UDC 004.8:612.8:616.89

Computational Approaches for Emotional Burnout Detection: Machine Learning and Deep Learning Evaluation

Mushta S. A.[1], *Mushta I. A.*[1], *Popov A. O.*[1,2], *Lysenko O. M*[1], *Tukaiev S. V.*[3]

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2]Ukrainian Catholic University, Lviv, Ukraine
[3]Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

E-mail: *simonmushta@gmail.com*

This study focuses on the development and evaluation of computational techniques for classifying emotional burnout based on quantitative data analysis. Instead of relying on subjective psychometric questionnaires, we investigate the applicability of machine learning and deep learning algorithms to structured datasets. Several classical methods, including logistic regression, random forest, and gradient boosting, were systematically compared with a Deep Learning (DL) ensemble model. To enhance robustness, preprocessing steps such as feature selection, data balancing, and resampling were applied. The deep learning architecture, incorporating focal loss and adaptive threshold optimization, achieved the best performance. On 5-fold cross-validation, the proposed DL model obtained an overall accuracy of 86.3%, with precision of 0.815 / 0.887, recall of 0.786 / 0.904, and F1-scores of 0.800 / 0.895 for the negative and positive classes respectively. The results demonstrate that advanced computational models can provide scalable and generalizable tools for automatic detection tasks, forming a technical foundation for future integration into applied research and occupational health monitoring systems.

Keywords: emotional burnout; machine learning; deep learning; electroencephalography; maslach burnout inventory; feature selection; classification; mental health diagnostics

DOI: 10.64915/RADAP.2026.103.69-77

Introduction

Emotional burnout is a multidimensional psychological syndrome characterized by emotional exhaustion, depersonalization, and a reduced sense of personal accomplishment. Since its first description by H. J. Freudenberger in 1974, the concept has been widely studied, yet its diagnosis remains inconsistent due to the absence of universally accepted biomarkers and objective reference values. Traditional diagnostic approaches, such as the Maslach Burnout Inventory (MBI) and similar self-report questionnaires, provide valuable insights but are limited by subjectivity, cultural bias, and lack of proven validity.

Recent research has increasingly focused on exploring physiological and neurocognitive correlates of burnout. Structural brain imaging has revealed cortical thinning and altered subcortical volumes in individuals exposed to chronic occupational stress. Electrophysiological studies using Electroencephalography (EEG) have identified spectral alterations – such as reduced alpha activity and changes in Event-Related Potentials (ERPs) – that may reflect the neural basis of burnout-related emotional and cognitive impairments [1–3]. However, despite these advances, no consistent biological marker has been established, and burnout continues to overlap significantly with related conditions such as depression and stress-related disorders.

To address the limitations of conventional diagnostic methods, Machine Learning (ML) and Deep Learning (DL) approaches are emerging as powerful tools for analyzing complex, multidimensional data. ML methods such as logistic regression, random forests, and gradient boosting have demonstrated the ability to predict burnout levels from survey data with improved accuracy compared to traditional statistical methods [4,5]. At the same time, DL architectures offer the potential to capture non-linear dependencies in high-dimensional features, particularly in neurophysiological data such as EEG power spectral density [6, 7]. Ensemble models, threshold optimization, and advanced loss functions further enhance classification performance, enabling more robust and generalizable models.

This study explores the integration of ML and DL techniques for emotional burnout detection. By combining feature selection strategies, oversampling techniques, and optimized architectures, the work aims to improve classification accuracy and identify discriminative features that may serve as objective markers of burnout. The results contribute to the growing body of research seeking computational approaches for early detection and prevention of burnout, which is increasingly recognized as a critical occupational health concern.

1 Analysis of Research and Publications

Emotional burnout has been widely recognized as a multidimensional occupational phenomenon associated with chronic stress, emotional exhaustion, depersonalization, and reduced professional efficacy. Despite its prevalence, the diagnosis of burnout remains problematic due to the lack of standardized models, reference values, and biological markers. As highlighted in [8], the absence of a distinctive biomarker is not unexpected given the significant overlap between burnout and depressive syndromes. Thus, a major research direction involves the search for objective diagnostic tools, including psychometric, neuroimaging, and electrophysiological approaches.

The problem of emotional burnout detection increasingly relies on computational methods, as traditional psychometric instruments lack objectivity and reproducibility. Recent research has demonstrated that burnout is accompanied by measurable neurophysiological changes. For instance, a study in [9] revealed reduced functional connectivity in frontal and midline brain regions within the alpha3 range (11–13 Hz) during resting-state EEG with eyes open, pointing to impaired emotional regulation and cognitive fatigue.

Machine Learning (ML) approaches have been widely applied for stress and resilience detection. In particular, interpretable ML models based on EEG spectral features achieved high performance; for example, an XGBoost classifier obtained 95% accuracy in assessing cognitive resilience in military personnel, while SHAP analysis highlighted the importance of delta and alpha activity in frontal and parietal regions [10]. Classical algorithms such as logistic regression and other traditional classifiers applied to spectral EEG features derived from Fourier analysis have shown effective performance in stress detection tasks, achieving reliable classification on accuracy in experimental setting [11].

Deep Learning (DL) architectures are particularly effective for modeling nonlinear temporal and spatial dependencies in EEG data. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have demonstrated strong performance in mental stress classification, with LSTM-based models achieving accuracy of up to 94% using EEG signals, while CNNs are more effective when applied to spectral or topographical representations [12]. Further advances are evident in multimodal architectures: the Early Fusion Network (EF-Net) model combined EEG and functional Near-Infrared Spectroscopy (fNIRS) data to achieve superior cross-subject recognition of mental states, confirming the robustness of CNN-based designs in multimodal learning tasks [13]. Graph Neural Networks (GNNs) represent a novel trend in EEG analysis, enabling explicit modeling of connectivity between electrodes. For example, Graph Convolutional Networks (GCNs) combined with attention mechanisms improved classification accuracy and noise robustness when applied to EEG, Electromyography (EMG), and Electrocardiography (ECG) data [14]. Another study employed GCNs with EEG connectivity features to classify four mental states (rest, control-alert, stress, and stress mitigation), achieving 80.95% accuracy [15]. Beyond stress detection, multimodal GNN-based models have been shown to identify treatment-resistant depression by integrating functional Magnetic Resonance Imaging (fMRI) and EEG brain networks, highlighting their clinical relevance [16]. Moreover, transformer-based hybrid models such as EEG Mind-Transformer leverage graph structures with temporal attention to capture complex spatiotemporal EEG patterns, improving both scalability and interpretability [14].

Technological advances also expand practical applicability. It was developed a wearable, wireless, and skin-like electronic tattoo (e-tattoo) that adheres to the forehead and continuously measures brain and eye electrical activity (EEG and Electrooculography (EOG)). The device is lightweight and unobtrusive, allowing subjects to perform mental tasks without the restrictions of conventional EEG headsets. In validation studies, the e-tattoo was able to capture brain dynamics associated with increasing mental workload, showing characteristic shifts in delta, theta, alpha, and beta rhythms [17]. In parallel, low-cost EEG devices have enabled accessible experimentation and real-time ML/DL integration in applied settings.

Existing research shows that burnout is linked to function changes of brain and altered EEG dynamics. Questionnaires give lack objectivity and reliability, motivating EEG- and ML-based methods. Yet no diagnostic protocols exist, underscoring the need for robust computational models. Machine and deep learning approaches – from interpretable ML to CNN, LSTM, GNN, and transformers – offer scalable and generalizable solutions for EEG-based burnout detection.

2 Formulation of the Research Task

Given the absence of a single reliable biological marker of burnout, recent studies increasingly emphasize combining psychometric assessment with neurophysiological signals and computational modeling. EEG, with its high temporal resolution and sensitivity to cognitive and emotional states, presents a particularly promising modality for identifying objective burnout-related markers. The current research builds on prior findings that EEG spectral characteristics – especially in the alpha, theta, and beta bands – correlate with stress and burnout symptoms [2,10–13]. Additionally, psychometric instruments such as the MBI questionnaire provide valuable but subjective estimates of burnout dimensions, including emotional exhaustion, depersonalization and personal accomplishment. To overcome the limitations of self-report tools, integrating these data with machine learning and deep learning methods may enable more accurate and reproducible diagnosis.

In our study, a feature table was constructed from MBI questionnaire scores and EEG spectral measures [19, 20]. Participants were dichotomized into two groups: those with developed emotional burnout (label "1") and those without it (label "0"). Following [12], scores of 0–36 were classified as “burnout is not formed” and scores of 67–100 as “burnout is formed”. For each participant, normalized Power Spectral Density (nPSD) values were computed across 19 EEG channels (Fp1, Fp2, F3, F4, F7, F8, T3, T4, C3, C4, T5, T6, P3, P4, O1, O2, Fz, Cz, Pz) and 14 predefined frequency ranges (delta, theta1, theta2, alpha1, alpha2, alpha3, beta1, beta2, gammaLOW, gammaHIGH, theta, alpha, beta, gamma). The frequency ranges for each band are presented in Table 1.

TABLE 1 Frequency boundaries of EEG rhythms

High frequency boundaries	
‘delta’ rhythm	(1 - 4) Hz
‘theta’ rhythm	(4 - 7.5) Hz
‘alpha’ rhythm	(7.5 - 13) Hz
‘beta’ rhythm	(13 - 35) Hz
‘gamma’ rhythm	(35 - 100) Hz
Low frequency boundaries	
‘theta1’ rhythm	(4 - 6) Hz
‘theta2’ rhythm	(6 - 7.5) Hz
‘alpha1’ rhythm	(7.5 - 9.5) Hz
‘alpha2’ rhythm	(9.5 - 11) Hz
‘alpha3’ rhythm	(11 - 13) Hz
‘beta1’ rhythm	(13 - 20) Hz
‘beta2’ rhythm	(20 - 35) Hz
‘gammaLOW’ rhythm	(35 - 50) Hz
‘gammaHIGH’ rhythm	(50 - 100) Hz

The research task is to develop and evaluate machine learning and deep learning models for the objective diagnosis of emotional burnout, based on EEG spectral characteristics and psychometric measures. The aim is to establish a computational framework that integrates subjective assessment tools with neurophysiological biomarkers, providing a foundation for more accurate and clinically applicable burnout diagnostics.

3 Description of the Research Material

3.1 Dataset Description and the Challenge of Small, Imbalanced Samples

This study used a combined dataset of psychometric and neurophysiological measures from participants screened for emotional burnout. EEG was recorded monopolarly using a 19-channel Neurocom system (XAI-MEDICA, Ukraine) during a 3-minute resting state with eyes closed. Silver/silver chloride electrodes were positioned according to the International 10–20 system at symmetrical anterior frontal (Fp1, Fp2), frontal (F3, F4, Fz, F7, F8), central (C3, C4, Cz), parietal (P3, P4, Pz), occipital (O1, O2), and temporal (T3, T4, T5, T6) sites, referenced to interconnected ear electrodes, with interelectrode impedance kept below 5 $k\Omega$. Signals were sampled at 500 Hz [21].

Initially, 752 participants (students and employees from Taras Shevchenko National University of Kyiv and the National Aviation University, Kyiv, Ukraine) were recruited (209 males, mean age = 19.2 ± 2.55 years; 543 females, mean age = 18.28 ± 2.19 years), but after excluding data from individuals with poor EEG quality and excessive artifacts, the final sample consisted of 257 participants. The study was approved by the Local Ethics Committee of the Educational and Scientific Centre “Institute of Biology and Medicine” at Taras Shevchenko National University of Kyiv, and written informed consent was obtained from each volunteer in accordance with the Declaration of Helsinki. Exclusion criteria included clinical manifestations of mental disorders or cognitive impairment, as well as the use of psychoactive medication, drugs, or alcohol [21].

Psychometric assessment was performed using the MBI burnout inventory [19, 20], where cumulative scores across the subscales of emotional exhaustion, depersonalization and personal accomplishment were calculated and normalized to a 0–100 range. After dichotomizing the participants into two groups (“stage not formed” and “stage formed”), a dataset was obtained. The pie charts Fig. 1 and Fig. 2 reveal that the dataset is predominantly composed of individuals diagnosed with Burnout, accounting for 65.0% of participants, while only 35.0% are healthy controls. This notable class imbalance must be addressed to ensure an accurate diagnosis. Additionally, the gender distribution shows a considerable disparity.

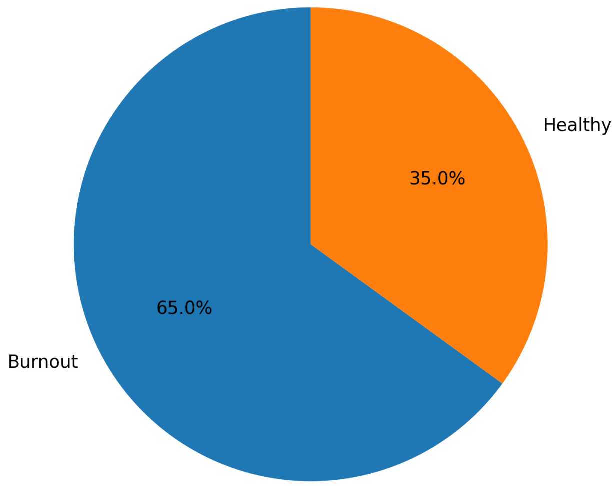

Fig. 1. Patients cohort

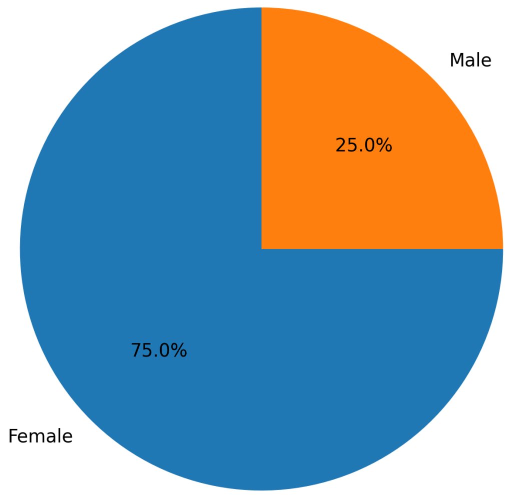

Fig. 2. Patients sex at birth

The final dataset consisted of 269 feature vectors with 80 predictors each, including PSD measures, participant sex, and class labels. However, the dataset posed two significant challenges: (i) its small sample size, which increases the risk of overfitting, and (ii) class imbalance, since the distribution between burnout and non-burnout cases was skewed. These issues are frequently reported in EEG-based diagnostic research [19–21], where the difficulty of collecting large, balanced samples often necessitates advanced preprocessing, feature selection, and resampling strategies. Addressing these limitations required careful experimental design.

3.2 Machine Learning Aproach

To establish a baseline and assess the predictive value of psychometric and EEG-derived features, a series of ML classifiers were implemented. The pipeline was designed to integrate feature selection, class rebalancing, and nested cross-validation for robust evaluation. The following methodological steps were undertaken:

Correlation-based features preselection: Initially, the absolute correlations between features and the target label were computed. The 20 features with the strongest correlation to the target variable were selected. To reduce multicollinearity, features with pairwise correlation coefficients greater than 0.85 were iteratively removed.

Recursive Feature Elimination (RFE) [19]: A logistic regression model with L2 regularization served as the base estimator for RFE, selecting the top five most informative predictors after standardization. This step has been shown to improve generalization in small-sample EEG datasets [20]. The global parameters for each step of feature selection, including the number of top correlated features and the number of features retained by RFE, were chosen based on preliminary experiments and optimized using a grid search approach [21]. Finally, the selected features are shown in Table 2.

TABLE 2 Selected Features for Burnout Classification

N.	Feature Name
1	T4_beta
2	F4_beta
3	T6_alpha1
4	F4_beta2
5	T5_delta

Class rebalancing: To mitigate class imbalance, a hybrid resampling strategy combining Synthetic Minority Oversampling Technique (SMOTE) with Tomek link removal was applied [19]. This method simultaneously generates synthetic minority class samples and removes borderline instances, producing cleaner decision boundaries for classifiers.

Classifier selection and parameter choices: Three classifiers were evaluated in the pipeline due to their complementary strengths and suitability for small datasets [20, 21]:

Logistic Regression (LogReg): A linear baseline model with interpretable coefficients. In the pipeline, the regularization strength parameter C was tuned over the range $[0.01, 0.1, 1, 10]$. Smaller C values impose stronger regularization to prevent overfitting, while larger values allow the model to better fit the data. This range was chosen to explore a balance between bias and variance given the small dataset size.

Random Forest (RF): An ensemble of decision trees capable of capturing nonlinear feature interactions. Hyperparameters tuned included number of estimators (number of trees) from range $[50, 100, 200]$ and maximum tree depth from range $[3, 5, \text{None}]$. A lower depth prevents overfitting, which is important with only 80 samples, while tuning the number of trees balances predictive stability and computational cost.

Gradient Boosting (GB): A sequential ensemble method that improves predictive performance iteratively. Hyperparameters tuned were number of estimators from range $[50, 100]$ and learning rate from range$[0.01, 0.1, 0.2]$. A smaller learning rate slows down learning to improve generalization, while the number of estimators controls model complexity. The chosen ranges reflect a compromise between accuracy and overfitting risk on a small dataset.

Model evaluation: To ensure unbiased and robust performance assessment, nested Cross-Validation (CV) was employed. The outer loop used 5-fold repeated stratified CV (5×5) to split the dataset into training and test sets, maintaining the original class distribution in each fold. Within each training fold, the inner loop performed hyperparameter optimization via grid search for each classifier, preventing information leakage from the test data. Model performance was quantified using macro-averaged F1, recall, and precision, metrics suitable for imbalanced medical datasets, and predictions from all outer folds were aggregated to provide overall evaluation. The choice of 5 folds in the outer cross-validation loop was motivated by the small size of the dataset (80 samples). Five folds represent a practical balance between obtaining robust, low-variance performance metrics and retaining a sufficient number of samples in each test fold for meaningful evaluation.

3.3 Deep Learning Approach and Network Architecture

Building upon the limitations of traditional ML, a DL framework was developed to exploit the nonlinear and hierarchical relationships within the EEG feature space. The DL pipeline incorporated the following key components:

- Feature preprocessing: As with ML, the input features were standardized and subjected to RFE to reduce dimensionality to the five most informative predictors shown in Table 2. This ensured comparability between the two approaches.

  Data augmentation: SMOTE oversampling was applied at the training stage of each fold to address class imbalance. Unlike in ML, Tomek link removal was omitted, as the DL models demonstrated resilience to borderline instances.

- Network architecture: The final architecture was a feedforward neural network comprising:

  1) Input layer matching the dimensionality of the RFE-selected features;
  2) Three hidden layers with 64, 32, and 16 neurons, respectively, each using ReLU activations and L2 kernel regularization to prevent overfitting;
  3) Dropout layers (p=0.2) interleaved between hidden layers for additional regularization;
  4) A single sigmoid output unit for binary classification.

- Loss function and model compilation: The neural network was compiled using the Adam optimizer with a learning rate of 5×10^{-4}, binary focal loss as the objective function, and multiple evaluation metrics – accuracy, precision, and recall. The Adam optimizer [22] adaptively adjusts learning rates for each parameter, allowing stable and efficient convergence.

  Binary focal loss [23] emphasizes learning from misclassified samples, which is particularly useful in imbalanced datasets, with parameters set to $\alpha = 0.5$ and $\gamma = 2$ to balance classes and focus on harder-to-classify instances. For a single instance with true label $y \in \{0, 1\}$ and predicted probability $\hat{y} \in [0, 1]$, the binary focal loss is defined as:

  $$\mathcal{L}_{FL}(y, \hat{y}) = -\alpha\,(1 - p_t)^{\gamma}\,\log(p_t), \tag{1}$$

  where

  $$p_t = \begin{cases} \hat{y}, & \text{if } y = 1, \\ 1 - \hat{y}, & \text{if } y = 0, \end{cases} \tag{2}$$

  $\alpha \in [0, 1]$ is a weighting factor for balancing positive and negative classes, and $\gamma \geq 0$ is the focusing parameter that reduces the contribution of well-classified examples. This approach has been successful in other biomedical classification contexts where minority class detection is critical [23–26].

- Model training and decision threshold optimization:

  To ensure a reliable and unbiased evaluation of the proposed DL model, we employed a comprehensive training and validation pipeline. First, a 5-fold stratified cross-validation scheme was used. Stratification preserved the proportion of burnout and non-burnout cases in each fold, thereby ensuring that both classes were adequately represented in training and validation subsets. This is particularly important for relatively small and moderately imbalanced datasets, where conventional cross-validation may produce folds that are unrepresentative of the population distribution. Within each fold, the training set was augmented using SMOTE, which generates synthetic minority class examples. This step mitigated the bias toward the majority class and allowed the model to better learn class-specific decision boundaries without leaking synthetic samples into the validation partitions.

On top of this structure, we implemented an ensemble learning strategy [27]. For each fold, three identical neural networks were trained independently, each starting from different random initializations. Their outputs were then averaged to form the final ensemble prediction. Such an approach reduces prediction variance and increases robustness, as individual models may converge to different local minimum. Ensemble averaging thus provides a more stable decision process, especially in small datasets where single networks may otherwise be sensitive to noise or initialization effects.

Instead of applying a fixed classification threshold of 0.5, we performed a dynamic threshold search in the range of 0.3–0.7 for each fold. The threshold that maximized the macro F1-score on the validation set was selected as optimal. This adaptive calibration improved the trade-off between sensitivity and specificity, ensuring balanced performance across both classes and addressing the skew introduced by class imbalance.

- Performance evaluation: Model performance was evaluated comprehensively using accuracy, precision, recall, and F1-score, with macro F1 chosen as the primary metric due to its ability to equally weight both classes. Confusion matrices were further analyzed to visualize classification patterns and misclassifications. The ensemble achieved an overall accuracy of 86.3% with class-specific F1-scores of 0.800 for the non-burnout group and 0.895 for the burnout group, outperforming conventional ML baselines.

TABLE 3 Nested CV Classification Report

Logistic Regression (LogReg)		
Metric	Class 0	Class 1
Precision	0.491	0.758
Recall	0.607	0.662
F1-score	0.543	0.706
Support	140	260
Accuracy	0.642	
Macro avg	0.625	0.634
Weighted avg	0.664	0.642
Random Forest (RF)		
Metric	Class 0	Class 1
Precision	0.414	0.697
Recall	0.500	0.619
F1-score	0.453	0.656
Support	140	260
Accuracy	0.578	
Macro avg	0.556	0.560
Weighted avg	0.598	0.578
Gradient Boosting (GB)		
Metric	Class 0	Class 1
Precision	0.398	0.686
Recall	0.486	0.604
F1-score	0.437	0.642
Support	140	260
Accuracy	0.562	
Macro avg	0.542	0.545
Weighted avg	0.585	0.562

TABLE 4 ML Nested CV Results (averaged across folds)

Model	Test F1 Macro	Test Recall	Test Precision	Mean Score
LogReg	0.625	0.662	0.758	0.681
RF	0.554	0.619	0.697	0.624
GB	0.540	0.604	0.686	0.610

4 Results

Nested CV classification reports for LogReg, RF and GB are shown in Table 3 and Table 4. The evaluation of both ML and DL models demonstrates clear differences in predictive performance for the burnout classification task. Among the ML models, LogReg consistently outperformed ensemble methods such as RF and GB, achieving the highest accuracy (0.642) and macro F1-score (0.625). RF and GB, while capable of identifying the majority class, struggled with the minority class, reflected in lower precision and recall values, indicating a tendency toward class imbalance bias. These results suggest that simpler linear models, when paired with careful feature selection, can be more robust in small, imbalanced datasets than more complex tree-based methods.

In contrast, the DL ensemble substantially outperformed all traditional ML models. As shown in Table 5 the model achieved an overall accuracy of 0.863, with F1-scores of 0.800 for class 0 and 0.895 for class 1, indicating high sensitivity and precision across both classes. This balanced performance reflects the ensemble's capacity to capture complex nonlinear relationships that conventional ML models cannot, while the combination of SMOTE oversampling, dropout, and L2 regularization effectively mitigated overfitting despite the limited dataset size.

The confusion matrices for the top-performing ML classifier (LogReg, Fig. 3) and the proposed DL model (Fig. 4) reinforce these findings. Misclassifications in the ML model were skewed toward the minority class, whereas the DL ensemble produced markedly fewer errors across both classes, demonstrating more balanced classification. The training and validation loss curves (Fig. 5) further support this conclusion, showing smooth convergence and a minimal gap between training (0.04) and validation (0.06) loss, indicating that the model generalizes well to unseen data and provides robust performance for objective burnout detection.

TABLE 5 DL Classification Report (5-fold CV with Ensemble & Threshold)

Per-class metrics		
Metric	Class 0	Class 1
Precision	0.815	0.887
Recall	0.786	0.904
F1-score	0.800	0.895
Support	28	52
Overall metrics		
Accuracy	0.863	
Macro avg	0.851	0.845
Weighted avg	0.862	0.863

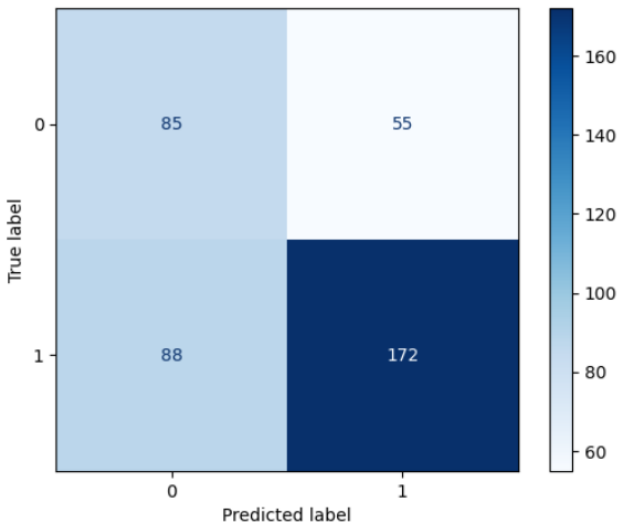

Fig. 3. ML (LogReg) Confusion Matrix

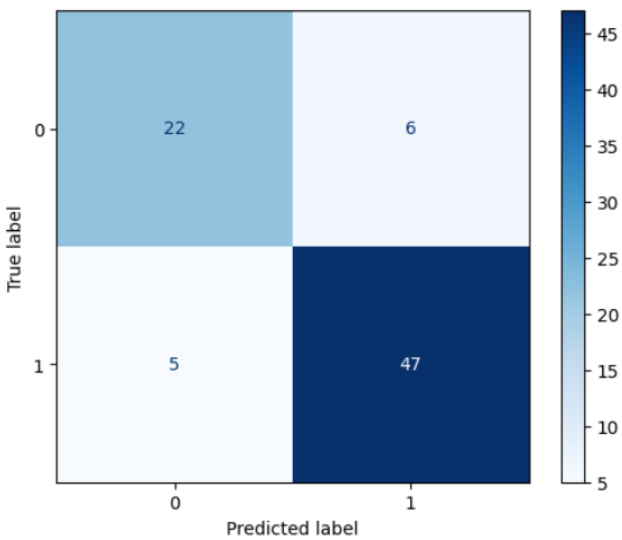

Fig. 4. DL Confusion Matrix

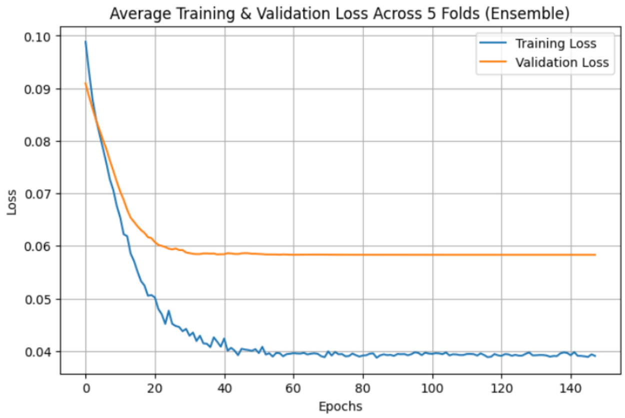

Fig. 5. Train and Validation Loss

Conclusions

The present study demonstrates that both ML and DL approaches can effectively discriminate between individuals with and without emotional burnout using combined psychometric and EEG-derived features. The most informative power spectral density changes were observed in the beta band (temporal T4, frontal F4), the beta2 sub-band (frontal F4), the alpha1 sub-band (temporal T6), and the delta band (temporal T5). Among ML models, Logistic Regression provided the best performance, achieving moderate accuracy and F1-scores, while ensemble tree-based methods (Random Forest and Gradient Boosting) were less effective, particularly in identifying the minority class. This suggests that, for small and imbalanced datasets, simpler linear models with careful feature selection can yield reasonable baseline performance.

DL, on the other hand, offered a substantial improvement across all metrics. The ensemble of neural networks with focal loss and threshold optimization achieved an overall accuracy of 86.3% and high F1-scores for both burnout and non-burnout classes, indicating balanced sensitivity and precision. The minimal gap between training and validation losses further confirms the model's robustness and generalizability, despite the limited sample size. These results underscore the advantage of DL in capturing complex, nonlinear relationships within high-dimensional EEG data, which traditional ML models may fail to exploit.

Overall, the findings support the potential of computational approaches for objective burnout assessment. While ML models can provide interpretable baseline predictions, DL frameworks offer superior performance and robustness, suggesting their suitability for clinical applications where accurate and reliable detection is critical. Future research should aim to expand dataset size, incorporate longitudinal measurements, and explore multimodal inputs to further enhance diagnostic accuracy and generalizability.

References

[1] Van Luijtelaar G. V., Verbraak M., van den Bunt M., Keijsers G., and Arns M. (2010). EEG findings in burnout patients. *The Journal of Neuropsychiatry and Clinical Neurosciences*, Vol. 22, No. 2, pp. 208–217. DOI: 10.1176/jnp.2010.22.2.208.

[2] Golonka K., Gawlowska M., Mojsa-Kaja J., and Marek T. (2019). Psychophysiological characteristics of burnout syndrome: Resting-state EEG analysis. *BioMed Research International*, 2019, Article ID 3764354. DOI: 10.1155/2019/3764354.

[3] Tement S., Pahor A., and Jaušovec N. (2016). EEG alpha frequency correlates of burnout and depression: The role of gender. *Biological Psychology*, Vol. 114, pp. 1–12. DOI: 10.1016/j.biopsycho.2015.11.005.

[4] Guo Y., Park J., and Kang S. (2024). Effects of job crafting and leisure crafting on nurses' burnout: A machine learning approach. *International Journal of Environmental Research and Public Health*, Vol. 21, No. 7, Article 1191915. DOI: 10.1155/2024/9428519.

[5] Van Zyl-Cillié L., Le Roux D. B., and Engelbrecht A. (2024). Machine learning model to predict burnout risk factors among South African nurses. *Healthcare*, Vol. 12, No. 16, Article 1662. DOI: 10.1186/s12913-024-12184-5.

[6] Badr Y. (2024). A review on evaluating mental stress by deep learning using EEG. *Neural Computing and Applications*, Vol. 36, No. 6, pp. 2849–2871. DOI: 10.1007/s00521-024-09809-5.

[7] Kim H.-G., Jeong D.-K., and Kim J.-Y. (2022). Emotional stress recognition using EEG signals based on a three-dimensional convolutional gated self-attention deep neural network. *Applied Sciences*, Vol. 12, No. 21, Article 11162. DOI: 10.3390/app122111162.

[8] Bianchi R., Schonfeld I. S., and Laurent E. (2017). Biological research on burnout-depression overlap: Long-standing limitations and on-going reflections. *Neuroscience and Biobehavioral Reviews*, Vol. 83, pp. 238–239. DOI: 10.1016/j.neubiorev.2017.10.019.

[9] Afek N., Harmatiuk D., Gawlowska M., Ferreira J. M. A., Golonka K., Tukaiev S., Popov A., and Marek T. (2025). Functional connectivity in burnout syndrome: a resting-state EEG study. *Frontiers in Human Neuroscience*, Vol. 19, Article 1481760. DOI: 10.3389/fnhum.2025.1481760.

[10] Kakkos, I., Tzavellas, E., Feleskoura, E., Mourtakos, S., Kontopodis, E., et al. (2025). EEG-Based Assessment of Cognitive Resilience via Interpretable Machine Learning Models. *AI*, 6(6), 112. DOI: 10.3390/ai6060112.

[11] Hamid, N. H. A., Sulaiman, N., Aris, S. A. M., Murat, Z. H., Taib, M. N. (2010). Evaluation of human stress using EEG power spectrum. *Proceedings of the 6th International Colloquium on Signal Processing & Its Applications (CSPA)*. DOI: 10.1109/CSPA.2010.5545282.

[12] Khan, T., Javed, H., Amin, M., Usman, O., Hussain, S. I., Mehmoood, A., and Maple, C. (2021). EEG Based Aptitude Detection System for Stress Regulation in Health Care Workers. *Scientific Programming*, Article ID 4620487. DOI: 10.1155/2021/4620487.

[13] Arif, A., Wang, Y., Yin, R., Zhang, X., Helmy, A. (2024). EF-Net: Mental State Recognition by Analyzing Multimodal EEG-fNIRS via CNN. *Sensors*, 24(6), 1889. DOI: 10.3390/s24061889.

[14] Liu Z. and Zhao J. (2025). Leveraging deep learning for robust EEG analysis in mental health monitoring. *Frontiers in Neuroinformatics*, Vol. 18, Article 1494970. DOI: 10.3389/fninf.2024.1494970.

[15] Al-Shargie F., Badr Y., Tariq U., Babiloni F., Al-Mughairbi F., and Al-Nashash H. (2023). Classification of mental stress levels using EEG connectivity and convolutional neural networks. *Proceedings of the 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, DOI: 10.1109/EMBC40787.2023.10340398.

[16] Wang, X., He, Y., Zhao, Y., Zhang, Y., and Li, Y. (2021). Multi-modal brain network analysis using graph convolutional networks for major depressive disorder classification. *Medical Image Analysis*, 72, 102123. DOI: 10.1016/j.media.2021.102123.

[17] Yeo, W.-H., Kim, Y.-S., Lee, J., and Rogers, J. A. (2012). Multifunctional Skin-Like Electronics for Long-Term Health Monitoring. *Proceedings of the ASME 2012 International Mechanical Engineering Congress and Exposition*. DOI: 10.1115/IMECE2012-85698.

[18] Mushta S. A., Palamar S. P., Harmatiuk D. V., Popov A. O., Alves Ferreira J. M., Makarchuk M. Yu., and Tukaiev S. V. (2025). EEG frequencies pattern as brain signature of emotional burnout: Anxiety-tension stage formation. *Clinical and Preventive Medicine*, No. 1, pp. 92–100. DOI: 10.31612/2616-4868.1.2025.12.

[19] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. DOI: 10.1613/jair.953.

[20] Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, 180, 68–77. DOI: 10.1016/j.neuroimage.2017.06.061.

[21] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, pp. 2825–2830.

[22] Kingma D. P., Ba J. (2015). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.

[23] Lin T.-Y., Goyal P., Girshick R., He K., and Dollár P. *Focal Loss for Dense Object Detection*, IEEE International Conference on Computer Vision (ICCV), 2017.

[24] Zunair H. and Ben Hamza A. (2020). Melanoma Detection using Adversarial Training and Deep Transfer Learning. *arXiv preprint arXiv:2004.06824*.

[25] Yeung M., Sala E., Schönlieb C.-B., and Rundo L. (2021). Unified Focal Loss: Generalising Dice and Cross Entropy-Based Losses to Handle Class Imbalanced Medical Image Segmentation. *arXiv preprint arXiv:2102.04525*.

[26] Pandiyaraju V., Malarvannan S., Venkatraman S., Abeshek A., Priyadarshini B., and Kannan A. (2024). A Novel Adaptive Hybrid Focal-Entropy Loss for Enhancing Diabetic Retinopathy Detection Using Convolutional Neural Networks. *arXiv preprint arXiv:2411.10843*.

[27] Ganaie M. A., Hu M., Malik A. K., Tanveer M., and Suganthan P. N. (2021). Ensemble deep learning: A review. *arXiv preprint arXiv:2104.02395*, pp. 1–35.

Діагностика емоційного вигорання із застосуванням методів машинного та глибокого навчання

Мушта С. А., Мушта І. А., Попов А. О., Лисенко О. М., Тукаєв С. В.

У роботі розглянуто розробку та оцінку обчислювальних методів для класифікації емоційного вигорання на основі кількісного аналізу даних. На відміну від суб'єктивних психометричних опитувальників, досліджується застосування алгоритмів машинного та глибокого навчання до структурованих наборів даних. Було проведено систематичне порівняння класичних методів, зокрема логістичної регресії, випадкового лісу та градієнтного бустингу, з ансамблевою моделлю глибокого навчання. Для підвищення надійності застосовано попередню обробку даних, включаючи відбір ознак, балансування та ресемплінг. Архітектура глибокого навчання (Deep Learning, DL), що поєднує функцію втрат типу focal loss та адаптивну оптимізацію порогу класифікації, продемонструвала найкращі результати. За 5-кратної крос-валідації запропонована DL-модель досягла загальної точності 86.3%, з показниками precision 0.815 / 0.887, recall 0.786 / 0.904 та F1-мірою 0.800 / 0.895 для негативного та позитивного класів відповідно. Отримані результати підтверджують, що сучасні обчислювальні моделі можуть забезпечувати масштабовані та узагальнені інструменти для задач автоматизованого виявлення, створюючи технічне підґрунтя для подальшої інтеграції у прикладні дослідження та системи моніторингу професійного здоров'я.

Ключові слова: емоційне вигорання; машинне навчання; глибоке навчання; електроенцефалографія; опитувальник вигорання Маслач; відбір ознак; класифікація; діагностика психічного здоров'я