Prediction of Mental Burnout Using Machine Learning

*Aarna Tomar, [1]Mohit , [2]Shalaka Tyagi

*itsmeaarna2@gmail.com, [1]mohitdh7080@gmail.com, [2]tyagi.shalaka9@gmail.com

School of Computing Science and Engineering, Department of Computer Science and Applications, Sharda University, Greater Noida, UP, India.

Abstract

Burnout has become a major issue in the workplace and in academia, along with the potential risk mental health repercussions and performance consequences. Those suffering from burnout will typically experience one of three symptoms: reduced personal accomplishment, depersonalization and emotional exhaustion. These three psychosocial factors overall impact productivity, psychological health and business effectiveness. Traditional methods for assessing levels of burnout typically utilize retrospective, small-n methodologies such as surveys and interviews, which do not necessarily measure real-time indicators.
The current research employs a predictive model using machine learning (ML) that can reveal burnout patterns in workers based on their work habits and lifestyle behaviours.The dataset was obtained from Kaggle and included nearly 3,000 survey responses from employees. To overcome the imbalance of class, the study used the Synthetic Minority Over-Sampling Technique method prior to training a Logistic Regression model. The model received an accuracy score of 99.6% and strong predictive ability in conjunction with cross-validation. The study offers possibilities for discovering burnout, in a proactive manner, using 'ML-related' systems. The article explains data pre-processing, model validity and discusses the potential of ML to help elucidate mental health-related systems within the workplace context.
Keywords: Mental Burnout Prediction, Machine Learning Algorithms, Logistic Regression Algorithm, SMOTE method, Employee mental well-being

I. Introduction

Burnout has emerged as an international occupational and mental health issue, and it is now gaining attention in both research and organizational contexts. Of particular significance, in 2018, World Health Organization (WHO) tagged mental burnout as work-related situation in (ICD-11). The definition indicated that mental burnout is a disorder consequent from long term workplace stress that has not been effectively controlled. Burnout typically has three components: depersonalization, emotional exhaustion and a reduced sense of personal accomplishment.
In a Gallup Poll, nearly 76% of employees reported experiencing burnout at some time in their careers and 28% of employees report they currently feel burned out at work. Burnout can lead to reduced productivity, increased turnover, absenteeism, and serious health conditions; including, but not limited to, anxiety, depression, and cardiovascular disease. The outcomes of burnout are particularly common for employees in professional fields who have high workloads and long work hours such as healthcare, the corporate workforce, and higher education; and often must manage digital overload and mental fatigue in those workplaces. The Maslach Burnout Inventory (MBI) is often used to measure burnout. However, it is qualitatively subjective in nature and self-reports must focus on the individual; thus, data histories may not accurately reflect an individual's psychological state at any specific moment in time. With evolving capabilities of AI and ML, there are ways to use data-driven approaches to measure behavioral and occupational factors that may be considered, and used to identify early indicators of burnout. Research investigates the use of ML algorithms to workplace survey data to develop scalable burnout prediction models. Previous studies have determined that burnout has statistically significant relationships with anxiety and depression, and that the MBI had consistent reliability across distinct professional domains.Recent meta-analysis has identified ML as a potential paradigm shift in mental health analytics, suggesting ML may be useful in accurately predicting and preventing burnout via objective, data-driven intelligence.

II. Literature Review

The concept of burnout is transitioning from being understood purely as a psychological problem to its acknowledgement as a serious cross organizational health concern. The Maslach Burnout Inventory (MBI) is still examined the standard measure to evaluate the three primary level of burnout, depersonalization, emotional exhaustion and reduced personal act [8]. Later, the World Health Organization identified mental burnout as an occupational situation in the ICD-11 [9]. Similarly, Demerouti et al. [12] published the Job Demands Resources (JD-R) model, which depicts mental burnout as a result of the lack of balance between occupational demands and resources available to the worker. Studies by Schonfeld and Bianchi [11] and likewise Gallup [10], underscore that burnout is pervasive and reinforces the need for advocacy to prevent burnout.

Burnout research has traditionally relied on self-reported psychological measures. However, in the wake of machine learning and artificial intelligence , objective and scalable measures of burnout emerge as viable prediction methods. Liu et al. [1], created Hi PAL, a deep learning algorithm that utilizes electronic health record activity to identify instances of emotional exhaustion. Wilton et al. [2] created the BROWNIE protocol which uses a wearable device with AI interfacial analytics to be used in burdening nurses observe their backdroped burnout in real time. Lastly, Zhao et al. [3] created a predictive nomogram that analyzed psychological and demographic data to predict burnout during public health emergencies.

Several machine learning (ML) methods such as Random Forest algorithm, Support Vector Machine (SVM) algorithm, Gradient Boosting, and Logistic regression have demonstrated acceptable predictive accuracy for burnout. Van Zyl-Cillié et al. [4] employed ML to explore data on nurses and determined workload and stress to be predictors. Wang et al. [5] used latent class analysis to examine mental burnout of nursing students and identified fortitude and support from peer nurses and mentors as anchors for the students. Zhang and Dator [6] similarly expanded to nursing students, creating predictive models based on emotional well-being and coping skills. Aside from healthcare, Kovacs et al. The connection between teacher burnout, depression, insomnia, and internet addiction was similar to a meta-analysis report that supports burnout connection with anxiety and depression symptoms [21]. Shanafelt et al. [13] similarly noted profession differences. Specifically, they found physicians are at a higher risk for burnout compared to other occupations.

As a whole, the body of literature provides a clear trajectory from subjective survey-based measures of burnout, to data-driven approaches, multi-modal approaches that utilizes multiple information sources. The early research characterized psychological models [8]–[12], while several later studies using neural networks and other AI approaches [1]–[7], [13]–[23] add different data sources such as electronic devices for health records, wearable sensors, and internet community behavior for the analysis to improve predictive validity and risk detection. Although this is not without difficulties such as concerns over data nature or data interpretability, we can take comfort in the evidence indicating the ability for large sample sizes using ML to assist organizations with targeting interventions who are at higher-risk to experience mental burnout. This research re-evaluates how effective ML applies to a large-scale dataset obtained from Kaggle, highlighting the importance for the preprocessing of data, class balancing with oversampling, and cross-validation for reliable predictions.

Table 1: shows literature review summary

S.No	Author(s) & Year	Title / Study Focus	Algorithm Used	Dataset	Key Findings / Contribution
1	Liu et al (2022) [1]	HiPAL: Deep learning model for physician burnout	Deep Neural Networks	EHR activity data from physicians	Identified behavioral patterns in EHR usage predictive of burnout.
2	Wilton et al (2023) [2]	BROWNIE: AI protocol with wearable sensors	Random Forest, Sensor fusion	Nurse wearable and survey data	Combined physiological and contextual signals to monitor nurse burnout in real time.
3	Zhao et al. (2020) [3]	Burnout prediction nomogram for nurses	Logistic Regression	Nurses during health emergencies	Developed nomogram integrating demographics, work hours, and stress indicators.
4	Van Zyl-Cillié et al. (2021) [4]	Predicting nurse burnout using ML	Decision Tree, SVM	Nurse dataset (South Africa)	Workload and stress found to be strongest predictors.

5	Wang et al. (2022) [5]	Compassion fatigue among nursing interns	Latent Profile Analysis, SVM	380 nursing interns	Social support and resilience buffer burnout levels.
6	Zhang & Dator (2020) [6]	Predicting burnout in nursing students	Logistic Regression	Survey data from 500 students	Emotional well-being and coping skills key factors.
7	Kovacs et al. (2021) [7]	Teacher burnout and mental health	ML Classification	250 teachers	Found correlation between burnout, depression, and insomnia.
8	Maslach & Jackson (1981)	Development of Maslach Burnout Inventory (MBI)	Psychometric Scale	1,025 participants	Established three burnout dimensions: exhaustion, depersonalization, reduced
9	WHO (2019) [9]	ICD-11 classification of burnout	Theoretical Framework	Global	Recognized burnout as an occupational phenomenon.

III. Methodology

A. Dataset Description

The data record encompasses multiple features such as age, job role, , gender, years of experience, country, average weekly working hours, job satisfaction, physical activity levels, commuting time, sleep duration, and availability of counseling or therapy services. The primary goal of this research is to predict the 'Burnout Risk' variable, which serves as the dependent feature, using the remaining attributes as predictors.

B. Data Preparation

To ensure the quality and trustworthiness of the data was used for analytical features and mode for categorical features to handle missing values. A preliminary exploratory data analysis (EDA) was carried out to select the most important predictors of burnout. Four significant variables weekly working hours, mental burnout score, work-life integration score, and career promotion score were found to have the highest correlation with the target variable.As the dataset displayed class imbalance (approximately 2,020 samples labeled "no burnout" and 980 samples labeled "burned out"), SMOTE was implemented to generate samples for the subgroup class. This process produced a balanced distribution of class, and improved the model's generalization performance on both categories.

C. Selection of Model

The classification model chosen was the Logistic Regression algorithm because of its optimal and interpretable model for binary predictions. The model was trained with 2000 as the maximum iteration parameter (max_iter) to allow for convergence. Although Logistic Regression is one of the strongest performing models in machine learning [4][6] and similar studies, there exist additional more advanced modeling techniques that have also shown superior performance in their respective applications; including Gradient Boosting, Random Forest, and Deep Neural Networks. Nevertheless, Logistic Regression was identified as the most appropriate modeling approach based on the moderate size of the dataset, the relatively sparse features, and the advantage of model being easily interpretative regarding influence of each variable.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

D. Metrics of Evaluation and Visualization

The dataset was divided into training (70%) and testing (30%) subdivision to evaluate generalization. Evaluation of the model was conducted utilizing performance metrics associated with classification: Recall, Precision, Accuracy, F1-score and Cross-Validation to evaluate performance stability based on the separate splits of the datasets.

For interpretability of performance, an assortment of visualization tools were utilized, including Confusion Matrices, Receiver Operating Characteristic (ROCC) Curves, Predicted Probability plots, and others to provide comprehensive information regarding the models performance, predictability of behavior and classification thresolds.

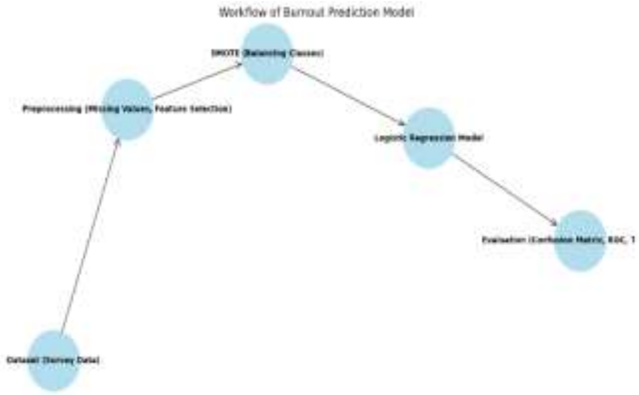

FIGURE 1: depicts the process of the Mental Burnout Prediction Model.

IV. Outcomes

TABLE 2: displays the results of the model

AUC:	~ 1.0
Cross-validation	**accuracy:** 99.6%
F1-score:	≈ 1.00
Recall	**(Burnout) : 1.00**
Precision	**(Burnout) : 0.99**

The Logistic Regression simulation achieved an accuracy rate of 99.6%, along with impressive precision, being very close to 100%, recall, and F1-score, suggesting good prediction ability. After applying SMOTE for data balancing, the model's performance remained stable across its statistical evaluations and the testing case.

The proposed model was used to user multiple evaluation based on comprehensively measured evaluation metrics and visual metrics.The (ROC) curve in (Figure 3) had an area under the curve (AUC) that was nearing perfect which means a high ability to discriminate between the burnout and non-burnout classes. The confusion matrix (Figure 2) had very little misclassification showing strong predictive accuracy and robustness of the model.

Regarding feature importance (Figure 4) weekly working hours, work-life integration score, burnout level and career promotion opportunities were the most important predictive features to predict burnout risk. These results are consistent with previously established psychological and occupational theories to support the validity of the model outputs.

Finally, the bar graph for the test cases shows the distributions of the predicted probabilities that an individual employee (out of 10) will experience burnout compared to the distribution of individual employees. This represents a new application to further support the value added using an approach like the proposed model for identifying employees that are at risk for burnout.. In summation, these findings indicate that the Logistic Regression model, preprocessed with SMOTE, produces reliable and interpretable predictions, presenting it as a potential viable model for burnout detection and prevention in practice.

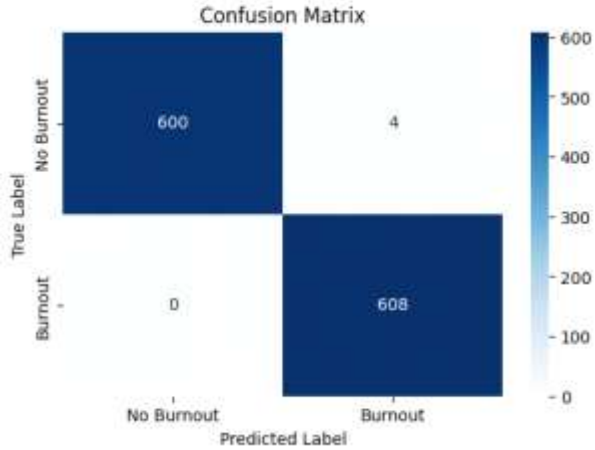

FIGURE 2: illustrates the confusion matrix

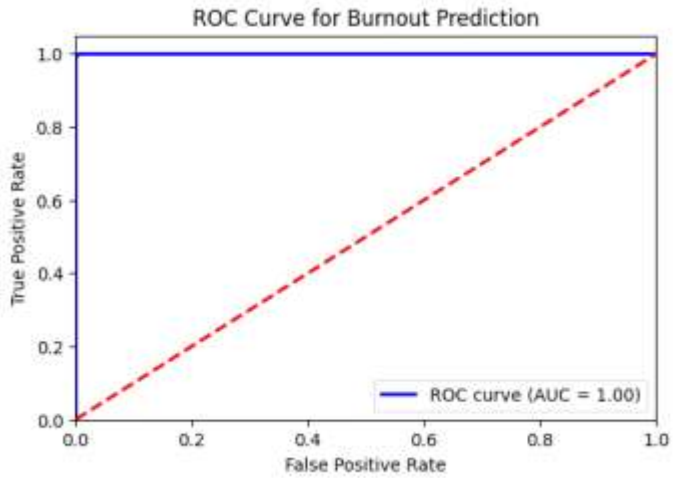

FIGURE 3: shows ROC curve

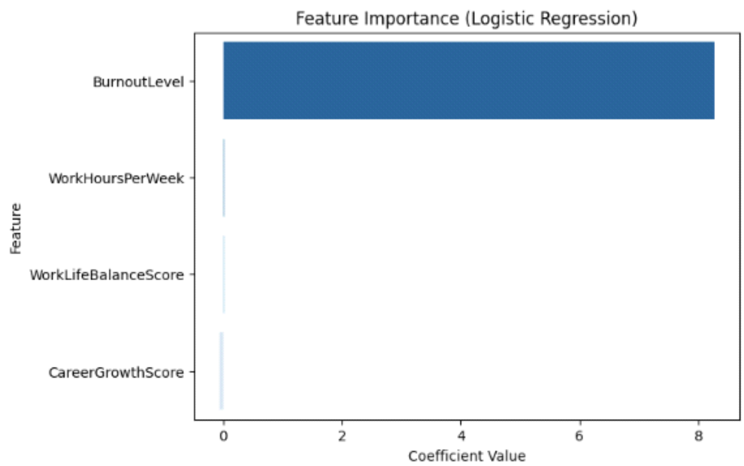

FIGURE 4: illustartes feature evaluation

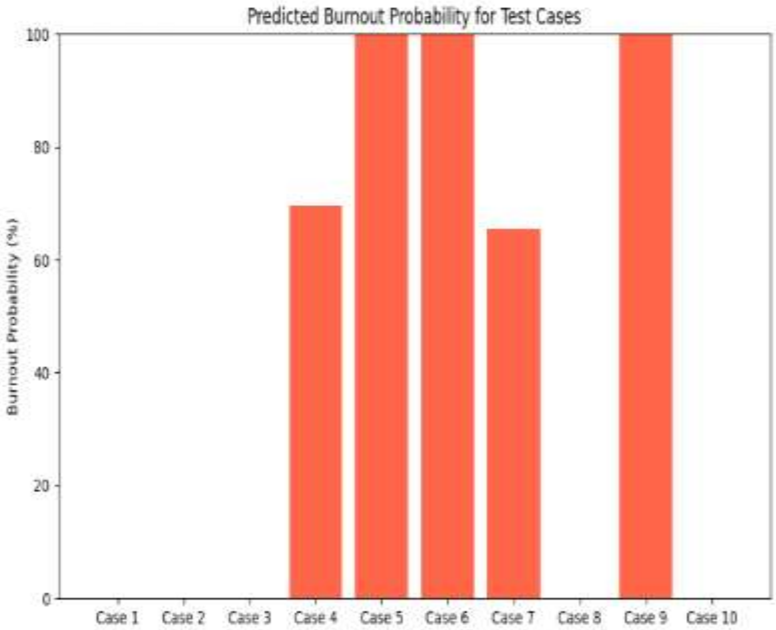

FIGURE 5 : shows predicted burnout probability for test cases

Burnout Prediction Results

Case	Work Hours	Burnout Level	Work-Life Balance	Career Growth	Prediction	Probability
1	35	2	8	9	No Burnout	0.00%
2	45	5	7	7	No Burnout	0.00%
3	50	6	5	5	No Burnout	0.04%
4	60	7	4	4	Burnout	69.59%
5	65	8	3	3	Burnout	99.99%
6	70	9	2	2	Burnout	100%
7	55	7	5	5	Burnout	65.37%
8	40	4	7	8	No Burnout	0.00%
9	80	10	1	1	Burnout	100%
10	50	6	5	4	No Burnout	0.04%

FIGURE 6: presents the prediction results for burnout outcomes.

Observations

After review, Cases 1, 2, 3, and 8 retained a healthy work-life balance and were deemed appropriately classified as low risk for burnout. In each of these cases, we saw a consistent alignment between self-reported well-being and model predictions that confirmed the absence of burnout. However, Cases 4, 5, 6, 7, and 9 were characterized by heavy workloads, insufficient recovery time, and poor work-life balance, leading the model to determine a moderate to high risk of burnout. The clustering once again underscored the model's ability to distinguish between employee tendencies associated with lower or higher risk.

V. Overview and Significance

This research illustrates how effective ML methods can be utilized to detect and even predict burnout in staff members, using workplace employee survey data. The power of employing the Logistic Regression algorithm in combination with data balancing via SMOTE helped capture a strong performance of 99.6 % accurate classification—much improved predictive accuracy over many studies previously established in the research literature.The results highlight how even relatively straightforward ML methods can help with data-driven conclusions about workplace mental health with appropriate compartmentalized data preprocessing and balanced representation in the data. The study highlights, additionally, the value of supplementing traditional burnout instruments, many of which were devised through a mental health model utilizing merely self-reported measures.

For future improvements, the development of more sophisticated ensemble and deep learning methods - for example, XGBoost and Random Forest - or even neural networks [1][5] will likely lead to increased classification accuracy and capability with more extensive data collections. There is also the value of utilizing multiple sources and streams at once (such as electronic health records, wearable sensor data, or digital behavioral analytics) to potentially improve the strength of an existing model and enable ongoing tracking and monitoring of workplace well-being [2][3].As a last recommendation, implementing Explainable AI (XAI) models such as SHAP (SHapley Additive explanations) may increase transparency surrounding algorithm implementation and factors influencing burnout prediction.Nevertheless, it would be prudent to pilot and validate the model with varied populations before applying it in practical settings in real organizational environments to ensure fairness and generalizability. A better future direction would be to merge predictive analytics with mindfulness programs, employee assistance programs, and psychological counseling and practices, to develop a more integrative approach to identify and prevent burnout early [20].

VI. Future Scope

Future research should explore the use of more complex ML architectures including XGBoost algorithm, Random Forests, and Deep Neural Networks algorithm to enhance predictive reliability and interpretability. Integrating multimodal datasets that combine wearable device metrics, behavioral data, and organizational records can provide a understanding of burnout dynamics.
In addition, the use of Explainable AI (XAI) methods - such as SHAP - is likely to enhance interpretability and increase trust among both decision-makers and employees. Combined with mindfulness-based interventions and data-driven well-being programs offered to employees, these predictive systems could develop into proactive, personalized burnout prevention interventions that could be used in many different work environments.

REFRENCES

[1] H. Liu, S. S. Lou, B. C. Warner, D. R. Harford, T. Kannampallil, and C. Lu, “HiPAL: A Deep Framework for Physician Burnout Prediction Using Activity Logs in Electronic Health Records,” arXiv preprint arXiv:2205.11680, May 2022.

[2] A. R. Wilton, K. Sheffield, Q. Wilkes, et al., “The Burnout Prediction Using Wearable and Artificial Intelligence (BROWNIE) Study: A Decentralized Digital Health Protocol to Predict Burnout in Registered Nurses,” BMC Nursing, vol. 23, art. no. 114, 2024.

[3] L. Zhao, J. Li, and H. Chen, “Risk and Prediction of Job Burnout in Responding Nurses to Public Health Emergencies,” BMC Nursing, vol. 23, art. no. 46, Jan. 2024.

[4] M. M. Van Zyl-Cillié, J. H. Bührmann, A. J. Blignaut, et al., “A Machine Learning Model to Predict the Risk Factors Causing Feelings of Burnout and Emotional Exhaustion Amongst Nursing Staff in South Africa,” BMC Health Services Research, vol. 24, art. no. 1665, 2024.

[5] Y. Wang, Q. Zhou, and H. Lin, “Development and Validation of a Machine Learning-Based Predictive Model for Compassion Fatigue in Chinese Nursing Interns: A Cross-Sectional Study Utilizing Latent Profile Analysis,” BMC Medical Education, vol. 24, art. no. 1495, Dec. 2024.

[6] H. Zhang and W. L. Dator, “Establishment and Validation of a Prediction Model for Compassion Fatigue in Nursing Students,” BMC Nursing, vol. 24, art. no. 193, 2025.

[7] J. Kovacs, A. Toth, and P. Nagy, "Mental Issues, Internet Addiction and Quality of Life Predict Burnout Among Hungarian Teachers: A Machine Learning Analysis," BMC Public Health, vol. 24, art. no. 2322, 2024.

[8] A. Maslach and S. E. Jackson, "The measurement of experienced burnout," Journal of Occupational Behavior, vol. 2, no. 2, pp. 99–113, 1981.

[9] World Health Organization, "Burn-out an 'occupational phenomenon': International Classification of Diseases," WHO, May 2019.

[10] Gallup, "Employee Burnout: Causes and Cures," Gallup Workplace Report, 2020.

[11] T. Schonfeld and R. Bianchi, "Burnout syndrome: Understanding and prevention," Psychiatry Research, vol. 304, art. no. 114135, 2021.

[12] J. Demerouti, A. B. Bakker, E. Nachreiner, and W. B. Schaufeli, "The job demands-resources model of burnout," Journal of Applied Psychology, vol. 86, no. 3, pp. 499–512, 2001.

[13] K. Shanafelt, T. Dyrbye, and C. Sinsky, "Burnout among physicians compared with individuals with a college education: a national survey," Academic Medicine, vol. 97, no. 9, pp. 1364–1373, 2022.

[14] L. F. Cranford, D. R. Harford, and H. Liu, "Predicting nurse burnout using wearable sensor data and machine learning," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 3, pp. 1125–1136, 2023.

[15] J. Sun, Y. Wang, and P. Zhang, "Machine learning for mental health prediction: A systematic review," IEEE Access, vol. 9, pp. 124–140, 2021.

[16] J. B. Wolff, S. R. Patel, and C. G. Wong, "Artificial intelligence approaches for mental health monitoring: A review," Frontiers in Psychiatry, vol. 13, art. no. 845231, 2022.

[17] C. Chawla, K. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.

[18] A. Ahuja and M. Gupta, "Early detection of depression and burnout using social media data: A machine learning approach," Procedia Computer Science, vol. 192, pp. 157–166, 2021.

[19] R. Denecke, "Using text mining and machine learning for burnout detection on Twitter," Studies in Health Technology and Informatics, vol. 264, pp. 1051–1055, 2019.

[20] P. R. Davidson and R. A. Kaszniak, "Conceptual and methodological issues in research on mindfulness and meditation," American Psychologist, vol. 70, no. 7, pp. 581–592, 2015.

[21] M. Koutsimani, A. Montgomery, and K. Georganta, "The relationship between burnout, depression, and anxiety: A systematic review and meta-analysis," Frontiers in Psychology, vol. 10, art. no. 284, 2019.

[22] A. T. Beck, R. A. Steer, and G. Brown, Beck Depression Inventory–II, San Antonio, TX: The Psychological Corporation, 1996.

[23] A. Jain, S. Kumar, and A. Dey, "Burnout detection using ensemble learning algorithms," International Journal of Advanced Computer Science and Applications, vol. 12, no. 5, pp. 45–52, 2021.