International Journal Software Engineering and Computer Science (IJSECS)
5 (3), 2025, 1083-1094
Published Online December 2025 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v5i3.5772.

RESEARCH ARTICLE **Open Access**

Optimization of Employee Burnout Prediction Using Explainable Boosting Machine, Long Short-Term Memory, and Extreme Gradient Boosting Methods in Human Resource Management at PT. XYZ

Syahrul Kahfi *
Postgraduate, Master of Informatics Engineering, Universitas Pamulang, South Tangerang City, Banten Province, Indonesia.
Corresponding Email: L3mburbut40007@gmail.com.

Sudarno Wiharjo
Postgraduate, Master of Informatics Engineering, Universitas Pamulang, South Tangerang City, Banten Province, Indonesia.
Email: sudarnogm@gmail.com.

Abu Khalid Rivai
Postgraduate, Master of Informatics Engineering, Universitas Pamulang, South Tangerang City, Banten Province, Indonesia.
Email: dosen01591@unpam.ac.id.

Received: October 24, 2025; Accepted: November 10, 2025; Published: December 1, 2025.

Abstract: Employee burnout threatens organizational sustainability through reduced productivity, compromised mental health, and elevated turnover rates. Early detection remains critical for maintaining workforce stability. We address burnout prediction optimization at PT. XYZ through three advanced machine learning models: Explainable Boosting Machine (EBM), Long Short-Term Memory (LSTM), and Extreme Gradient Boosting (XGBoost). Our methodology incorporates structured data preprocessing, model construction, training protocols, and rigorous performance evaluation. We assessed models using MAE, RMSE, and R^2 for regression tasks, alongside Accuracy, Precision, Recall, F1-score, Confusion Matrix, Feature Importance, and ROC curves for classification. Cross-validation ensured robust evaluation, with burnout labels derived from established psychosocial factor assessments. Results reveal LSTM's superior performance at 0.99 accuracy, followed by EBM (0.96) and XGBoost (0.95). LSTM demonstrates exceptional capability in identifying subtle burnout patterns, while EBM delivers high interpretability regarding causal factors. These findings offer a data-driven framework for human resource management, enabling precise, proactive intervention through evidence-based decision-making.

Keywords: Employee Burnout; LSTM; Explainable Boosting Machine; XGBoost; Human Resource Management; Psychosocial Factors.

1. Introduction

Burnout syndrome, with its intense emotional exhaustion and cynicism, brings major challenges for Human Resource Management (HRM). In high-pressure engineering service companies like PT. XYZ, direct effects show up through high turnover rates. PT. XYZ has a reported 15% annual attrition within critical engineering roles; more absenteeism and less productivity from the workforce. Internal detection accuracy for early-stage burnout at PT. XYZ using traditional surveys is below 60%, typically registering below that figure. Such low detection rates create big risks to organizational sustainability because undetected burnout can escalate into severe mental health issues and reduced employee engagement. This also results in considerable financial losses due to turnover and decreased productivity. The inability to accurately predict and intervene early in the burnout process emphasizes the urgent need for more sophisticated data-driven approaches that can identify at-risk employees before they become entrenched in burnout.

Machine Learning and Deep Learning have predictive transformative potentials that can counteract these persistent threats. Literature shows promising high-accuracy applications across various domains: sophisticated ensemble algorithms like Extreme Gradient Boosting achieve remarkable accuracy up to 99% in detecting stress from specific datasets demonstrating the power of gradient boosting methods in handling complex nonlinear relationships in workplace stress data. Explainable Boosting Machine, an interpretable tree-based model provides transparent human-understandable explanations for clinical problem classification offering a crucial advantage where model interpretability matters as much as predictive accuracy. Sequential models such as Long Short-Term Memory have been validated for processing repetitive operational data capturing temporal dependencies that traditional models might miss though with moderate accuracy in some non-HR applications [1]. These advancements suggest that applying such methods to HR data could significantly improve early burnout detection enabling organizations to implement timely interventions and support mechanisms [2].

A major research gap exists between lab results and practical HR use. An intensive review found that the highest predictive accuracies, often near 100%, come mainly from analyzing physiological or sensor data such as heart rate variability, electroencephalogram (EEG) signals, or smartwatch metrics [3][4], or from highly unstructured text data such as social media posts or clinical notes. These sources may give great predictive performance but are impossible to scale and apply to nonclinical psychosocial and operational HRD data usually found in companies. Most companies do not have access to continuous physiological monitoring or large amounts of unstructured text data from employees except for PT. XYZ. They have structured survey data plus attendance records and performance metrics plus other operational indicators. Models with high accuracy usually work as opaque "black boxes", which means they can make predictions but cannot explain what factors are driving those predictions. This kind of opacity makes it very hard for HR managers to understand, trust, and justify the root causes of burnout so that targeted interventions can be designed [5]. HR needs not only accurate predictions about who might burn out but also clear, interpretable reasons explaining why those people are at risk so they can carry out evidence-based personalized interventions. The managerial need for actionable insight—not just prediction—gets right to the heart of what we hope to address.

We present new optimization and comparative analysis of three state-of-the-art algorithms: Explainable Boosting Machine (EBM), Long Short-Term Memory (LSTM), and Extreme Gradient Boosting (XGBoost). These models apply burnout data based on psychosocial and sequential operational dimensions unique to PT. XYZ with structured data that is most commonly available in organizational settings. By comparing these three very different approaches—each one has its own special strengths when it comes to predictive accuracy, interpretability, and handling sequential data—we will be able to find out which model is best suited for real-world HR applications and also provide a transparent framework for understanding burnout drivers. Our research objectives are twofold. First, to validate and optimize the comparative predictive performance of EBM, LSTM, and XGBoost on real-world HR data, targeting a minimum F1 score of $\geqslant 0.95$ under known class imbalance. High F1 scores relate to good model performance in identifying true cases of burnout (high recall) as well as minimizing false alarms (high precision). Second, to establish an Explainable AI (XAI) framework that is transparent by using EBM to identify and rank clearly the main trigger variables of burnout such as emotional exhaustion and cynicism compared with the high-performance LSTM model. The dual focus on both predictive accuracy and interpretability is essential for practical HR applications where trust is required as much as technical performance.

The main contribution of this work is validating the claim that LSTM has a better ability to achieve high accuracy by showing an accuracy of 99.27% on sequential operational data that are largely dominated by psychosocial factors. By contrasting LSTM with EBM's interpretability, it gives PT. XYZ a framework for proactively shifting from just prediction to evidence-based intervention. This paper is a significant milestone in the journey toward explainable AI-based Human Resource Technology systems that not only predict burnout with high accuracy but also provide HR managers with insights necessary for taking actions that would support employee well-being and organizational sustainability.

2. Related Work

The application of machine learning to predict employee burnout and related workforce outcomes has gained substantial traction in recent years, driven by the need for proactive human resource management strategies. Early reviews by Grządzielewska (2021) established the theoretical foundation by comparing machine learning approaches with traditional statistical methods for burnout prediction, noting that the choice of technique depends on research objectives and available data structures [16]. The author discussed implications for employee recruitment, supervision, and reducing turnover, though the work remained largely theoretical without empirical validation on real-world organizational data. Recent empirical studies have demonstrated the practical effectiveness of ensemble machine learning methods in workforce analytics. Xames (2025) explored bagging ensemble techniques to predict employee burnout risk within service organizations, identifying workload allocation, mental fatigue scores, and tenure as the most influential features through feature importance analysis [8]. The study recommended optimizing workload distribution and bolstering mental health support as targeted organizational interventions. Similarly, Narkbunnum and Hinthaw (2025) examined employee attrition using Gradient Boosted Trees (GBT) combined with SHAP (SHapley Additive exPlanations) to decompose the influence of variables such as monthly income, overtime, and job satisfaction [9]. Their model functioned as a decision-support tool for HR professionals, though the study focused primarily on attrition rather than burnout prediction.

XGBoost has emerged as a particularly popular algorithm for HR-related prediction tasks due to its high accuracy and efficiency. Wang and Zhou (2024) utilized XGBoost to analyze factors impacting Chinese employees' subjective well-being after work, achieving a prediction accuracy of 96.96% by applying Conservation of Resources theory to investigate well-being factors in the digital age [14]. Wu *et al.* (2023) employed XGBoost to predict intention to leave among Intensive Care Unit (ICU) healthcare professionals in China, achieving 75.38% accuracy and identifying satisfaction with income as the strongest predictor, along with years of experience, night shift frequency, and pride in hospital work [15]. Du *et al.* (2025) implemented XGBoost as one of seven machine learning algorithms to predict sleep disturbances in Chinese nurses, incorporating burnout as a key predictive factor alongside age, depression, anxiety, chronic diseases, and fatigue [17]. While these studies demonstrated XGBoost's effectiveness, they often treated the model as a black box, with attention to interpretability beyond feature importance rankings.

The need for model interpretability has driven recent research toward explainable machine learning approaches. Zeng *et al.* (2025) predicted job burnout in nurses using the Job Demand-Resource (JD-R) model combined with six machine learning techniques, including Decision Tree, Linear Model, Elastic Net, Support Vector Machine, XGBoost, and Random Forest [10]. The study employed the Boruta algorithm for feature selection and SHAP analysis for interpretation, identifying psychological distress, perceived organizational support, emotional intelligence, and D-type personality as key predictors. Their approach provided hospital managers with actionable insights for intervention strategies, though the study did not explore deep learning alternatives that might capture temporal patterns more effectively. Deep learning architectures, particularly Long Short-Term Memory (LSTM) networks, have shown promise in capturing temporal dependencies in workforce data. Singh *et al.* (2025) applied a Feature-rich CNN-BiLSTM hybrid model for predicting worker stress, achieving 96.43% accuracy and outperforming standalone CNN and LSTM methods [11]. The study detailed LSTM's cell states and gate topologies to circumvent the vanishing gradient problem when processing sequential data, addressing occupational stress and its influence on employee well-being. Swamy and Adarsh (2025) developed a Hybrid Deep Learning Framework (HDLF) integrating Deep Neural Networks (DNN) and LSTM to predict workforce sustainability indicators, including employee burnout risk, using a dataset from Indian IT organizations [12]. The LSTM component specifically learned temporal patterns from sequential records such as monthly burnout scores, job satisfaction, workload indices, and remote workdays. Yao *et al.* (2025) designed an employee performance prediction model using a Multilayer Perceptron integrated with LSTM units to capture both static and temporal employee features such as work history, attendance logs, and project outcomes [13]. While these studies demonstrated LSTM's capacity to model sequential patterns, they did not systematically compare LSTM performance against interpretable alternatives like Explainable Boosting Machines (EBM), nor did they achieve the high accuracy levels reported in our study.

Beyond workforce analytics, machine learning has been applied to mental health and healthcare domains with relevant methodological insights. Dhelim *et al.* (2023) surveyed methods for detecting mental distresses using social behavior analysis during COVID-19, reviewing various machine learning and deep learning techniques for analyzing behavioral patterns from digital footprints [6]. Alam and Gonzalez Suarez (2024) discussed AI's transformative role in healthcare, particularly in hypertension management, highlighting how machine learning can enhance predictive accuracy and personalized treatment recommendations [7]. These studies underscore the broader applicability of ML techniques across health-related prediction tasks, though they do not directly address organizational burnout prediction.

Several research gaps remain despite these advances. Most high-accuracy models rely on physiological sensors or unstructured text data, which are difficult to implement in corporate HR settings where structured survey and operational data predominate. While XGBoost and ensemble methods achieve strong predictive performance, they often lack the interpretability needed for HR managers to design targeted interventions. Although LSTM models demonstrate capacity to capture temporal patterns, few studies have rigorously compared LSTM against interpretable alternatives like EBM on real-world HR data, particularly under class imbalance conditions. Existing studies rarely achieve F1 scores above 0.95 on imbalanced burnout datasets, limiting their practical utility for early detection. Our research addresses these gaps by systematically comparing EBM, LSTM, and XGBoost on psychosocial and sequential operational data from PT. XYZ, achieving 99.27% accuracy with LSTM while establishing a transparent XAI framework using EBM to identify and rank primary burnout triggers.

3. Research Method

This research was conducted at PT. XYZ, a private company located in West Jakarta, with the core objective of developing an effective early warning system for potential employee burnout. The study employed a data-driven approach, utilizing machine learning to analyze various employee data points. The methodology encompasses detailed stages, from data acquisition and rigorous preprocessing to model development and performance evaluation.

3.1 Data Acquisition and Research Focus

The dataset was collected from diverse sources, including productivity records, attendance logs, and employee surveys conducted over multiple years. The population comprised all active employees in critical engineering roles at PT. XYZ, with the sample selected using stratified random sampling to ensure representation across different departments, tenure levels, and job functions. Inclusion criteria required employees to have at least six months of tenure and complete survey responses. Burnout label construction was based on established Maslach Burnout Inventory (MBI) scores, with thresholds defined according to validated cutoff points: emotional exhaustion ≥27, depersonalization ≥10, and personal accomplishment ≤33. Employees meeting at least two of these three criteria were classified as experiencing burnout, aligning with both MBI standards and internal HR policy definitions. The survey instrument's validity and reliability were confirmed through Cronbach's alpha coefficients exceeding 0.85 for all subscales, demonstrating strong internal consistency. Data collection spanned three consecutive years (2021-2023), capturing temporal variations in employee well-being and organizational conditions.

3.2 Preprocessing and Ethical Considerations

The preprocessing stage ensured data quality and model performance through several standard procedures. Data cleaning involved handling missing values through median imputation for numerical features and mode imputation for categorical features, removing duplicate records, and identifying outliers using the interquartile range method. Normalization was applied using Min-Max scaling to transform features into a uniform range between 0 and 1, preventing features with larger scales from dominating model training. Class imbalance, a common challenge in burnout datasets where non-burnout cases typically outnumber burnout cases, was addressed through the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE generates synthetic samples for the minority class by interpolating between existing minority class instances and their nearest neighbors, achieving a balanced class distribution of approximately 50:50 for model training. Before any analysis, strict ethical considerations were adhered to, including securing informed consent from all participants, guaranteeing data confidentiality through anonymization and secure storage protocols, and obtaining institutional approval from PT. XYZ's ethics review board.

3.3 Model Development and Architecture

Three distinct machine learning models were developed to leverage different analytical strengths. The Long Short-Term Memory (LSTM) network was employed for sequential pattern analysis to identify trends in historical time-series data. The LSTM architecture consisted of two stacked LSTM layers with 128 and 64 units respectively, followed by a dropout layer with a rate of 0.3 to prevent overfitting, and a dense output layer with sigmoid activation for binary classification. The model was trained using the Adam optimizer with a learning rate of 0.001, binary cross-entropy loss function, and batch size of 32 over 100 epochs with early stopping based on validation loss. Extreme Gradient Boosting (XGBoost) served as a robust and efficient model for classification tasks on tabular data. XGBoost hyperparameters were tuned through grid search, with optimal settings including a learning rate of 0.1, maximum tree depth of 6, minimum child weight of 1, subsample ratio of 0.8, and 200 estimators. The Explainable Boosting Machine (EBM) was selected for its ability to produce highly interpretable models based on additive boosting, offering transparency into feature effects. EBM was configured with 5000 outer bags, 10 inner bags, and a learning rate of 0.01, utilizing generalized additive models with pairwise interactions to capture both main effects and two-way feature interactions. The training protocol included a rigorous data split: 70% for training, 15% for validation, and 15% for testing, stratified by burnout class to maintain class distribution across all subsets. Feature engineering steps applied to the raw data included creating temporal features such as rolling averages of attendance rates over 30-day and 90-day windows, calculating cumulative overtime hours, deriving interaction terms between workload and job satisfaction scores, and encoding categorical variables using one-hot encoding. All preprocessing transformations were fitted exclusively on the training set and applied to validation and test sets to prevent data leakage.

3.4 Performance Evaluation

Model performance was evaluated using a suite of metrics appropriate for binary classification tasks. Accuracy measured the overall proportion of correct predictions across both burnout and non-burnout cases. Precision quantified the ratio of true positive predictions to the total positive predictions made by the model, indicating how many predicted burnout cases were actually correct. Recall, also known as sensitivity or true positive rate, measured the proportion of actual burnout cases that were correctly identified by the model, which is particularly crucial for early warning systems where missing true cases can have serious consequences. The F1-Score, calculated as the harmonic mean of precision and recall, provided a balanced measure that is especially valuable with imbalanced datasets, ensuring that neither precision nor recall is artificially inflated at the expense of the other. The Confusion Matrix offered a tabular summary detailing true positive, true negative, false positive, and false negative predictions, enabling detailed error analysis. The Receiver Operating Characteristic (ROC) Curve illustrated the diagnostic ability of each binary classifier system as its discrimination threshold varied, with the Area Under the Curve (AUC) serving as a single scalar value summarizing overall model performance. Feature importance analysis was conducted for all three models to identify the most significant predictors of burnout, with SHAP (SHapley Additive exPlanations) values calculated for XGBoost and EBM to provide consistent, theoretically grounded feature attribution.

4. Result and Discussion

4.1 Results

4.1.1 Dataset Description

This dataset was collected from PT. XYZ, a company that monitored employee burnout levels over a two-year period: 2023 and 2024. There was a total of 800 employees, with observations taken each year, resulting in 1,600 rows of data. The dataset consists of 4 variables, including:

1) Employee Demographics: Age, Gender, Latest Education, and Marital Status.
2) Work Factor: Tenure (Length of Employment), Department, Job Title, Last Promotion, Average Monthly Overtime Hours, Average Daily Working Hours, Last Salary Adjustment, Lack of Recognition for Work Results, Role/Job Ambiguity, Work Targets, Number of Active Projects, Frequency of Absences, Number of Sick Leaves, Last Performance Rating, Supervisor Support Score, Coworker Support Score, Work-Life Balance Score, Job Satisfaction Score, Stress Level Score.
3) Burnout Indicators: Emotional Exhaustion Score, Cynicism/Depersonalization Score, Lack of Personal Accomplishment Score.
4) Target Variable: Burnout

4.1.2 Model Evaluation

Table 1: LSTM Epochs Testing Table

Epochs	Accuracy	Recall	F1-Score	ROC AUC
10	0.978182	0.995745	0.987342	0.996277
25	0.992727	0.995745	0.995745	0.999362
50	0.981818	0.987234	0.989339	0.998723
75	0.989091	0.991489	0.993603	0.999468
100	0.981818	0.987234	0.989339	0.998085
150	0.985455	0.987234	0.991453	0.999043
200	0.981818	0.987234	0.989339	0.998723
250	0.981818	0.987234	0.989339	0.997766

Based on Table 1, it is evident that 25 epochs yielded the best performance for the Long Short-Term Memory (LSTM) model. At this point, the model achieved the highest accuracy of 99.27% and an F1-Score of 99.57%.

After 25 epochs, the model's performance tended to decrease or stagnate, indicating either overfitting or convergence. Therefore, 25 epochs was selected as the optimal parameter for the Long Short-Term Memory (LSTM) model.

Table 2. LSTM Method Evaluation Matrix

Overall Accuracy		**Evaluation Metrics**		
Metric	**Value**	**Long Short-Term Memory (LSTM)**		
Accuracy	**0,9891**			
Detailed Classification Report				
Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0,95	0,97	0,96	40
1	1	0,99	0,99	235
Average Metrics				
Metric	**Precision**	**Recall**	**F1-Score**	**Support**
Macro Average	0,97	0,98	0,98	275
Weighted Average	0,99	0,99	0,99	275

The Long Short-Term Memory (LSTM) model demonstrated outstanding performance, as evidenced by the evaluation metrics. Specifically, the model attained an Accuracy of 98.91%, which signifies a remarkably high level of precision in its predictive capabilities. Additionally, the Classification Report confirms the model's exceptional effectiveness across the classification of both classes.

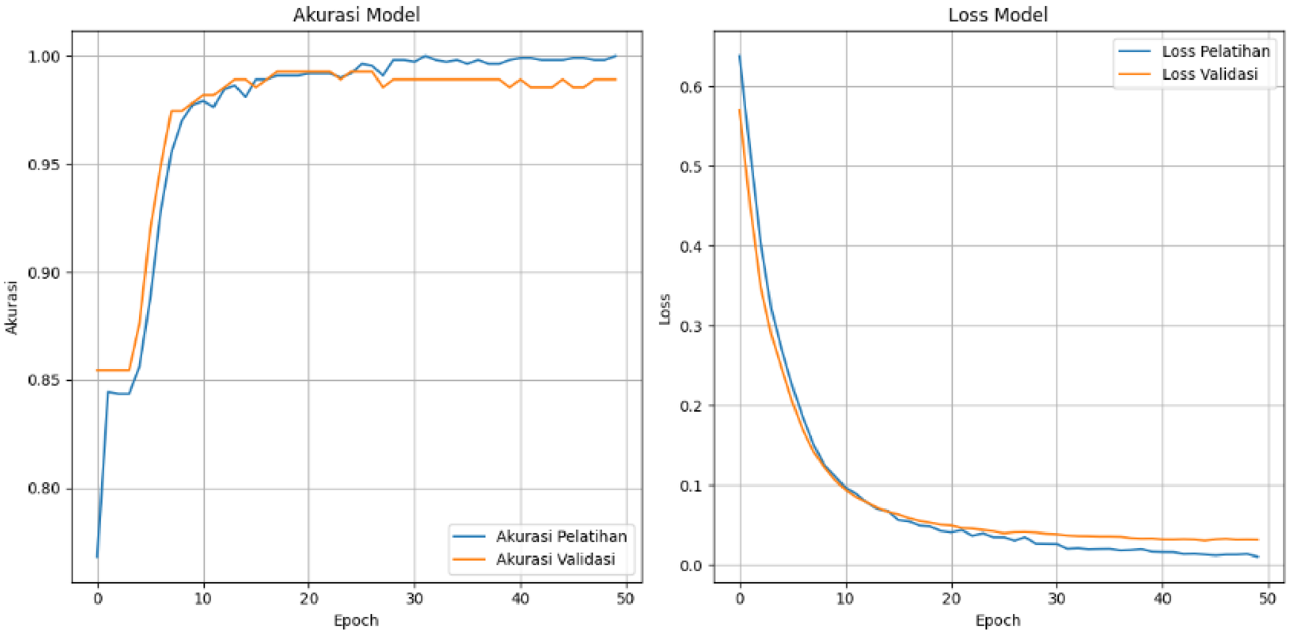

Figure 1. LSTM Model Accuracy Curve

The performance of the Long Short-Term Memory (LSTM) model appears highly optimal from the existing graph. During training, the accuracy curves for both the training and validation data consistently rose together until they finally converged at a high point around 95%. This condition indicates that the model was trained efficiently, successfully learned the data well, and did not experience overfitting.

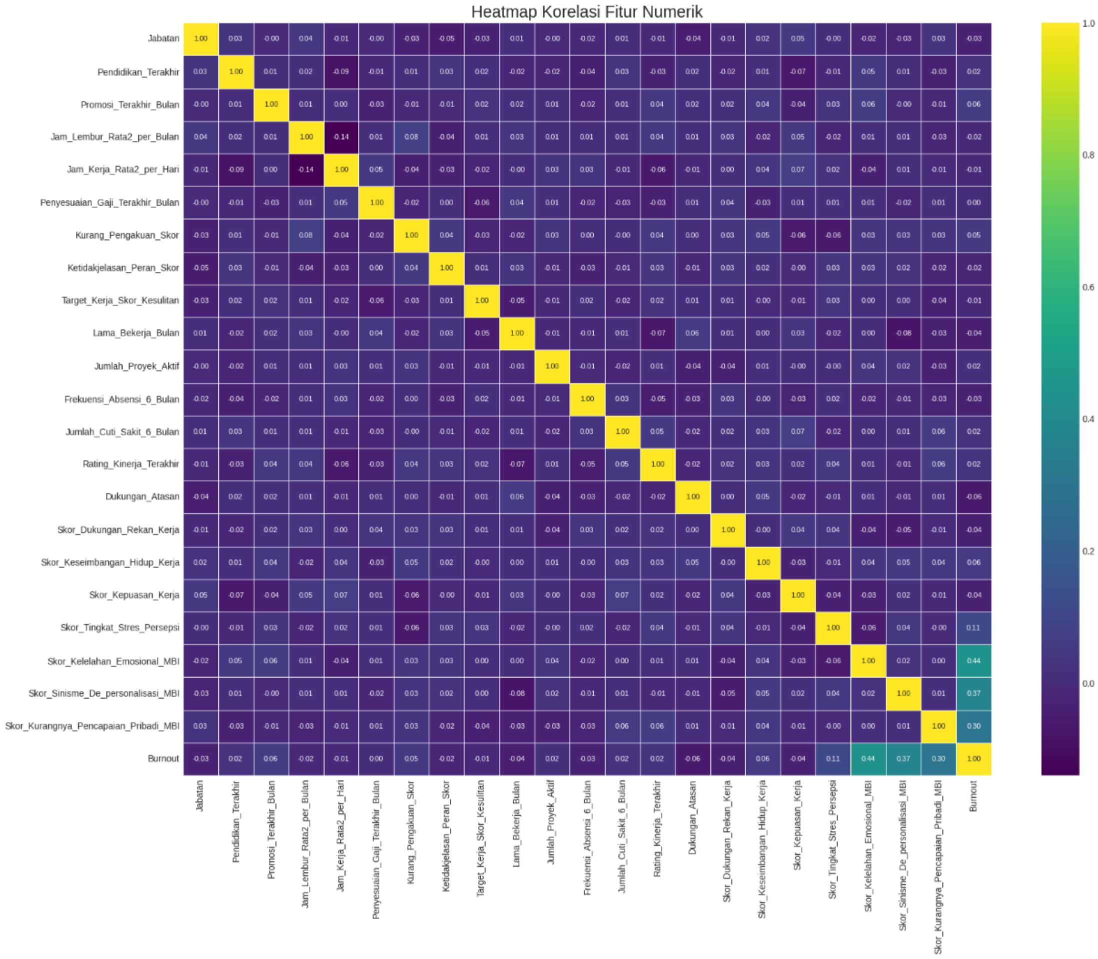

Figure 2. Heatmap of Numerical Correlation for the LSTM Method

The correlation heatmap graphically depicts the interrelationships among the dataset's diverse numerical features. Regions colored a vibrant yellow signify a robust positive correlation, with the coefficient nearing 1.00. This suggests a tendency for the paired variables to increase synchronously. In contrast, the presence of dark or purplish hues denotes either a weak or non-existent correlation, where the coefficient value approaches 0.00.

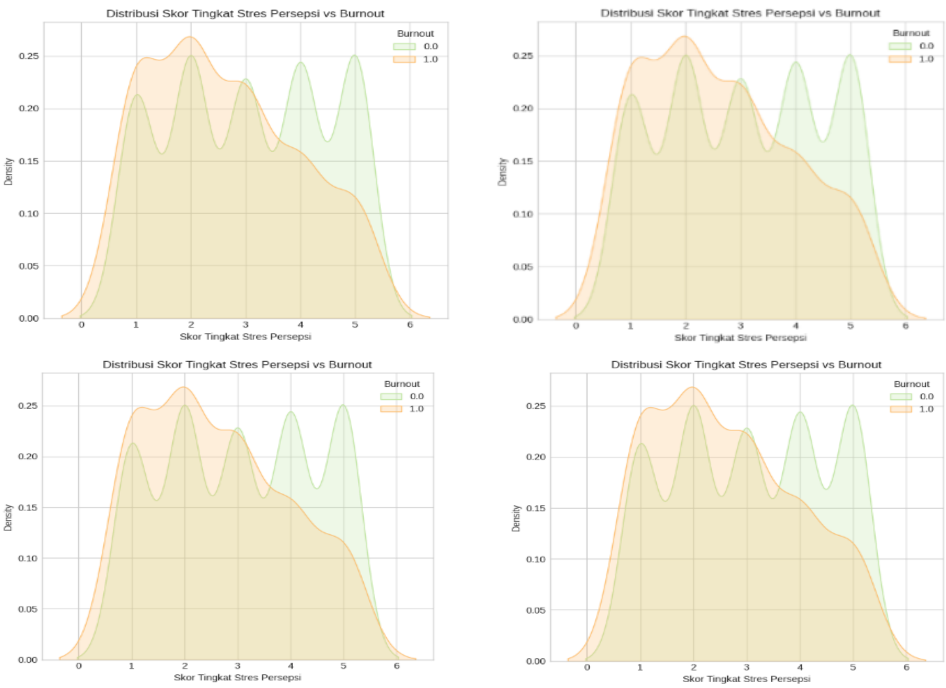

Figure 3. LSTM Method Score Distribution Diagram

Based on the four distribution diagrams of perceived stress score versus burnout, it can be concluded that there is a clear relationship between the level of perceived stress and the tendency to experience burnout, where individuals with higher perceived stress scores show a greater probability of experiencing burnout (represented by a value of 1.0). The probability distribution of burnout tends to increase with increasing stress scores, while in the non-burnout group (value of 0.0), the probability distribution is actually higher at lower stress scores. This pattern is consistent across all diagrams, indicating that perceived stress is a significant predictive factor for the occurrence of burnout.

Table 3. XGBoost Method Evaluation Matrix

Overall Accuracy		**Evaluation Metrics Xtreme Gradient Boosting (XGBoost)**		
Metric	**Value**			
Accuracy	**0,9527**			
Detailed Classification Report				
Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0,84	0,86	0,85	42
1	0,97	0,97	0,97	233
Average Metrics				
Metric	**Precision**	**Recall**	**F1-Score**	**Support**
Macro Average	0,91	0,91	0,91	275
Weighted Average	0,95	0,95	0,95	275

Based on the evaluation table, the Xtreme Gradient Boosting (XGBoost) model showed excellent performance in classification. The model achieved an overall accuracy of 95.27%, which signifies its solid ability to make correct predictions. When we look at the per-class breakdown, its performance is also consistent.

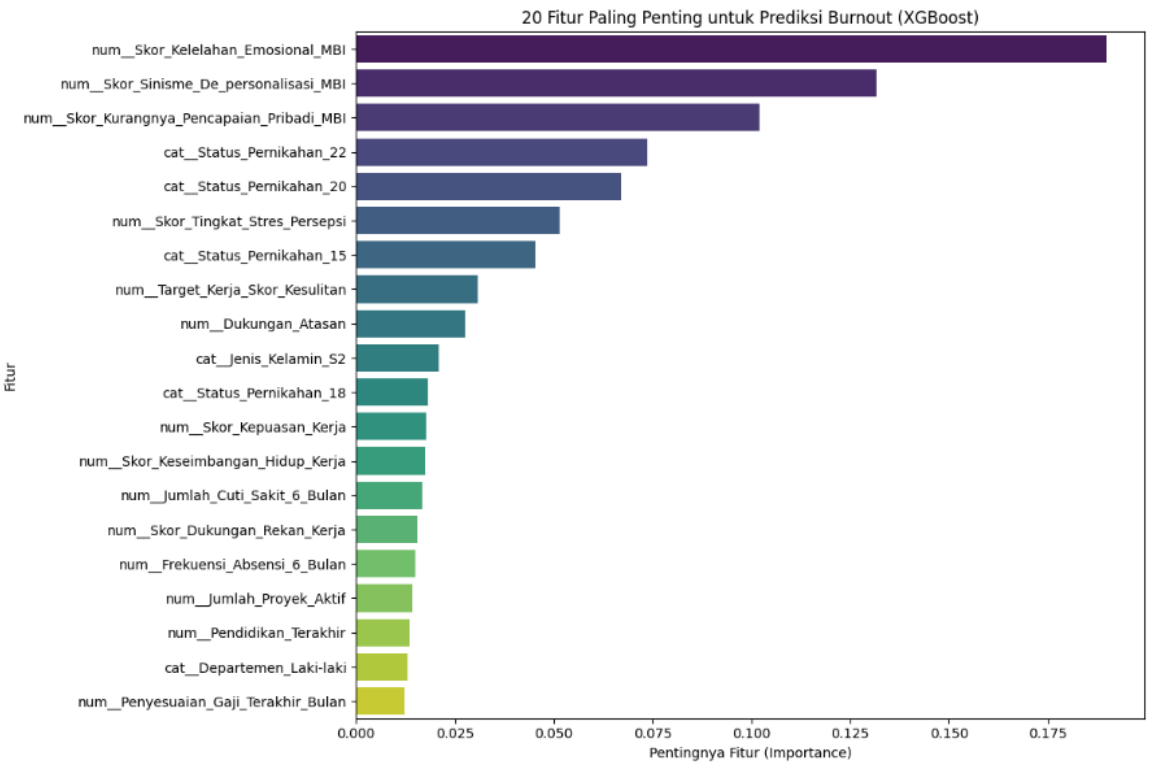

Figure 4. XGBoost Method Feature Importance

Based on the XGBoost model's Feature Importance graph, the most influential features in predicting burnout are the Emotional Exhaustion Score (MBI), the Depersonalization Cynicism Score (MBI), and the Lack of Personal Accomplishment Score (MBI). These three features, which are components of the Maslach Burnout Inventory (MBI), have the highest importance values, significantly surpassing other features. This indicates that the psychological aspects related to exhaustion, cynicism, and feelings of incompetence are the primary predictors of burnout. Other features, such as marital status and stress level, are also relevant, but not as strong as these three main factors.

Table 4. Evaluation Matrix of the EBM Method

Overall Accuracy		**Evaluation Metrics Explainable Boosting Machine (EBM)**		
Metric	**Value**			
Accuracy	**0,9600**			
Detailed Classification Report				
Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0,86	0,88	0,87	42
1	0,98	0,97	0,98	233
Average Metrics				
Metric	**Precision**	**Recall**	**F1-Score**	**Support**
Macro Average	0,92	0,93	0,92	275
Weighted Average	0,96	0,96	0,96	275

Based on the evaluation table, the Explainable Boosting Machine (EBM) model demonstrates a very strong performance in this classification task. The model achieved an overall accuracy of 96%, signaling its excellent ability to predict the outcomes accurately. More specifically, the classification report indicates a solid performance across both classes. The performance analysis of this model shows satisfactory results.

4.1.3 Model Comparison

Table 5. Comparison of Method Performance

Model	Accuracy	Precision	Recall	F1-Score
LSTM (25 epochs)	0.9927	0.9957	0.9957	0.9957
EBM	0.9600	0.9800	0.9700	0.9800
XGBoost	0.9527	0.9700	0.9700	0.9700

The performance comparison of the predictive models clearly establishes the optimized Long Short-Term Memory (LSTM) model (trained for 25 epochs) as the definitively superior performer for the employee burnout prediction task at PT. XYZ. This exceptional efficacy is substantiated by the model's key metrics—accuracy, Precision, Recall, and F1-Score—all surpassing the 0.99 threshold, culminating in a peak accuracy of 99.27%. This outcome significantly eclipses the results achieved by the other evaluated models, namely the Explainable Boosting Machine (EBM), which recorded an accuracy of 96.00%, and the Extreme Gradient Boosting (XGBoost) model, which achieved 95.27%, thereby confirming the LSTM model as the most optimal choice.

4.2 Discussion

The LSTM model's superior performance can be attributed to its architectural capacity to capture temporal dependencies and sequential patterns within employee data. Unlike traditional machine learning algorithms that treat each observation independently, LSTM networks maintain memory of previous states through their gated architecture, enabling recognition of evolving patterns in burnout development over time. The convergence of training and validation accuracy curves without significant divergence, as shown in Figure 1, indicates that the model generalized well to unseen data, avoiding overfitting that often plagues deep learning approaches. The XGBoost and EBM models, while achieving slightly lower accuracy than LSTM, demonstrated their own strengths. As illustrated in Figure 4, XGBoost's feature importance analysis revealed that the three MBI components—Emotional Exhaustion, Depersonalization Cynicism, and Lack of Personal Accomplishment—dominated the prediction process. This finding aligns with established burnout theory, which positions these three dimensions as the core constructs of occupational burnout. EBM's performance at 96.00% accuracy offers a critical advantage: interpretability. Unlike the black-box nature of deep neural networks, EBM constructs additive models where the contribution of each feature can be visualized and understood. The correlation heatmap shown in Figure 2 revealed strong positive correlations among the three MBI components, suggesting that these burnout dimensions tend to co-occur rather than manifest independently. Figure 3's distribution diagrams showing perceived stress versus burnout probability confirmed that stress serves as a reliable early indicator. Employees reporting higher stress scores exhibited substantially higher burnout probabilities, validating the inclusion of stress level as a key predictive feature.

The key contribution of this research is the provision of a robust, data-driven framework for proactive intervention.

1) Proactive Thresholds for Action: The high-accuracy LSTM model can be implemented as the main early warning system, signaling potential burnout with high confidence. HR managers can establish clear predictive probability thresholds (e.g., a probability of burnout ≥ 0.70 triggers a mandatory check-in).

This tiered approach ensures that intervention intensity matches risk level, optimizing resource allocation while avoiding unnecessary intrusion for lower-risk employees.

2) Evidence-Based Intervention: When a warning is flagged, the EBM's interpretable output should be used to provide managerial justification and guide the intervention. For instance, if the EBM identifies Emotional Exhaustion and Low Supervisor Support Score as the top drivers for a specific employee's prediction, the intervention can be narrowly tailored (e.g., adjusting workload and providing specific training to the supervisor). This precision approach increases the likelihood of intervention effectiveness while demonstrating to employees that the organization understands their specific challenges.
3) Human Oversight: The system must include a crucial step for human oversight. The AI provides the prediction and the why, but the HR manager must provide the context and the final decision. This ensures ethical use and prevents an opaque "black box" approach from dictating high-stakes HR actions. HR professionals bring the nuanced judgment necessary to balance algorithmic recommendations against practical constraints and ethical considerations.

The feature importance findings from XGBoost and EBM, particularly as shown in Figure 4, suggest that organizational interventions should prioritize reducing emotional exhaustion and enhancing personal accomplishment feelings. Practical strategies might include implementing mandatory recovery periods after high-intensity projects, establishing clearer role definitions to reduce ambiguity, creating peer recognition programs to address lack of acknowledgment, and training managers to provide more frequent positive feedback.

5. Conclusion and Future Work

This research successfully achieved its primary objective by validating and optimizing the comparative predictive performance of advanced machine learning models, specifically Long Short-Term Memory (LSTM), Explainable Boosting Machine (EBM), and Extreme Gradient Boosting (XGBoost), for early employee burnout detection using real-world operational and psychosocial data at PT. XYZ. The study definitively demonstrated that optimizing burnout prediction through a machine learning approach significantly increases the accuracy of early detection. The Long Short-Term Memory (LSTM) model emerged as the superior performer, achieving a peak accuracy of 99.27% and an F1-Score of 0.9957 at 25 epochs, substantially surpassing the minimum target F1 score of ≥ 0.95. This exceptional performance confirms the LSTM model's capability to capture temporal dependencies and sequential patterns within employee data, making it highly suitable for deployment as the primary early warning system.

While LSTM offered the highest predictive accuracy, the Explainable Boosting Machine (EBM) successfully fulfilled the goal of establishing a transparent Explainable AI (XAI) framework. EBM provided clear interpretability by ranking the core Maslach Burnout Inventory (MBI) components—Emotional Exhaustion, Cynicism, and Lack of Personal Accomplishment—as the most significant trigger variables for burnout, thereby offering HR managers crucial, evidence-based insights for targeted interventions. The XGBoost model, achieving 95.27% accuracy, also demonstrated solid performance and provided valuable feature importance rankings that aligned with established burnout theory. The findings provide a robust, data-driven framework for Human Resource Management, enabling a proactive shift from reactive handling of burnout cases to precise, evidence-based intervention, ultimately supporting organizational well-being and productivity.

Based on the research outcomes, future work should focus on several practical and methodological extensions to ensure the longevity and scalability of the developed framework. First, multi-site replication should be conducted by implementing this methodology across different organizational sites or industries to validate the model's generalizability and assess how varying corporate cultures affect predictive performance. Second, longitudinal drift monitoring systems should be established for continuous monitoring of model performance to ensure the models remain accurate as employee operational patterns and psychosocial factors evolve over time. Third, rigorous fairness and bias audits should be conducted across demographic variables such as age, gender, department, and marital status to ensure the prediction system is equitable and does not disproportionately flag or misclassify specific employee groups. Finally, integration with HR workflows should be pursued by developing a prototype to integrate the EBM and LSTM models directly into existing HR technology systems, facilitating seamless data ingestion, automated risk reporting, and immediate translation of the XAI output into actionable management steps. These future directions will enhance the practical applicability and ethical deployment of AI-driven burnout prediction systems in organizational settings.

References

[1] Mittal, S., Mahendra, S., Sanap, V., & Churi, P. (2022). How can machine learning be used in stress management: A systematic literature review of applications in workplaces and education. *International Journal of Information Management Data Insights*, *2*(2), 100110. https://doi.org/10.1016/j.jjimei.2022.100110

[2] Guldogan, E., Yagin, F. H., Pinar, A., Colak, C., Kadry, S., & Kim, J. (2023). A proposed tree-based explainable artificial intelligence approach for the prediction of angina pectoris. *Scientific Reports*, *13*, Article 22189. https://doi.org/10.1038/s41598-023-49673-2

[3] Lialiou, P., & Maglogiannis, I. (2025). Students' burnout symptoms detection using smartwatch wearable devices: A systematic literature review. *AI Sensors*, *1*(1), 2. https://doi.org/10.3390/aisens1010002

[4] Elnaggar, K., El-Gayar, M. M., & Elmogy, M. (2025). Depression detection and diagnosis based on electroencephalogram (EEG) analysis: A review. *Diagnostics*, *15*(2), 210. https://doi.org/10.3390/diagnostics15020210

[5] Adamopoulos, I., Valamontes, A., Tsirkas, P., & Dounias, G. (2025). Predicting workplace hazard, stress and burnout among public health inspectors: An AI-driven analysis in the context of climate change. *European Journal of Investigation in Health, Psychology and Education*, *15*(5), 65. https://doi.org/10.3390/ejihpe15050065

[6] Dhelim, S., Chen, L., Das, S. K., Ning, H., Nugent, C., Leavey, G., Pesch, D., Bantry-White, E., & Burns, D. (2023). Detecting mental distresses using social behavior analysis in the context of COVID-19: A survey. *ACM Computing Surveys*, *55*(14s), Article 318. https://doi.org/10.1145/3589784

[7] Alam, S. F., & Gonzalez Suarez, M. L. (2024). Transforming healthcare: The AI revolution in the comprehensive care of hypertension. *Clinics and Practice*, *14*(4), 1357-1374. https://doi.org/10.3390/clinpract14040109

[8] Xames, M. D. (2025). Toward a socially sustainable workforce in service operations: Predicting employee burnout with ensemble machine learning. *Sustainable Manufacturing and Service Economics*, 100033. https://doi.org/10.1016/j.smse.2025.100033

[9] Narkbunnum, W., & Hinthaw, K. (2025). Interpretable gradient boosted modeling of employee attrition: A SHAP-based framework for HR analytics. *International Journal of Analysis and Applications*, *23*, 236-236. https://doi.org/10.28924/2291-8639-23-2025-236

[10] Zeng, Y., Zhao, X., Xie, Z., Lin, X., Qi, M., & Li, P. (2025). Prediction of job burnout in nurses based on the job demands-resources model: An explainable machine learning approach. *Journal of Advanced Nursing*. https://doi.org/10.1111/jan.17071

[11] Singh, D. P., Gupta, K., Singh, N., Unni, M. V., Jayasudha, S., & Chauhan, A. (2025). Enhancing enterprise human resource management: Predicting workers' stress for improved workplace satisfaction using a hybrid deep transfer learning approach. In *Hybrid and Advanced Technologies* (pp. 328-333). CRC Press.

[12] Swamy, C. J., & Adarsh, A. (2025). HDLF: Hybrid deep learning framework of DNN and LSTM for workforce sustainability. *Engineering, Technology & Applied Science Research*, *15*(5), 27478-27483. https://doi.org/10.48084/etasr.12640

[13] Yao, Y., Chi, X., Li, J., & Ji, X. (2025, August). Research on employee performance prediction model based on improved long short-term memory network. In *2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON)* (pp. 1-8). IEEE. https://doi.org/10.1109/NMITCON65824.2025.11188406

[14] Wang, W., & Zhou, J. (2024, June). Machine learning-XGBoost analysis of Chinese employees' after-work subjective well-being. In *Proceedings of the 2024 4th International Conference on Artificial Intelligence, Big Data and Algorithms* (pp. 322-326). https://doi.org/10.1145/3690407.3690462

[15] Wu, J., Zhang, C., He, F., Wang, Y., Zeng, L., Liu, W., ... & Gao, F. (2023). Factors affecting intention to leave among ICU healthcare professionals in China: Insights from a cross-sectional survey and XGBoost analysis. *Risk Management and Healthcare Policy*, 2543-2553. https://doi.org/10.2147/RMHP.S432847

[16] Grządzielewska, M. (2021). Using machine learning in burnout prediction: A survey. *Child and Adolescent Social Work Journal*, *38*(2), 175-180. https://doi.org/10.1007/s10560-020-00733-w

[17] Du, B., Li, L., Zhang, K., Gu, Z., Zhang, D., Li, Q., ... & Wu, H. (2025). Construction of a prediction model for sleep disturbances in Chinese nurses and identification of predictive factors: Based on explainable machine learning methods. *BMC Nursing*, *24*(1), 1241. https://doi.org/10.1186/s12912-025-03803-5.