ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 13, Issue 05, May 2026

The Digital Empathy Engine: An AI-Driven Framework for Automated Burnout Detection and Intervention in HR Analytics

[1] Seelam Dikshitha, [2] Harshini PJ, [3] Deeksha Gottapu, [4] Chamakura Saisree, [5] Lakshman Raj Lokesh, [6] Vinodhkumar S

[1] [2] [3] [4] [5] [6] School of Business, Woxsen University, Hyderabad, India
Corresponding Author's Email: [1] seelam.dikshitha_2026@woxsen.edu.in, [2] harshini.pj_2026@woxsen.edu.in, [3] deeksha.gottapu_2026@woxsen.edu.in, [4] saisree.chamakura_2026@woxsen.edu.in, [5] lakshman.chapana_2026@woxsen.edu.in, [6] vinodhkumar.s_2026@woxsen.edu.in

Abstract*— Burnout amongst employees has been recognized as an emerging challenge for organizations in remote/hybrid work settings, but current solutions to this problem are limited by their dependence on self-reported measures, making it difficult to capture early stage burnout and also by inequities inherent in the use of these self-report based approaches in culturally diverse workforces. In this paper, we propose Digital Empathy Engine, an artificial intelligence framework which uses passive analysis of digitally created information, culturally normalized risk scoring, and personalized Just-In-Time Adaptive Interventions (JITAIs) to mitigate the challenges presented by the issues. Digital Empathy Engine analyzes employee metadata extracted from regular workplace communications such as after-hours email volume, meeting schedule, lag time, network density, and weekend work behavior, normalizing these inputs with respect to country-specific baselines derived using datasets collected from six different national contexts namely, India, USA, Germany, UK, China, and Japan. Four supervised machine learning classifiers such as Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting were tested and validated against 5,000 daily records of employees. All four classifiers yielded ROC-AUC scores above 0.95, with the latter achieving an F1-score for the burnout-class prediction task of 0.9355. An extensive ablation analysis was conducted which proves cultural normalization's superior performance compared to raw feature representation in terms of increasing burnout-class F1 score by 7.8 percentage points, thereby confirming our primary novelty proposition. The JITAIs engine component generates culturally sensitive recommendations for employees, supervisors, and HR personnel.

Index Terms: burnout detection, cultural normalisation, digital exhaust, gradient boosting, HR analytics, just-in-time adaptive intervention, machine learning, remote work.

I. INTRODUCTION

Employee burnout has become a central problem of modern human resource management. Employee burnout was described by Maslach and Jackson [1] as feeling emotionally exhausted, depersonalisation and lack of personal achievement. Remote and hybrid work has blurred the lines between professional and personal life even more, and recent evidence from Gallup [2] suggests that many employees suffer from burnout at some point. The World Health Organisation's recognition of burnout as an occupational phenomenon in ICD-11 [3] has made it more pressing for organisations to move from reactive wellness initiatives to proactive, system-level detection of early risk signals.

Uncontrolled burnout results in high organisational costs. The American Institute of Stress claims that absenteeism and turnover losses attributed to burnout cost the US economy hundreds of billions of dollars yearly. Burnout-related absenteeism and turnover losses are particularly high in intellectually-intensive industries like technology, consulting, and finance, which depend heavily on their intellectual assets. During the initial stages of burnout, employees perform well, but they are mentally disconnecting from their work. Traditional methods of detection, such as annual performance reviews, pulse surveys, and exit interviews, are insufficiently effective for identifying burnout.

There are three main drawbacks to traditional approaches to measuring burnout. First, they employ intermittent information gathering, which fails to identify the gradual changes in behavior that take place daily. Second, they depend on self-reporting, which means that stressed employees will under-report their burnout. Finally, they ignore the cultural context. Behavior considered normal in one nation may indicate boundary crossing in another country, leading to distorted assessments when using standardised assessment tools in multinational organisations without correcting for differences in culture. We introduce the Digital Empathy Engine (DEE), an AI-powered framework that addresses these shortcomings with three interrelated contributions. (1) Passive digital exhaust such as

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 05, May 2026

email timestamps, calendar usage and collaboration activity, can be analysed on an ongoing basis with no additional reporting burden [6]. (2) Risk is not considered in terms of absolute levels of activity, but rather as deviation from culturally appropriate patterns, based on cross-national research on workplace norms [5], and the cultural normalisation of behavioral indicators. (3) The ML risk scores generated by current models [8], [9], [11] are transformed into privacy-preserving, role-based recommendations for workers, supervisors, and HR personnel using the JITAI approach [10].

The remainder of the paper is devoted to discussing pertinent literature, presenting the methodology, experimental results, implications and limitations, and concluding remarks for future work.

II. LITERATURE REVIEW

A. Employee Burnout: Theory and Organisational Impact

Freudenberger [4] coined the definition of burnout in 1974 as being in a state of mental and physical exhaustion due to an intense exertion of efforts at one's workplace. Further development was carried out by Maslach and Jackson [1], who designed the Maslach Burnout Inventory, which defined three distinct dimensions of the state of burnout: emotional exhaustion (being exhausted from work-related emotional stress); depersonalisation (adopting a detached and/or cynical attitude towards other people, both colleagues and clients); and having lower feelings of personal accomplishment (i.e. lack of self-effectiveness).

The Maslach Burnout Inventory is currently considered the gold standard when diagnosing burnout and has served as the basis for designing the features used in the DEE system. As far as organisation is concerned, burnout is associated with high voluntary turnover intentions; reduced job productivity; increased absence and sick leaves; higher healthcare utilization; and long-term health issues such as major depression disorder and cardiovascular diseases. The current inclusion of the burnout concept into the ICD-11 classification has placed additional responsibility on organisations to solve systemic issues leading to burnout excessive workload, unclear role definition, low autonomy, and lack of social support.

B. Limitations of Traditional HR Analytics

Conventional HR analytics aggregate lagging indicators of disengagement rather than leading predictors of burnout. Pulse surveys, while more frequent than annual reviews, suffer from systematic response bias and survey fatigue: the employees most urgently in need of support are least likely to self-identify through self-report mechanisms. Exit interview data is definitionally retrospective, capturing insights only after the organisation has already lost the employee. Performance management data measures the consequences of burnout reduced output quality, missed deadlines rather than its early-stage behavioural precursors detectable weeks or months earlier.

A fundamental limitation underappreciated in the HR analytics literature is the failure to account for cultural heterogeneity in risk assessment models. Hofstede [5] demonstrates through extensive cross-national research that attitudes toward work-life balance, willingness to report workplace difficulties upward, and interpretations of appropriate work intensity vary substantially across national cultures along dimensions including power distance, uncertainty avoidance, and individualism-collectivism. A standardised monitoring instrument deployed across a multinational workforce without cultural calibration produces systematically biased risk assessments, over-flagging employees from high-baseline cultures as at-risk based on normative behaviour and under-flagging those from low-baseline cultures whose moderate absolute levels represent severe relative deviations.

C. Digital Exhaust as a Continuous Behavioural Signal

Digital exhaust refers to the passively generated metadata produced as a by-product of routine workplace communication activity email timestamps, calendar records, reply to latencies, and collaboration network edge counts. Unlike survey-based data, digital exhaust is continuous, objective, verifiable, and does not depend on employee self-disclosure. Prior research has established the feasibility of using email metadata to infer cognitive workload and stress states [6], calendar fragmentation patterns to measure attentional switching costs, and communication network topology to detect early social isolation signals. These passive markers are very similar to the behavioral manifestations associated with each of the MBI dimensions: out-of-work hours emailing and meeting overloading mirror the boundary blurring and role overload that cause emotional exhaustion; decreasing density of collaborative networks symbolizes the social isolation that is characteristic of depersonalization; and high response latency measures the cognitive depletion that accompanies low personal accomplishment.

D. Machine Learning in HR Analytics

Machine learning application in HR analytics has expanded substantially over the past decade. Bhatt et al. [7] demonstrated the feasibility of predicting employee attrition using supervised classifiers on standard HRIS data. Further research has considered topics such as the natural language processing of emails, temporal modeling of workload, analyzing collaboration graphs, and the sentiment analysis of performance review texts as proxies for engagement and stress, respectively. Among the supervised algorithms, ensemble-based techniques like Random Forest [8] and Gradient Boosting [9], which have outperformed other methods in their ability to capture non-linear relationships

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 05, May 2026

between features in the context of imbalanced classes, such as in cases of burnout prediction, have proved superior.

A persistent limitation of existing ML approaches to burnout detection is the absence of cultural calibration. Models trained on multinational data without normalisation conflate the burnout signal with cross-cultural variation in baseline communication behaviour, degrading both predictive accuracy and fairness across national employee populations. The DEE addresses this gap directly through its Stage 3 cultural normalisation pipeline.

E. Just-In-Time Adaptive Interventions

JITAIs were initially developed in mobile health to deliver behavioural support at moments of peak individual vulnerability [10]. Unlike static programmes applying uniform support regardless of individual state, JITAIs continuously assess contextual signals and deliver tailored support precisely when it is most likely to produce the desired behavioural response. Nahum-Shani et al. [10] define three core JITAI components: the decision point (when to intervene), the tailoring variable (what contextual information determines the intervention), and the intervention option (what support to deliver). In the DEE's HR analytics context, the decision point is the risk band threshold, the tailoring variable is the set of five active driver flags identifying the specific behavioural patterns driving elevated risk, and the intervention options are role-differentiated recommendations calibrated to the employee, manager, and HR stakeholder tiers. Adapting JITAI principles to organisational wellbeing at scale represents an unexplored frontier that the DEE directly addresses.

F. Research Gap

A review of the literature confirms that no existing HR analytics framework simultaneously integrates: (1) continuous passive digital exhaust monitoring requiring no additional employee reporting burden; (2) cross-cultural behavioural normalisation using country-specific baselines; and (3) a JITAI-based intervention delivery architecture with role-differentiated, privacy-preserving outputs. The DEE addresses all three dimensions within a single validated pipeline.

III. RESEARCH METHODOLOGY

A. System Architecture

The Digital Empathy Engine is structured as a five-stage modular pipeline. Each stage is independently deployable, enabling organisations to integrate individual components into existing HR information systems. Stage 1 (Digital Exhaust Collection) passively ingests communication metadata from enterprise email servers, calendar APIs, and collaboration platforms such as Microsoft Teams or Slack operating on metadata only, never reading communication content, ensuring regulatory compliance. Stage 2 (Feature Engineering) derives five behavioural metrics from the collected raw data. Stage 3 (Cultural Normalisation) divides each metric by the country-specific baseline to produce dimensionless deviation ratios. Stage 4 (ML Risk Classification) applies a trained supervised classifier to produce a 0–100 burnout risk score and Low/Medium/High band assignment. Stage 5 (JITAI Engine) translates the risk score and active driver flags into role-differentiated interventions routed to the appropriate stakeholder tier.

B. Synthetic Dataset Construction

Due to the non-availability of public longitudinal datasets on individuals containing validated labels of burnout through digital exhaust, it became necessary to create a synthetic dataset to prove the concept. The strategy employed herein is supported by existing protocols in the context of HR analytics where use of actual communication metadata by researchers has been restricted owing to stringent regulations under GDPR, the Data Protection Act of the United Kingdom, India's Digital Personal Data Protection Act (DPDP), as well as other country-specific data protection acts in China and Japan.

The dataset comprises 5,000 daily employee records generated for 250 simulated employees spanning the full calendar year 2024 (January 1 through December 31) across six countries: India, USA, Germany, UK, China, and Japan. Each record represents one employee-day observation and contains 16 variables: a unique employee identifier, country of employment, organisational role, department, a continuous seniority index (0–1), the observation date, a weekend day indicator, five raw digital exhaust features, five culturally normalised features, and the binary burnout label (0 = No Burnout, 1 = Burnout) serving as the ground-truth target. Employees were assigned one of five roles (Analyst, Manager, Engineer, Consultant, Executive) and one of six departments (Engineering, Sales, HR, Finance, Operations, Marketing). A seniority index modulates baseline communication volume, reflecting the empirical pattern that senior employees attend more meetings and interact with wider networks.

To simulate realistic burnout heterogeneity, 15% of employees were designated as burnout-prone at the simulation design stage, generating burnout-positive records on 40% of their observation days. The remaining 85% exhibited a 2% base-rate probability. This design produces a final burnout prevalence of 7.82% (391 burnout-positive; 4,609 non-burnout) epidemiologically realistic at the day-level observation granularity. Burnout-positive records received elevated stress multipliers of 1.5–2.8× the cultural baseline applied to the raw feature generation process, consistent with behavioural escalation patterns documented in the occupational health literature [1]. Non-burnout records received near-normal variation multipliers of 0.6–1.25×, ensuring the majority class is not unrealistically homogeneous. The simulation was implemented in Python

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 05, May 2026

3.11 using NumPy's default random generator with seed 42 for complete reproducibility.

C. Feature Engineering

Five behavioural features are derived from raw digital exhaust signals. These were selected based on alignment with the three MBI burnout dimensions and practical derivability from communication metadata without content access.

(1) after_hours_pct: Ratio of emails sent outside standard hours (09:00–18:00 local time) to total daily emails, clipped to [0,1]. This directly operationalises boundary erosion — the primary precursor of emotional exhaustion identified by Maslach and Jackson [1]. Burnout records show a mean normalised value of 2.11× the cultural baseline versus 1.22× for non-burnout records.

(2) meeting_load: Total calendar meeting duration per day in minutes. Meeting overload is a key antecedent of the role overload component of depersonalisation. Seniority modulates this feature, as senior employees attend more meetings on average. Burnout records show 1.65× versus 1.01× for non-burnout records.

(3) response_delay: Mean email reply latency in minutes. This serves as a proxy for cognitive overload and attentional fragmentation: employees overwhelmed by competing demands respond more slowly as attentional resources are depleted. Interpretation requires cultural normalisation — Germany's 45-minute baseline reply time versus China's 20-minute baseline reflects structural cultural norms, not individual distress.

(4) collab_index: Count of unique individuals communicated with per day. Unlike the other four features, collab_index is expected to be inversely associated with burnout risk: declining social network activity reflects the withdrawal characteristic of depersonalisation. This protective-direction feature is confirmed by LR coefficient analysis (−2.07) and operationalised in the JITAI engine as the Low Collaboration driver flag.

(5) weekend_load: Total documented work activity on Saturday and Sunday in minutes. Weekend work is the strongest single discriminator in this dataset (burnout-to-non-burnout ratio: 3.95×), consistent with the recovery failure literature identifying psychological detachment from work during off-hours as a key burnout moderator in remote and hybrid work environments [2].

D. Cultural Normalisation

Cultural normalisation is the primary methodological novelty of the DEE. For each of the five features, the observed employee-day value is divided by the corresponding country-specific baseline from the cultural registry, as formalised in Equation (1):

$$\text{norm}(i,c,t) = \text{raw_feature}(i,t) / (\text{baseline}(c) + \varepsilon) \quad \ldots (1)$$

where raw_feature(i,t) is the observed value for employee i on day t, baseline(c) is the country-specific normative value for country c, and $\varepsilon = 1\times10^{-6}$ prevents division by zero. The resulting ratio is dimensionless and culture-invariant: 1.0 indicates behaviour exactly at the cultural norm; 2.0 indicates twice the norm. The same ratio means the same relative risk burden regardless of which country the employee is in.

Table I presents the six-country baseline specification. Values derive from Hofstede's cultural dimensions framework [5], OECD Better Life Index working hours data, and Eurofound reports on digital workplace norms. Japan's high after-hours (30%) and weekend activity (50 min) baselines reflect deeply embedded corporate commitment norms; Germany's substantially lower values (8%, 8 min) reflect the strong legal and cultural protections for work-life separation in German labour relations. Without this normalisation, a Japanese employee and a German employee exhibiting identical absolute after-hours volumes would receive identical risk contributions despite representing radically different relative deviations from their respective cultural contexts.

Table I. Country-Level Cultural Baseline Values

Country	AH%	Mtg	Reply	Collab	Wknd
USA	18%	120	28	9	22
India	22%	95	35	8	35
Germany	8%	75	45	6	8
UK	14%	100	32	8	15
China	28%	110	20	10	45
Japan	30%	130	22	7	50

AH%=After-Hours email %; Mtg=Meeting minutes/day; Reply=Avg reply latency (min); Collab=Unique collaborators/day; Wknd=Weekend work minutes/day.

Two robustness mechanisms are embedded in the normalisation layer. First, divide-by-zero protection via ε handles any zero-valued baseline entry. Second, employees from countries absent from the registry are assigned global median baselines computed across all six registered countries, ensuring the system degrades gracefully when geographic metadata is incomplete.

E. Class Imbalance Handling

The dataset's 92.18%/7.82% class split presents a severe imbalance characteristic of real-world burnout monitoring data. Left unaddressed, this causes all classifiers to maximise accuracy by predicting the majority class universally achieving ~92% accuracy while detecting zero burnout cases. All four classifiers were configured with class_weight='balanced', which adjusts each class's loss contribution in inverse proportion to its frequency. Concretely, burnout-class errors are weighted approximately 11.8× more heavily than non-burnout errors during training, mathematically equivalent to oversampling the minority class without introducing the synthetic data artefacts of SMOTE. SMOTE was specifically excluded because synthetic

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 05, May 2026

interpolation between structured digital exhaust records risks producing behaviourally implausible instances, and because SMOTE-inflated training sets can produce optimistic cross-validation estimates when synthetic records appear in validation folds.

F. Machine Learning Model Configuration

Four supervised classification algorithms were selected to span a deliberate spectrum from interpretable linear models to high-performance ensemble methods. This enables simultaneous evaluation of predictive accuracy, stability, and explainability all critical for HR analytics deployment contexts where outputs must be trusted by and communicated to non-technical practitioners. All models used scikit-learn 1.4 [11], Python 3.11, random seed 42, and the five culturally normalised features as inputs.

Logistic Regression (LR): Pipeline of StandardScaler followed by LogisticRegression (class_weight='balanced', max_iter=500). StandardScaler prevents data leakage during cross-validation. LR's standardised coefficients provide direct log-odds-unit feature importance, making it the primary explainability companion in the recommended dual-model production architecture.

Decision Tree (DT): DecisionTreeClassifier (max_depth=6,min_samples_leaf=10,class_weight='balanced'). Maximum depth of six ensures any prediction path is auditable by a non-technical HR practitioner in at most six decision steps. The minimum leaf size prevents minority-class overfitting.

Random Forest (RF): RandomForestClassifier (n_estimators=300, class_weight='balanced', n_jobs=-1). Bootstrap aggregation across 300 diverse trees reduces variance and generates stable Gini-based feature importance scores free of the linear assumptions constraining LR coefficients. RF is the high-precision candidate for deployment contexts where false-positive minimisation is the primary operational concern.

Gradient Boosting (GB): GradientBoostingClassifier (n_estimators=200, learning_rate=0.05, max_depth=4). The conservative learning rate reduces overfitting; 200 trees provide capacity to capture non-linear feature interactions. GB's sequential residual correction mechanism [9] assigns asymmetric loss weight to misclassified minority-class instances, making it particularly well-suited to imbalanced burnout detection. GB is the recommended primary production scorer based on its symmetric precision-recall profile demonstrated in the results.

G. Evaluation Strategy

All models were evaluated on a 75/25 stratified train-test split (3,750 training / 1,250 test records), preserving the 7.82% burnout prevalence in both partitions. Stratification is critical for imbalanced datasets: random splitting risks placing insufficient minority-class records in the test partition, rendering precision-recall estimates statistically unreliable. To assess generalisation stability beyond a single split, all models were additionally evaluated using 5-fold stratified cross-validation on all 5,000 records, with CV AUC mean and standard deviation reported alongside test-set metrics.

Primary metric: ROC-AUC, measuring the probability that a randomly selected burnout record receives a higher score than a randomly selected non-burnout record across all thresholds. ROC-AUC is threshold-independent and robust to class imbalance. Secondary metrics include burnout-class precision, recall, and F1-score. Accuracy is reported for completeness but de-emphasised: a degenerate majority-class predictor achieves 92.18% accuracy with zero burnout identification capability, making accuracy misleading for this task.

H. JITAI Engine Design

The JITAI engine translates continuous risk scores into structured interventions through two sequential operations. Risk band classification converts the calibrated Logistic Regression probability (selected for its well-calibrated probability estimates) to a 0–100 scale and applies fixed thresholds: Low (<30) no active intervention; Medium (30–59) employee nudges only; High (≥60) full three-tier cascade. Five binary driver flags identify specific behavioural root causes: (1) Meeting Overload: norm_meetings ≥ 1.3; (2) After-Hours Work: norm_after_hours ≥ 1.3; (3) Weekend Work: weekend_load ≥ 60 min; (4) Response Delay: norm_reply ≥ 1.4; (5) Low Collaboration: norm_collab ≤ 0.5.

The three-tier delivery architecture enforces strict privacy boundaries. Employees receive direct personalised nudges for their specific flagged patterns, framed as supportive suggestions. Managers receive abstracted team-level guidance with no individual identification or score disclosure. HR receives population-level trend signals and is escalated to only when High Risk persists for three or more consecutive days, using aggregate summaries only. Raw individual scores are never transmitted outside the employee tier without explicit opt-in consent a functional requirement for data integrity as much as an ethical one, since systems perceived as surveillance produce strategic impression management that corrupts the underlying signal.

I. Implementation Details

The complete DEE implementation comprises five Python scripts: (1) 01_generate_dataset.py synthetic data generation with country-stratified parametric simulation (NumPy 1.26, Pandas 2.2); (2) 02_ml_pipeline.py feature engineering, cultural normalisation, model training, 5-fold CV, and metric computation (scikit-learn 1.4); (3) 03_ablation_study.py four-condition ablation experiment with cross-cultural generalisation testing; (4) 04_jitai_engine.py risk scoring, driver flag computation, three-tier intervention generation and routing; (5) 05_visualisations.py publication-quality figure generation (Matplotlib 3.8, Seaborn 0.13). All trained

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 05, May 2026

model artefacts are serialised using joblib for reproducible inference. The codebase, dataset, and baseline configurations are provided in a self-contained reproducibility package with a master run_all.py script that executes the full pipeline end-to-end from a single command.

IV. RESULTS

A. Feature Discriminative Power

Comparing mean normalised feature values between burnout-labelled and non-burnout-labelled records reveals strong and consistent class separability. Burnout records exhibit normalised meeting loads of 1.65× the cultural baseline versus 1.01× for non-burnout records (ratio: 1.63×); normalised after-hours email proportion of 2.11× versus 1.22× (ratio: 1.73×); and normalised weekend load of 3.63× versus 0.92× (ratio: 3.95×). Weekend work provides the strongest single discriminator. The consistently above-baseline normalised values even in the non-burnout class confirm that all employees exhibit some after-hours activity; the burnout distinction lies in the cultural-context-relative severity of that deviation. Logistic Regression standardised coefficients quantify independent feature contributions: norm_meetings (+3.66), norm_after_hours (+1.59), norm_weekend (+1.52) are the dominant risk-increasing features; norm_collab (−2.07) confirms the protective association of social connectivity with burnout resistance, consistent with social support buffer theory [1].

B. Model Performance

Table II summarises the results of evaluating all four models against the 1,250 held-out testing records. Standard deviations for 5-fold cross-validation ROC-AUC values are all less than 0.009, indicating actual stability of generalisation.

Table II. Model Performance on Held-Out Test Set (n = 1,250)

Model	AUC	CV AUC	Prec.	Rec.	F1-B
Log. Reg.	0.9949	0.9986±0.0011	0.887	0.959	0.9216
Dec. Tree	0.9578	0.9817±0.0089	0.840	0.908	0.8725
Rand. Forest	0.9927	0.9977±0.0028	1.000	0.878	0.9348
Grad. Boost*	0.9911	0.9980±0.0011	0.989	0.888	0.9355

F1-B = Burnout-class F1-Score. *Recommended. CV AUC = 5-fold mean ± std.

All four models are significantly better than random classifiers with a ROC-AUC greater than 0.95. Gradient Boosting and Random Forest obtain 99.04% accuracy and the best burnout class F1 (0.9355 and 0.9348, respectively). Logistic Regression obtains the highest recall (0.959) while Random Forest obtains perfect precision (1.000). Gradient Boosting is preferred for deployment based on its balance between precision and recall to avoid either intervention fatigue due to false positives or missing true burnout patients.

C. Random Forest Feature Importance

The variable importance metrics calculated based on the Gini index reinforce the predominance of boundary transgressions across time. Variable norm_meetings accounts for 42.84% of model importance, followed by norm_after_hours at 29.06%, norm_weekend at 15.09%, norm_reply at 8.41%, and norm_collab at 4.60%. Combined, norm_meetings and norm_after_hours account for 71.9% of variance. This predominance comes with an obvious organisational implication, which is that changes in the structure of meetings and after-hours communication can effectively address the two most significant factors causing employee burnout without any individual behavioural adjustments.

D. Ablation Study: Causal Impact of Cultural Normalisation

Table III presents ablation results comparing Condition A (raw features) versus Condition B (normalised features), evaluated using 5-fold stratified cross-validation. All other experimental conditions model configurations, hyperparameters, folds, and seeds are held strictly constant.

Table III. Ablation Study Results — Raw vs. Normalised Features

Model	Cond.	CV AUC	F1-B	Rec.	ΔF1
LR	A — Raw	0.9910	0.7928	0.898	—
LR	B — Norm	0.9986	0.9216	0.959	+12.88pp
RF	A — Raw	0.9890	0.8634	0.806	—
RF	B — Norm	0.9977	0.9348	0.878	+7.14pp
GB	A — Raw	0.9918	0.8677	0.837	—
GB	B — Norm	0.9980	0.9355	0.888	+7.78pp

LR=Logistic Regression, RF=Random Forest, GB=Gradient Boosting. pp=percentage points. 5-fold stratified CV throughout.

Cultural normalisation consistently and materially improves all models. GB burnout-class F1 rises from 0.8677 to 0.9355 (+7.78pp); LR from 0.7928 to 0.9216 (+12.88pp); RF from 0.8634 to 0.9348 (+7.14pp). CV AUC also improves for all models. Since all other conditions are held constant, these improvements are entirely attributable to the normalisation step alone constituting causal empirical evidence for the primary novelty claim.

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 05, May 2026

E. JITAI Engine Output

When using the entire pipeline on all 5,000 data sets, we have the following results: 3,753 low risk (75.1%), 922 medium risk (18.4%) and 325 high risk (6.5%). High risk data sets align well with the previously defined burnout-positive prevalence (7.82%), thus indicating that the system retains its specificity while not burdening the HR department with too many tasks. The most frequent trigger flags include After Hours Work (41.6%), Meeting Overload (38.2%), Weekend Work (19.3%), Response Delay (14.7%), and Low Collaboration (8.7%). HR review fires high-risk data sets for those employees who demonstrate high risk on three or more consecutive days, which affects 4.1% of all employees. The sample JITAI is provided for one high-risk employee in Table IV.

Table IV. Sample JITAI Output — High Risk Employee (Score: 87.4/100)

Tier	Intervention Content
Employee	Your meeting load and after-hours activity are significantly above your usual baseline. Consider declining optional meetings this week and requesting advance agendas. Review your task list for non-urgent items that can be rescheduled to standard hours to protect your recovery time.
Manager	A team member shows elevated workload and boundary strain. Schedule a confidential check-in focused on support — not performance. Audit recurring meetings for necessity and reinforce no after-hours response expectations by modelling healthy boundaries yourself.
HR	Elevated meeting overload and after-hours work detected. If High Risk persists 3+ consecutive days, initiate a confidential HR review using aggregate summaries only. Consider a team-level meeting hygiene workshop.

V. DISCUSSION

A. Model Selection Rationale

Gradient Boosting is recommended as the primary production risk scorer based on its symmetric precision-recall profile (0.989 precision, 0.888 recall), which navigates between the two failure modes most costly in HR analytics. A high-recall, low-precision model generates excessive false-positive alerts, producing intervention fatigue analogous to alert fatigue documented in clinical decision support research. A high-precision, low-recall model misses genuine burnout cases the outcome with the highest individual and organisational cost. GB's balanced F1 of 0.9355 avoids both extremes.

Logistic Regression is retained as the explainability companion despite its lower balanced F1 because its standardised coefficients enable HR practitioners to audit and communicate the specific factors driving individual risk decisions a regulatory requirement under emerging AI governance frameworks including the EU AI Act, which classifies high-risk AI applications in employment contexts as requiring explainable decision records. The dual-model architecture (GB for scoring, LR for explanation) balances predictive performance with accountability requirements.

B. Cultural Normalisation: Evidence and Implications

The ablation study constitutes the central empirical contribution of this paper. Consistent 7–13 percentage point F1 improvements across three classifiers under strictly controlled conditions establish cultural normalisation as a substantive, causal, and replicable methodological contribution. The mechanism is analytically clear: without normalisation, a classifier must simultaneously learn both the burnout signal and the cross-cultural variation in baseline communication behaviour, conflating two independent sources of variation. Normalisation presents the model with culture-invariant deviation ratios, allowing the classifier to focus entirely on the burnout signal.

This has immediate practical implications for any multinational HR analytics deployment. Systems operating on raw communication data without cultural calibration systematically mis-classify employees: those from high-baseline cultures (Japan, China) are over-flagged based on normative behaviour, while those from low-baseline cultures (Germany) whose moderate absolute deviations represent severe relative violations are under-detected. The DEE eliminates this bias without requiring per-country models, which would be operationally impractical for organisations employing staff across many national contexts.

C. Behavioural Drivers and Organisational Interventions

The excessive number of meetings and working outside of working hours explain 71.9% of the RF features' importance, having the two biggest positive values of LR coefficients. This high weight indicates that specific structural measures such as mandatory agendas, invitations only, reserved periods of focus work time, and explicit rules against responding to messages during off-time would tackle the two highest drivers of burnout together. This solution involves no behavioural change of employees; it is based on organisational structure. The protective relationship between social interaction (LR -2.07) implies that well-being programmes should consider re-engaging peers, in addition to reducing workload.

D. Privacy and Signal Integrity

The privacy-by-design architecture is not merely an ethical constraint but a functional requirement for signal integrity. Passive monitoring systems perceived as management

surveillance reliably induce strategic impression management employees modifying their communication behaviour to appear less stressed. This corrupts the underlying digital exhaust signal, rendering the ML pipeline unreliable. With the guarantee that individual risk scores will never be disclosed to managers and HR without consent, the DEE ensures the necessary behavioural authenticity for accurate detection and provides employees with sufficient psychological safety to interact openly with the support functions provided by the system.

VI. LIMITATIONS AND FUTURE WORK

This paper suffers from four main limitations. Firstly, the generated synthetic dataset, although informed by relevant scientific literature, has not been cross-checked against empirical data from real-world organisations. Its applicability across different industries, organisational cultures, and organisational ranks is still unknown. Secondly, all burnout classifications were made artificially and cannot be validated with the help of the MBI measure due to the absence of clinical information. A clinical partnership collecting MBI-validated labels alongside digital exhaust data would be the single most impactful methodological advance. Thirdly, considering each day as an independent observation leads to loss of information about the development pattern since gradual burnout escalation can cross the daily threshold only but demonstrate weekly trends. Fourthly, the baseline dataset used covers only six countries while South America, Sub-Saharan Africa, and South-East Asia, among other areas, are not included.

Areas of future research could include: (1) time series modelling of burnout with the use of LSTM or TCNs; (2) multimodal approach using emails sentiment and calendar fragmentation metrics, as well as consented biometric measures from wearables; (3) expanding the dataset of the baseline to cover several more cultures; (4) implementation of federated learning technique; (5) conducting a longitudinal randomised study assessing the causal impact of JITAI on burnout rates, attrition, and wellbeing improvement.

VII. CONCLUSION

The Digital Empathy Engine is a system that shows what it means to really take care of employees. It uses intelligence to make sure everyone is treated fairly and that their privacy is respected. This system is special because it can detect when employees are getting burned out. It does this by looking at things like emails sent late at night or, on weekends.

The main idea here is that when people are getting burned out they do things that leave a kind of trail. For example they might send emails late at night or work on weekends when they should be resting. The Digital Empathy Engine looks at these things. Figures out what they mean for each person. It then offers help before things get too bad.

This system shows that we already have the technology to detect and help with burnout. We already have the data and the infrastructure we need. What we were missing was a way to put all these things together that's fair and respects peoples privacy. The Digital Empathy Engine does that.

It is a problem when organisations wait for employees to say they need help or only find out why people left after they are gone. This means they are acting late. This project shows that it is possible to act and that artificial intelligence can make a big difference. If we use intelligence in the right way it can help organisations catch people before they get too burned out. The Digital Empathy Engine is a framework that makes this possible. It helps organisations take care of their employees in a way that's fair and respectful.

Acknowledgment:

The authors gratefully acknowledge Dr. K V Rajesh Kumar, Associate Professor, Department of Analytics, School of Business, Woxsen University, for invaluable guidance and supervision throughout this research. The authors also thank Dr. Hemachandran K, Associate Dean, and Dr. Raul Villamarin Rodriguez, Vice President, Woxsen University, for their continued institutional support.

REFERENCES

[1] C. Maslach and S. E. Jackson, "The measurement of experienced burnout," Journal of Organizational Behavior, vol. 2, no. 2, pp. 99–113, 1981.
[2] Gallup, "State of the global workplace 2021 report," Gallup Press, Washington, DC, 2021.
[3] World Health Organisation, "Burn-out an 'occupational phenomenon': International Classification of Diseases," WHO, Geneva, 2019.
[4] H. J. Freudenberger, "Staff burn-out," Journal of Social Issues, vol. 30, no. 1, pp. 159–165, 1974.
[5] G. Hofstede, Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations, 2nd ed. Thousand Oaks, CA: Sage Publications, 2001.
[6] G. Mark, S. Iqbal, M. Czerwinski, and P. Johns, "Bored Mondays and focused afternoons: The rhythm of attention and online activity in the workplace," in Proc. ACM CHI, Toronto, 2014, pp. 3025–3034.
[7] P. Bhatt, S. Bhatt, and M. Bhatt, "Employee attrition prediction using machine learning," International Journal of Computer Applications, vol. 180, no. 45, pp. 20–26, 2019.
[8] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[9] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
[10] Nahum-Shani, S. N. Smith, B. J. Spring, L. M. Collins, K. Witkiewitz, A. Tewari, and S. A. Murphy, "Just-in-time adaptive interventions (JITAIs) in mobile health," Annals of Behavioral Medicine, vol. 52, no. 6, pp. 446–462, 2018.
[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and E. Duchesnay, "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.