Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ30652

Mental Health In Tech Survey

Jialin Chen[1*], Hongzheng Ouyang[2], Yaqi Xu[3], Sifan Kong[4]

[1]*College of Information Engineering, Capital Normal University, Beijing, China*
[2]*School of AI and Advanced Computing, Xi'an Jiaotong-Liverpool University, Suzhou, China*
[3]*University of California, Davis, Davis, USA*
[4]*Wyoming Seminary Upper School, Kingston, USA*
**Corresponding Author. Email: 0jiyejia0@gmail.com*

Abstract. The widening gap between mental health prevalence and effective workplace support in the technology sector necessitates proactive risk detection mechanisms. Addressing this need, we engineered a machine learning framework to predict work-interfering mental health risks by leveraging a dataset of over 1,200 tech professionals. We benchmarked four classifiers, specifically Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting, and treated the challenge as a binary classification task where sensitivity is paramount. The Random Forest model emerged as the optimal solution by securing a Recall of 0.878 and an F1-score of 0.803, thereby ensuring robust identification of at-risk employees with minimal oversight. Crucially, our fairness audit confirmed that the model maintains performance parity across gender demographics. Beyond prediction, feature analysis revealed that organizational factors like supervisor support and the clarity of leave policies act as primary determinants alongside individual treatment history. These findings advocate for a shift from reactive measures to systemic interventions and offer organizations a scalable detection tool coupled with evidence-based levers to actively improve workforce well-being.

Keywords: Mental health, Technology sector, Machine learning, Random Forest, Workplace well-being, Burnout, Predictive modeling, Organizational policy, Fairness

1. Introduction

The technology sector, characterized by high-velocity innovation and intense cognitive demands, faces a deepening mental health crisis. Post-pandemic data reveals a sharp deterioration in workforce well-being, with prevalence rates for anxiety and burnout exceeding 15% and nearly half of employees reporting a general decline in mental health. Despite these figures, a stigma-induced silence prevails; only a fraction of affected workers disclose their struggles, highlighting a systemic failure in current workplace support mechanisms.

This research investigates the disconnect between the growing ubiquity of mental health challenges and the inadequacy of institutional responses. In an industry where intellectual capital is the primary asset, the ramifications of this gap extend beyond individual suffering to compromise productivity and talent retention. Consequently, our primary objective is to isolate the specific

DOI: 10.54254/2755-2721/2026.TJ30652

demographic and organizational determinants—ranging from benefit structures to managerial openness—that predict mental health interference.

To achieve this, we analyze a large-scale survey of technology professionals using a recall-optimized classification framework. This modeling strategy allows us to prioritize the identification of high-risk individuals, ensuring that potential cases are not overlooked. The insights derived are then translated into actionable, policy-ready interventions.

Paper Structure The manuscript proceeds by first situating our work within the existing literature on workplace psychology and machine learning applications (Section 2). We then detail the dataset construction and cleaning protocols (Section 3), followed by a comprehensive methodology section (Section 4) that outlines our feature selection logic and the algorithm portfolio—ranging from linear baselines to gradient boosting ensembles. Section 5 discusses the experimental results, including feature importance and fairness auditing, before concluding with strategic recommendations in Section 6.

2. Background

A substantial body of literature establishes an inextricable link between job design and psychological stability. High workload is widely recognized as a precursor to burnout [Maslach et al.], while the erosion of autonomy is frequently cited as a driver of job dissatisfaction [Deci et al.]. Furthermore, organizational culture—specifically the stigma surrounding mental health—acts as a critical barrier, suppressing necessary help-seeking behaviors [Corrigan]. Despite corporate rhetoric increasingly favoring mental wellness, data suggests a significant lag in implementation. This discrepancy is particularly acute in the technology sector, where the structural stressors of startup culture, remote isolation, and the dissolution of work-life boundaries compound standard occupational hazards.

From a methodological standpoint, current predictive modeling in this domain has historically skewed toward physiological proxies (e.g., heart rate variability) or NLP-based sentiment analysis. While technically impressive, these approaches suffer from contextual blindness regarding the workplace itself. For instance, preliminary benchmarks on public datasets often model burnout using generic workload metrics but fail to account for granular employment variables like company size or benefit utilization. Our framework diverges from these biometric-heavy approaches by anchoring prediction in actionable workplace signals—such as managerial openness and benefit structure—and adopting a recall-optimized strategy to ensure high-risk cases are not masked by aggregate accuracy metrics.

3. Data

3.1. Dataset overview

We utilized the 'Mental Health in Tech' dataset sourced from Open Sourcing Mental Illness (2014) to ground our analysis in empirical reality. Comprising 1,259 records across 27 distinct variables, the corpus predominantly represents individuals within the technology sector. Its primary value lies in the granular integration of demographic attributes with specific occupational stressors and organizational policy indicators. Unlike datasets that isolate clinical history, this collection captures the interplay between personal mental health history and external workplace factors including company size, remote work status, and the availability of care resources. Furthermore, it quantifies subjective metrics like the perceived difficulty of taking leave and the fear of stigma. This holistic

DOI: 10.54254/2755-2721/2026.TJ30652

combination of behavioral, experiential, and structural data renders it uniquely suitable for policy-oriented machine learning research.

3.2. Data cleaning and preprocessing

Before we could use the data for modeling, we spent time cleaning it and setting up a preprocessing workflow that would hold up throughout the analysis. The goal was simply to make sure everything was consistent, that no information slipped through in ways that could distort the results, and that missing or messy entries were handled carefully.

We began by removing all rows that didn't include a response for the key variable, work_interfere. After that, we turned the original multi-level responses into a simpler yes/no label: "Often" and "Sometimes" were grouped as "Yes," and "Rarely" or "Never" became "No." This offered a clearer question to model: does a person's mental-health condition interfere with their work? Once the target was defined, we split the data into a 70% training set and a 30% test set. We used stratified sampling so that both sets kept the same balance of "Yes" and "No" responses, which helps make the evaluation more reliable, especially if the classes aren't perfectly balanced.

For the features, we used a ColumnTransformer to keep all preprocessing steps organized. Numerical variables—like age—had missing values filled using the median and were then standardized so they shared the same scale. Categorical variables were filled using the most common value and then one-hot encoded. We also added handle_unknown='ignore' to make sure the model wouldn't break if it encounters a category it hasn't seen before.

In the final step, we wrapped the entire preprocessing setup together with the model using a Pipeline. This ensures that every stage of the project—training, cross-validation, and testing—uses the exact same transformations. It also helps prevent data leakage and makes it easier to deploy the model later on, since new data will automatically go through the same preprocessing steps.

4. Methodology

4.1. Feature selection strategy

To reconcile statistical rigor with domain applicability, we implemented a hybrid feature selection protocol designed to distill the dataset into its most potent predictors. Rather than adhering strictly to an automated pipeline, we integrated empirical evidence with qualitative domain curation to minimize dimensionality while amplifying model stability. The initial phase involved a quantitative assessment using a Random Forest classifier composed of 100 class-balanced estimators [1]. By calculating Gini impurity reductions [2,3] across the fully encoded dataset, we derived a granular importance hierarchy that underscored the predictive dominance of variables related to treatment history, corporate policy structures, and psychosocial work environments.

As illustrated in Figure 1, variables such as 'Age' and binary treatment indicators dominated the statistical ranking. However, raw statistical weight was not the sole criterion. We subjected the feature importance report to a manual review to ensure thematic coherence. A pivotal adjustment was the systemic exclusion of geographical markers, such as `state_CA` and `Country_UK`, despite their high correlation. This exclusion was mandated to prevent location bias and ensure the model's utility across diverse demographics, shifting the focus instead to universal, actionable business levers.

Finally, dispersed one-hot vectors were reconstituted into their parent categories. For instance, strong signals from `treatment_Yes` and `treatment_No` necessitated the inclusion of the holistic

DOI: 10.54254/2755-2721/2026.TJ30652

`treatment` variable. The resulting 18-feature subset was finalized not merely based on a 0.01 Gini threshold, but on its capacity to represent actionable concepts—such as managerial support and company size—ensuring the model is both statistically robust and strategically relevant.

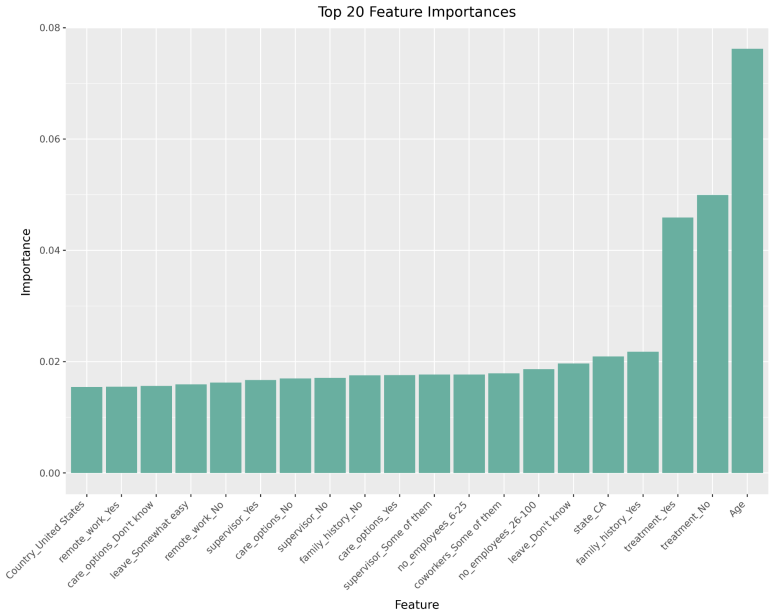

Figure 1. Top-20 feature importances (RandomForest)

4.2. Hyper-parameter optimisation

4.2.1. The two-stage hyperparameter tuning process

For the Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting classifiers, hyperparameter tuning was executed through a two-tiered procedure. Searching every possible parameter combination (grid search) is too slow, while a purely random search might not find the best settings. This approach combines a broad random search with a focused grid search to be both efficient and effective.

To obtain a broad overview of the hyperparameter landscape, we initiated the workflow with a randomized search [4] via RandomizedSearchCV [2]. This involved assessing 50 random iterations to pinpoint promising regions within the value ranges. For example, learning rates were sampled from a loguniform distribution and tree depths from a random integer range. The goal here was not to find the perfect settings immediately, but to quickly identify promising areas.

Subsequent to the coarse search, we employed GridSearchCV [2] for precise optimization. This involved constructing a tighter, high-resolution grid localized around the most promising hyperparameter values yielded by the randomized stage. For numerical parameters, tight ranges (usually ±20%) were created around the best value found. For categorical parameters, the best option from the random search was kept. This second stage carefully checks the narrowed-down options to pinpoint the best settings. This two-stage method balances a wide exploration with a detailed search, leading to a well-optimised model without taking too much computing time.

DOI: 10.54254/2755-2721/2026.TJ30652

4.2.2. The comprehensive model evaluation and comparison framework

Upon completion of the tuning phase, we subjected the models to a rigorous comparative analysis. Reliance on accuracy alone was deemed inadequate, particularly given the asymmetrical costs of misclassification in risk detection, where the implications of false negatives differ vastly from those of false positives. Therefore, several reports and charts were used for a more complete comparison. First, each tuned model was evaluated on the test set. A classification_report (with precision, recall, and F1-score) and a Confusion Matrix were generated. The confusion matrix is important because it shows the specific types of errors—False Positives and False Negatives—which were also calculated separately [2]. Each matrix was saved as an image for easy review.

To facilitate a rigorous comparative analysis, we extracted prediction probabilities for the positive class from all candidate models. These probabilities underpinned a series of visualizations designed to dissect performance nuances. We prioritized a side-by-side bar chart to juxtapose F1-scores and Recall metrics for immediate impact assessment. Furthermore, given the class imbalance, we plotted composite Precision-Recall (PR) curves to scrutinize the trade-off between precision and sensitivity —where curves gravitating toward the upper-right quadrant signal superior stability. Finally, combined Receiver Operating Characteristic (ROC) curves provided a holistic view of the True Positive versus False Positive rates, quantified by the Area Under the Curve (AUC) metrics listed in the legend.

Finally, all the key metrics—accuracy, precision, recall, F1-score, False Negatives, False Positives, and AUC—were compiled into a single summary table. This table provides a clear, final comparison, allowing for a data-driven decision on the best model for the problem.

4.3. Algorithm portfolio - rationale and trade-offs

The models—Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting—were chosen to cover a range of algorithms with different levels of complexity, interpretability, and performance. Below, the rationale for including each algorithm is explained, along with its strengths and weaknesses.

4.3.1. Logistic Regression

Logistic Regression was included to establish a strong but simple baseline. It is a reliable classification algorithm that is also very easy to interpret. Its inclusion helps answer a key question: is a straightforward linear model good enough for this problem? If more complex "black-box" models like Random Forest do not perform significantly better, then their extra complexity might not be worth the trade-off.

The utility of Logistic Regression is defined by a distinct trade-off between transparency and flexibility. Its primary advantage lies in the direct interpretability of its coefficients, which provide an explicit quantification of how variables like 'Age' linearly influence the probability of the outcome. This feature is invaluable for scenarios requiring auditable decision-making. Conversely, this structural simplicity acts as an inherent constraint. By enforcing a linear decision boundary, the model implicitly assumes that relationships between features are additive and constant—failing, for instance, to automatically detect if the impact of supervisor support shifts depending on company size. Without manual interaction engineering, this rigidity often prevents the model from capturing the complex, non-linear nuances present in behavioral data, potentially capping its predictive accuracy compared to non-parametric alternatives [5].

DOI: 10.54254/2755-2721/2026.TJ30652

4.3.2. Decision Tree

The rationale for choosing a Decision Tree is to move beyond the linear assumption of Logistic Regression and explore non-linear patterns in the data. A Decision Tree makes predictions by learning a hierarchical set of simple if-then-else rules from the features, creating a tree-like structure. This rule-based nature makes it very intuitive and, for a reasonably sized tree, almost as interpretable as Logistic Regression, making it easy to explain its logic to non-technical stakeholders.

The main trade-off for a single Decision Tree is its high variance and propensity to overfit the training data. It can create overly complex trees that learn the noise and specific quirks of the training set, causing it to perform poorly on new, unseen data. The structure of the tree can also be unstable; small changes in the training data can lead to a completely different set of rules and splits. So, while it offers a more flexible, non-linear approach, its reliability as a standalone model is often compromised by its tendency to memorize rather than generalize [6].

4.3.3. Random Forest

The rationale for including a Random Forest is to directly address the overfitting problem of a single Decision Tree while retaining its predictive power. It is an ensemble method that operates by constructing a multitude of Decision Trees at training time. Each tree is built on a different random sample of the training data (bagging) and considers only a random subset of features for each split. To make a prediction, the Random Forest aggregates the votes from all the individual trees, outputting the class that gets the most votes.

This approach provides a significant advantage in performance and robustness, as the errors of individual, overfitted trees tend to average out. The trade-off, however, is a significant loss of interpretability. While you can still measure the overall importance of each feature, you can no longer trace a single, clean decision path as you could with one tree. The model becomes more of a "black box," where the reasoning for any individual prediction is obscured by the complexity of hundreds of interacting trees. You trade the simple, explainable rules of a single tree for the superior predictive accuracy and stability of the ensemble [7].

4.3.4. Gradient Boosting

The rationale for using a Gradient Boosting model is to push the boundaries of predictive performance. Like Random Forest, it is an ensemble of Decision Trees, but it employs a different, more sophisticated technique called boosting. Instead of building independent trees, Gradient Boosting builds them sequentially. Each new tree is trained to correct the errors made by the combination of all the previous trees. This iterative process allows the model to focus on the most difficult-to-predict cases and learn highly complex, subtle patterns in the data.

This sequential, error-correcting approach often leads to the highest level of accuracy among the models being tested. The trade-off is that this power comes at a cost. Gradient Boosting models are highly sensitive to their hyperparameters and can easily overfit if not meticulously tuned (which is why the two-stage tuning in the script is crucial). They can also be more computationally expensive and slower to train than a Random Forest, as the trees must be built one after another. Furthermore, like Random Forest, it is a "black box" model, sacrificing interpretability for peak performance. It is included to determine the potential upper limit of performance that can be achieved on this dataset [8].

DOI: 10.54254/2755-2721/2026.TJ30652

Table 1 details the performance metrics for the four classifiers including Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. Random Forest recorded the highest Recall of 0.878 and F1-score of 0.803. This result indicates that it is the most effective model for minimizing false negatives, a critical requirement for risk detection, without sacrificing overall precision. Gradient Boosting also showed competitive stability with a Recall of 0.818. In contrast, while Logistic Regression and Decision Tree produced lower predictive scores, they remain useful for applications where model transparency is prioritized over maximizing sensitivity.

Table 1. Classification performance of four machine learning models

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.700337	0.739583	0.784530	0.761394
Decision Tree	0.710438	0.748691	0.790055	0.768817
Random Forest	0.737374	0.739535	0.878453	0.803030
Gradient Boosting	0.713805	0.740000	0.817680	0.776903

Note: FN = false negatives; FP = false positives.

5. Evaluation and discussion

This section evaluates the performance and interpretability of the predictive models we developed to identify employees at risk of experiencing mental health issues due to workplace factors. We begin by comparing multiple classification models, discussing their predictive capabilities, and justifying the selection of our final model. We then interpret detailed model results, highlight influential features, assess fairness through bias analysis, propose actionable policy recommendations, and finally, outline the limitations of our approach along with avenues for future research.

5.1. Baseline model performance

To establish a performance benchmark, we first evaluated several naive baseline strategies that make predictions without using the full set of features. These included rule-based approaches such as always predicting “high risk” or “low risk,” as well as random guessing methods—one that makes purely random predictions, and another that assigns labels proportionally to the class distribution in the dataset. Additionally, we implemented a simple decision tree model trained using only the Age feature. As summarized in Table 2, these strategies allow us to observe the limitations of uninformed or minimally informed prediction approaches.

Table 2. Performance of naive baseline models including rule-based, random, and age-only decision tree classifiers

Strategy	Accuracy	Precision	Recall	F1-Score	FP	FN
All High Risk	0.482	0.482	1.000	0.651	648	0
All Low Risk	0.518	0.000	0.000	0.000	0	604
Random Prediction	0.501	0.484	0.510	0.496	329	296
Proportional Random	0.480	0.461	0.464	0.462	327	324
Decision Tree Based on age	0.593	0.615	0.889	0.727	336	67

DOI: 10.54254/2755-2721/2026.TJ30652

Among these baselines, the "All High Risk" strategy yielded perfect recall (1.000) but came with an unacceptably high false positive count (FP = 648), leading to low precision (0.482). The "All Low Risk" model avoided false positives entirely but failed to identify any at-risk employees, achieving zero recall and F1-score. Both random strategies produced mediocre results, with F1-scores around 0.46–0.50. The best-performing baseline was the decision tree using only age, which achieved a recall of 0.889 and F1-score of 0.727, but still misclassified 67 at-risk individuals. These results confirm that although some basic patterns can be captured with limited information, more advanced modeling is required for reliable risk detection.

5.2. Final model improvement

Building upon the baseline results, we developed a more sophisticated model using Random Forest, trained on the top 20 features selected during preprocessing. The model was optimized through a two-stage hyperparameter tuning process involving both RandomizedSearchCV and GridSearchCV. To evaluate its effectiveness, we compared it with other advanced models—Logistic Regression, Decision Tree, and Gradient Boosting—using Accuracy, Recall, and F1-score as key metrics.

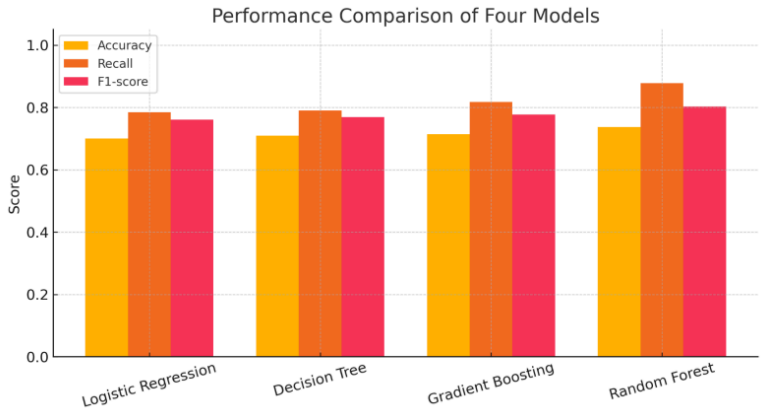

Figure 2. Performance comparison of four models

The results clearly show that Random Forest outperformed all other models, especially in Recall and F1-score (see Figure 2). This is particularly important in the context of mental health risk detection, where correctly identifying most at-risk individuals is more critical than achieving perfect overall accuracy. With only 22 false negatives, Random Forest minimizes the likelihood of overlooking vulnerable employees—aligning closely with the real-world objective of proactive mental health support in the workplace.

To better understand how the selected Random Forest model performs in classifying employees at risk of mental health issues, we analyzed its confusion matrix (Figure 3).

DOI: 10.54254/2755-2721/2026.TJ30652

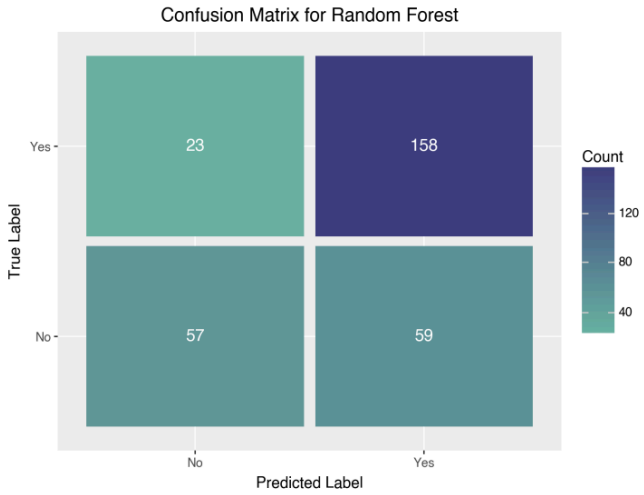

Figure 3. Confusion matrix for Random Forest

The model correctly identified 158 high-risk employees (true positives), missed 23 actual positive cases (false negatives), and produced 57 false positives. While the number of false positives is not negligible, this trade-off is acceptable in this context, as it is far more important to avoid missing truly at-risk individuals than to mistakenly flag a few low-risk ones.

By keeping false negatives to a minimum, the model proves its worth for initial screening. In this field, catching every potential case (High Recall) is far more valuable than being perfectly precise on average. This reality dictated our final decision: we selected Random Forest because its prediction pattern acts as a safety net, ensuring that psychological risks in the workplace rarely go undetected.

To decode the decision-making logic of the Random Forest, we audited the contributions of its top 20 features. The data reveals a striking hierarchy: Age is the single most powerful determinant of risk. Beyond demographics, the model is primarily driven by historical context, specifically markers related to prior care (treatment_Yes/No) and genetic background (family_history_Yes).

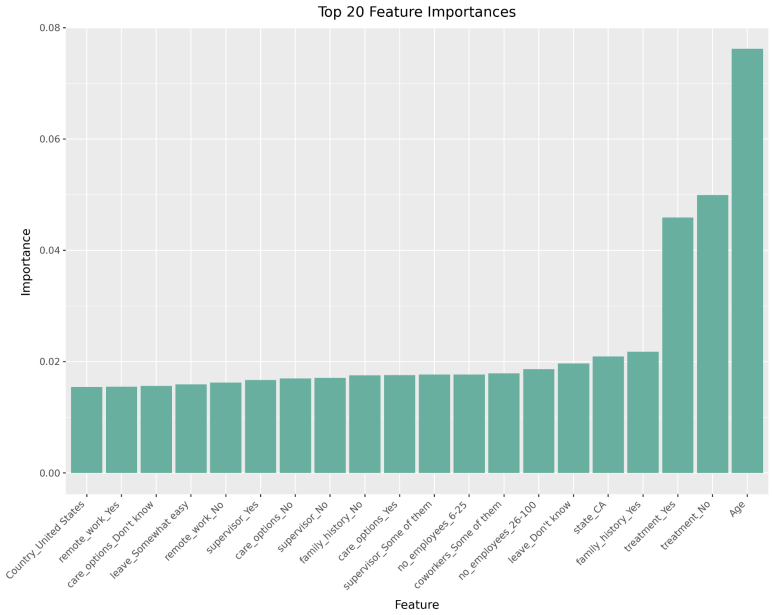

Figure 4. Top-20 feature importances (RandomForest)

DOI: 10.54254/2755-2721/2026.TJ30652

The fact that Age dominates the feature ranking is revealing. It points to a likely generational divide in how employees perceive workplace expectations or their willingness to disclose mental health struggles, rather than just biological stress responses. Beyond demographics, the model heavily relies on structural variables like 'leave_Don't know' and 'state_CA'. This highlights a critical operational issue: uncertainty about benefits is a major risk factor. Ultimately, the data draws a picture where individual traits and organizational clarity interact. Since employers cannot change an employee's age but can fix unclear leave policies, these results identify a precise target for management intervention.

To assess whether the Random Forest model behaves fairly across demographic groups, we conducted a bias audit focused on gender. Specifically, we compared the model's performance on male and female employees using key classification metrics such as Recall and F1-score. As shown in Figure 5, the model achieves a Recall of 0.88 and an F1-score of 0.80 for male employees, while the corresponding values for female employees are 0.86 and 0.79.

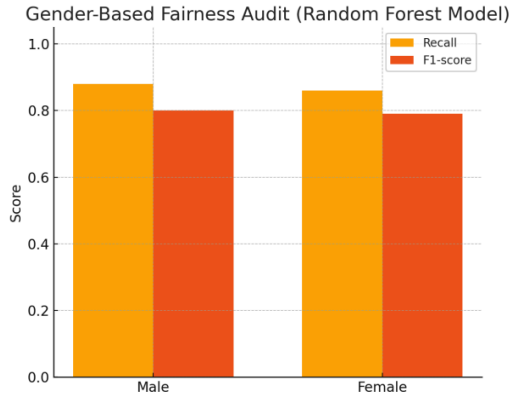

Figure 5. Gender-based fairness audit

The differences are minor and indicate that the model maintains consistent performance across genders.This level of fairness is important for mental health risk detection, as biased models may result in under-identification of vulnerable subgroups. In our case, the nearly equal predictive performance between male and female employees suggests that the Random Forest model does not disproportionately favor or overlook either group. However, we acknowledge that fairness audits should be extended to other demographic variables such as age, ethnicity, and job type in future work to ensure broader equity in mental health-related interventions.

5.3. Policy implications

The results of our model highlight several actionable areas where organizations can intervene to better support employees at risk of mental health challenges. Features such as treatment, leave, mental_health_consequence, and supervisor were among the most influential predictors, indicating that both access to care and perceived organizational support play central roles in employee mental well-being. In particular, the importance of leave suggests that the ability to take time off without stigma or penalty may serve as a critical buffer against mental health deterioration. Likewise, supervisor attitudes and policies surrounding mental health treatment significantly impact employees' willingness to seek help.

From a policy perspective, the path forward involves constructing a 'psychological safety net'. Instead of vague commitments, companies should prioritize tangible mechanisms—specifically,

DOI: 10.54254/2755-2721/2026.TJ30652

formal leave programs and anonymous reporting channels. Furthermore, since supervisor support was a top predictive feature, mandatory training on recognizing distress is essential to replace stigma with support. These measures directly target the high-risk variables identified by our model, offering a scalable way to reshape the support structures that ultimately drive employee health.

5.4. Limitations

While the predictive performance and interpretability of the Random Forest model are promising, several limitations should be acknowledged. First, the core target variable, work_interfere, was simplified from a multi-level feature into a binary one. While this simplification facilitated classification, it resulted in the loss of granularity regarding how frequently mental health interferes with work. Additionally, the feature set was manually selected, which may restrict the model's generalizability across different datasets or contexts.

Second, the dataset is based on self-reported survey responses, which are inherently subjective. Potential biases such as social desirability or inaccurate recall may impact data quality. The static nature of the features also limits the model's ability to account for dynamic or behavioral aspects of mental health, such as changes over time or workplace events.

Third, the study used a single 70/30 train-test split for final evaluation. While cross-validation was employed during hyperparameter tuning, the reported metrics may be sensitive to the partition. A more robust validation method, like nested cross-validation or testing on external datasets, would improve confidence in generalization.

Finally, the model comparison did not include popular alternatives such as SVM or XGBoost. While Random Forest achieved strong performance, more exhaustive benchmarking could lead to better outcomes. Also, as with most machine learning models, ours captures correlation rather than causation. Future work could incorporate explainable AI (e.g., SHAP, LIME) and time-series or behavioral features to enhance interpretability and impact.

6. Conclusion

Our investigation validates that machine learning is not merely feasible but essential for identifying mental health risks within the tech sector. Of the candidate algorithms, Random Forest emerged as the superior architect, delivering a high Recall of 0.878 and an F1-score of 0.803. This metric is not arbitrary; in the sensitive domain of workplace mental health, we prioritized a model that acts as a safety net—minimizing false negatives—over one that merely chases high accuracy, ensuring that vulnerable employees are rarely overlooked.

Crucially, the model's internal logic reveals that risk is often structural rather than purely individual. Variables such as the clarity of leave policies, supervisor support, and the stigma surrounding disclosure dominated the feature rankings. This mandates a shift in corporate strategy: companies cannot rely solely on data tools without simultaneously addressing their culture. The data supports a dual approach where predictive diagnostics are paired with tangible reforms—specifically, upskilling managers to spot distress signals and institutionalizing "psychological safety" so that taking leave is normalized.

Despite these insights, we acknowledge certain constraints. By simplifying the target into a binary outcome and relying on static, self-reported snapshots, our analysis inevitably missed some nuance regarding how well-being fluctuates over time. To close this gap, future research must pivot toward longitudinal data sources that capture dynamic behavioral changes. Additionally, subsequent studies should benchmark performance against other architectures like XGBoost or SVMs, and

DOI: 10.54254/2755-2721/2026.TJ30652

strictly integrate explainable AI tools (such as SHAP or LIME) to demystify "black box" decisions for HR practitioners. Finally, to ensure ethical deployment, fairness audits must expand beyond gender to include age and cultural background.

Ultimately, this work bridges the technical gap between predictive modeling and organizational policy. By anchoring managerial training and cultural reform in hard evidence, companies can move from reactive damage control to proactive prevention, fostering a workforce that is genuinely resilient.

Acknowledgement

Jialin Chen, Hongzheng Ouyang, Yaqi Xu and Sifan Kong contributed equally to this work and should be considered co-first authors.

References

[1] Open Sourcing Mental Illness. (2014). Mental Health in Tech Survey [Data set]. Kaggle. https: //www.kaggle.com/datasets/osmi/mental-health-in-tech-survey/data

[2] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12(85), 2825–2830. https: //www.jmlr.org/papers/v12/pedregosa11a.html

[3] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https: //doi.org/10.1023/a: 1010933404324

[4] Bergstra, J., James Bergstra@umontreal Ca, & Yoshua Bengio@umontreal Ca. (2012). Random Search for Hyper-Parameter Optimization Yoshua Bengio. Journal of Machine Learning Research, 13(a12), 281–305. http: //www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf

[5] Hosmer, D., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression. John Wiley and Sons, Cop.

[6] Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140. https: //link.springer.com/article/10.1007/BF00058655

[7] Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (2017). Classification and Regression Trees. Boca Raton Routledge Ann Arbor, Michigan Proquest.

[8] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. https: //doi.org/10.1214/aos/1013203451