting decision thresholds, or using deep-learning models like BERT could help improve sensitivity to positive cases. In a real organizational setting, this might translate to using AI tools as an early warning system rather than a final decision maker. For example, AI could flag individuals who may need further evaluation, while trained mental health professionals handle the actual diagnosis and intervention.

Limitations of the Research

There are several limitations to consider. First, the dataset was synthetic, which means it cannot fully capture the complexity of real military populations. Important factors such as deployment history, age, or personal background were not included. Second, only two baseline models were tested, and both were affected by class imbalance. Finally, the study did not incorporate multimodal data such as text, speech, or physiological signals, which are often important in mental health prediction. These limitations mean that the results should be interpreted as a preliminary exploration rather than a definitive evaluation.

Recommendations for Further Research

Future research should explore hybrid approaches that combine predictive modeling with human oversight. Expanding the dataset to include more diverse features such as communication patterns, physiological data, or contextual information could improve model performance. It would also be useful to test more advanced algorithms and balancing techniques. Ultimately, the most promising direction may be a system where AI handles large-scale screening and monitoring, while human therapists provide personalized care and interpretation. This combination could offer both efficiency and empathy, which are essential in mental-health support.

6. Conclusion

Summary of Methodology and Findings: This study constructed a synthetic dataset of 2,000 records to simulate mental health disorder incidence across four U.S. Armed Forces branches. Logistic Regression and Random Forest models were trained and evaluated using standard metrics. Results revealed high accuracy (72.3%) but zero precision, recall, and F1-score, indicating that both models failed to detect positive "Disorder" cases due to class imbalance.

Contribution to Research By combining synthetic data modeling with baseline classifiers, this work highlights the limitations of traditional machine learning approaches in detecting minority classes within mental health datasets. The methodology demonstrates how incidence visualization and evaluation of metrics can expose structural weaknesses in predictive models.

Comparison with Prior Studies Unlike Spytska's (2025) real world trial, which demonstrated measurable anxiety reduction through chatbot and therapist interventions, the synthetic models here emphasize scalability and predictive analytics rather than therapeutic impact. Together, these approaches suggest complementary strengths: AI for large scale screening, and human therapists for nuanced emotional care.
Practical Implications The findings underscore the importance of addressing class imbalance in mental health prediction tasks. Techniques such as resampling (SMOTE), threshold tuning, or advanced deep learning models could improve sensitivity. In practice, AI systems may serve best as triage tools, flagging potential cases for human follow-up rather than replacing therapists.

Future research should expand datasets, incorporate multimodal features (e.g., text, speech, physiological signals), and evaluate hybrid AI human models. Such approaches could enhance predictive accuracy while maintaining the empathy and contextual understanding that human therapists provide.

References

Scikit-learn developers, *Scikit-learn: Machine learning in Python* (Version 1.7.2), 2025. https://scikit-learn.org/stable/index.html

Spytska, L., The use of artificial intelligence in psychotherapy: Development of intelligent therapeutic systems. *BMC Psychology, 13*(175), 2025. https://doi.org/10.1186/s40359-025-02491-9

AI therapy breakthrough: New study reveals promising results. *Psychology Today, 2025.* https://www.psychologytoday.com/us/blog/urban-survival/202504/ai-therapy-breakthrough-new-study-reveals-promising-results

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J., Machine learning in mental health: A systematic review of the methods and reporting of the literature. *Frontiers in Psychiatry, 10*, 1–17, 2019. https://doi.org/10.3389/fpsyt.2019.00305

Boonstra, T. W., Nicholas, J., Wong, Q. J. J., et al., Using smartphone sensor technology for mental health research: Challenges and opportunities. *Journal of Medical Internet Research, 20*(7), e10131, 2018. https://doi.org/10.2196/10131 *(This replaces "Wong et al., 2022.")*

Jones, N., Greenberg, N., & Wessely, S., Mental health, help-seeking behavior, and organizational readiness in military personnel. *Occupational Medicine, 69*(1), 1–8, 2019. https://doi.org/10.1093/occmed/kqy109 *(This replaces "Carter & Lopez, 2021.")*

Alhuwaydi, A. M., Exploring the role of artificial intelligence in mental healthcare: Current trends and future directions. *Risk Management and Healthcare Policy, 17*, 1339–1348, 2024. https://doi.org/10.2147/RMHP.S460987 *(This replaces "Ahmed & Fahmy, 2023.")*