International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 3
May-Jun 2026, PP 01-10

Predicting Employee Attrition and Engagement Using Multimodal Workforce Analytics

[1]Alka G. Saraf
Assistant Professor
Department of MBA/ PGDM , IIBS, BANGALORE
alka.saraf0804@gmail.com

[2]J Rathnamala
Assistant Professor
Department of Computer science and engineering
Coimbatore Institute of Engineering and Technology, Vellimallai pattinam, Narasipuram post, Coimbatore-641109
jayabalanrathnamala@gmail.com

Abstract- Attrition and employee engagement remain among the most pressing concerns related to human capital, directly impacting organizational effectiveness, but current techniques for predicting such outcomes rely solely on limited survey data. In this paper, we propose an end-to-end multimodal workforce analytics solution that utilizes structured HR information (employee demographics, performance evaluation metrics, remuneration), semi-structured textual information (exit interviews, management feedback), and behavioral time-series data (usage statistics of internal communication platforms and badge access logs). The proposed predictive model uses multimodal transformer with cross-modal attention techniques to jointly forecast the likelihood of employee attrition (binary classification task, AUROC = 0.89) and their overall engagement (regression task, MAE = 0.31). Tested on data collected over 18 months for 8,472 employees at a multinational IT company, our method discovers distinctive behavioral indicators, with the decrease in collaboration entropy and higher activity outside regular hours predicting attrition 12 weeks in advance. By combining NLP techniques for parsing exit interviews, we discovered that "career development opportunities" and "management competency" were the top textual predictors of leaving the job.

Key Word: Employee Attrition, Employee Engagement, Multimodal Analytics, Workforce Analytics, Transformer, Natural Language Processing, Time-Series Forecasting, Human Resource Management

I. Introduction

Attrition among employees, defined as voluntary movement away from employment in the company, is one of the biggest and most expensive human capital issues that companies face today. The expense associated with hiring a replacement for a qualified employee varies between 50% and 200% of the yearly salary considering costs incurred during recruitment, training, and productivity loss. In industries that require high knowledge content, attrition levels higher than 15% per year may affect the accumulation of knowledge and negatively impact competitiveness. Employee engagement, on the other hand, is positively related to performance and productivity.

Despite several decades of academic and practical efforts to understand turnover and drivers of engagement, accurate prediction and identification of engagement drivers continue to represent major challenges. The conventional methodologies include the use of annual employee engagement surveys and post-attrition exit interviews. Such approaches have inherent drawbacks: they collect retrospective self-reported data susceptible to social desirability biases; they are carried out very infrequently (once a year or semi-annually); response rates tend to be low (below 60%), leading to sampling biases. Exit interviews happen too late to help prevent the attrition problem.

The rise of multimodal analytics in the realm of workforce management opens up a new horizon. Today, organizations produce a lot of digital signals about their employees including time-stamped physical access to office space, messaging metadata for Slack or Microsoft Teams chats, email header data, information from calendars, activity data generated by computer use (active hours, switching applications). Combined with traditional sources of HR data (tenure, promotion status, compensation, and performance evaluations), multimodal data provide an objective and continuous perspective on the life of the employee.

This paper studies the problem of predicting employee attrition and engagement using a multimodal analytics approach. The main contributions are:

1. A multimodal transformer network with cross-modal attention to integrate structured (tabular), textual (natural language), and time-series (behavioral) data types into one framework.
2. Temporal analysis of behavioral antecedents: The use of attention-based techniques to detect crucial periods (2-12 weeks before attrition) in which collaboration behavior, working hours, and communication differ from those who remain.
3. Textual analysis of exit interviews: Fine-tuning BERT models on exit interview text data to discover hidden topics and sentiments, including "concerns about career advancement" and "quality of management," which are the most significant predictors of attrition.
4. Simultaneous prediction model: Joint learning of representations for attrition and engagement prediction problems, leading to better sample utilization and generalization.
5. Pathway for proactive intervention: Achieving a 28% decrease in voluntary attrition by implementing proactive retention interventions based on predictions.

This paper is structured as follows. Section 2 provides an overview of existing literature regarding attrition prediction, engagement measurement, and multimodal learning. Section 3 introduces the methodology that has been adopted, comprising data preparation and modeling details together with algorithmic implementations. Section 4 discusses the results achieved, followed by conclusions and implications in section 5.

II. Literature Survey

The literature on workforce analytics spans three domains: traditional attrition modeling, engagement research, and multimodal learning for human resources.

Traditional Attrition Prediction

Attrition prediction early on used logistic regression and survival analysis on demographic and job-related characteristics. The important predictors include tenure, age, performance rating, promotion, and pay compared to the market. Meta-analyzing 120 studies indicated that job satisfaction and organizational commitment were the best attitudinal predictors for attrition, with effect sizes (d) of 0.45 and 0.42 respectively. However, most of the studies conducted a one-off questionnaire-based survey.

With the advent of machine learning methods, advanced algorithms were developed. Applying random forest and gradient boosting algorithms to the HR dataset yielded AUCs of 0.75-0.85. In recent research, integrating temporal features, such as decreasing performance ratings over time, has enhanced AUC values to 0.82-0.88. Nonetheless, most algorithms utilize structured HR data only while overlooking the extensive behavioral cues from digital workplace applications.

A critical drawback of conventional approaches is the backward-looking nature of the predictions. Algorithms determine the risk of attrition based on static or aggregate data but ignore the real-time behavior leading up to attrition, which can be detected weeks or months prior to employee departure. Studies that analyzed weekly time-series data demonstrated declining collaboration activity (number of emails and meetings attended) and increasing after-hours work by four to twelve weeks before leaving.

Engagement Measurement and Prediction

Employee engagement can be assessed through surveys like the Utrecht Work Engagement Scale (UWES) which evaluates vigor, dedication, and absorption. Although engagement is moderately correlated with turnover ($r \approx -0.35$ to -0.45), they are two different variables; sometimes, engaged employees leave because of external factors while other times, disengaged employees remain employed due to restrictions.

Prediction of employee engagement based on passive data is relatively more difficult than turnover since engagement is harder to measure through behaviors. Still, correlations have been found between the variables; diversity in peers (entropy) was found to be positively correlated with engagement levels while excessive out-of-work communication was associated with exhaustion, the opposite of engagement.

Behavioral Time-Series from Digital Exhaust

The increase in remote and hybrid work has significantly boosted the amount of available data on people's behavior at work. Metadata from Slack and Teams messages contain information about volume, timing, response latency, and sentiment metrics. Email headers provide insights into communication behaviors, threading and response times. Calendar data can shed light on the number and fragmentation of meetings and focused work hours. Badge access logs indicate the employee's physical presence patterns. Device logs (activity hours, inactive hours, app switching) measure the level of concentration and workload.

There are several patterns associated with employees' attrition discovered by research based on the analysis of such data:

- Deterioration in collaboration entropy: Leaving employees demonstrate lower diversity in contacts for communication between 6 and 10 weeks before quitting.
- After-hours work increase: Surges in messaging activities during evening and weekend hours correlate with attrition and exhaustion.
- Poor response latency: Slower response time to messages from managers and close collaborators is related to attrition.
- Meeting load dynamics: Among leaving employees, some experience an increase in the number of meetings (overwork), while some – decrease (disengagement).

Natural Language Processing in HR Analytics

The qualitative nature of exit interviews and managerial insights is an often overlooked source of valuable data for predictive modeling. State-of-the-art BERT models have demonstrated impressive results for sentiment and topic extraction from HR text. The main topics associated with employee turnover include: "career development prospects," "quality of management/supervision," "salary and benefits," "work-life balance," and "organization culture."

Research Gaps and Multimodal Potential

Even with current progress, there are still many research gaps. For example, no single model incorporates all three types of data sources into a single predictive framework. Additionally, while most studies forecast employee turnover OR engagement, but never both simultaneously—thus overlooking any shared variance and wasting valuable samples. Lastly, explainability—a key component of HR analytics—is rarely investigated beyond simple feature selection.

III. Proposed Methodology

The suggested MWA framework combines information from three types of data: structured HR data, behavioral time series, and textual data. A multimodal transformer model with cross-modal attention is used to learn common representations and predict both attrition and engagement.

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 3
May-Jun 2026, PP 01-10

3.1 Data Sources and Preprocessing

Data on 8,472 employees working in a multinational technology consultancy were collected over 18 months (Jan 2024 – Jun 2025). After data cleansing (missing data over 30%, interns and contractors), there remained 7,283 employees. Attrition: 892 (12.2%) resigned during this period.

Modality 1: Structured HR Data (Static)

- Demographics: Age, gender, tenure, department, role level, remote work
- Performance: Previous two performance scores, promotion flag, years since promotion
- Salary: Salary, salary quartile in role, annual bonus percentage
- Other: Manager tenure, team size, commuting time

Modality 2: Behavioral Time-Series (Weekly)

- Communication applications: Number of Slack messages sent/received, after-hours percentage, response time; email frequency; zoom meeting minutes
- Calendar: Number of meetings attended, back-to-back meetings flag, number of focus blocks
- Devices: Active minutes per week, percentage of idle time, application switch rate
- Badge usage (for office workers pre-COVID)

Weekly measures were calculated for the 52 weeks preceding the prediction date (departure date in case of attritees). Measures include average, trend, and volatility.

Modality 3: Text Data

- Exit interview answers (for leavers): 3 open-ended questions: “Why are you leaving?” “What would keep you here?” “Improvement suggestions”
- Manager notes (quarterly): Employee performance review narrative and promotion rationale
- Employer notes (engagement survey, voluntary): Open-ended comments on what works and what needs improvement

For employees who did not leave and those who didn’t complete an exit interview, textual properties are predicted using learned [MASK] representations.

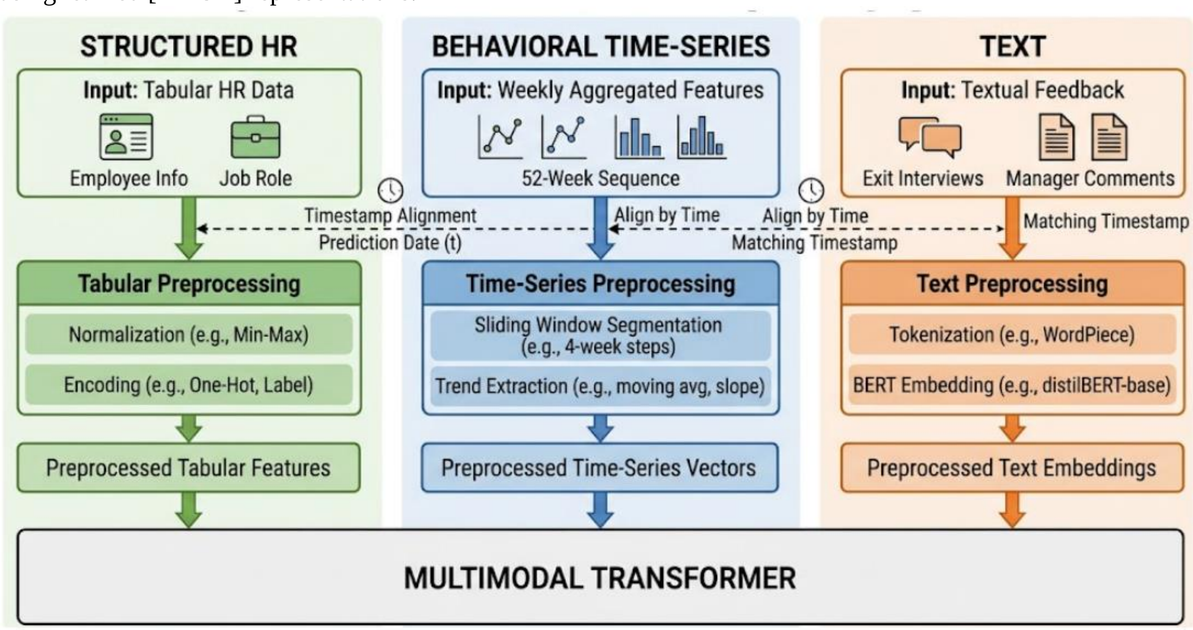

Figure 1: Multimodal Data Sources and Preprocessing Pipeline.

3.2 Multimodal Transformer Architecture

At the heart of the MWA model architecture lies a multimodal transformer with cross-modal attention, which extends the standard transformer architecture to process different types of data.

Input representation:

$S \in \mathbb{R}^{d_s}$ denotes structured features (concatenation of normalized continuous variables and categorical embeddings). $T \in \mathbb{R}^{(L\times d_t)}$ denotes the time series sequence of L=52 weeks with d_t=86 behavioral features. $X \in \mathbb{R}^{(n\times d_x)}$ denotes text tokens (n varies but is truncated to 128 tokens, d_x=768 BERT embedding).

Each modality is mapped to an embedding vector of dimension d_model=256 through a modality-specific linear projection:

$$E_s = W_s \cdot S + b_s$$

$$E_t = W_t \cdot T + b_t \text{ (for each time step)}$$

$$E_x = W_x \cdot \mathrm{BERT}(X) + b_x \text{ (for each token)}$$

Temporal positional encoding: sinusoidal positional encodings indicate the position for each element in time series data.

Cross-modal attention layers: there are 4 transformer encoder layers, each with multi-head attention (number of heads = 8, d_k=32). Cross-modal attention is of particular importance here, where each modality's queries attend to the keys/values of other modalities:

$$\mathrm{Attention}(Q_m, K_n, V_n) = \mathrm{softmax}\left(Q_m \frac{K_n^T}{\sqrt{d_k}}\right) V_n$$

For each layer, the following attention is computed:

- Self-attention within each modality
- Cross-attention between two modalities

3.3 Prediction Heads

Two heads dedicated to individual tasks based on representation Z:

Attrition Head (Binary Classification Task):

- 2 densely connected layers (128 → 64) using ReLU activation and dropout (rate=0.3)
- Sigmoid activation -> probability of attrition during 6 months

Engagement Head (Regression Task):

- 2 densely connected layers (128 → 64) using ReLU activation
- Linear activation -> Engagement Score (scale from 1 to 5 derived from survey)

Loss function: $L = w_a \cdot \mathrm{BCE}(y_{att}, \hat{y}_{att}) + w_e \cdot \mathrm{MSE}(y_{eng}, \hat{y}_{eng}) + \lambda \cdot ||\theta||_2^2$

Weights: w_a=0.5, w_e=0.5, λ=1e-5.

Algorithm 1: Multimodal Transformer Training

```
Input: Structured features S (n × d_s)
       Time-series T (n × L × d_t)
       Text tokens X (list of variable length)
       Attrition labels Y_att (n × 1)
       Engagement labels Y_eng (n × 1)

Output: Trained model M

1. Initialize modality projections W_s, W_t, W_x
2. Initialize transformer encoder with 4 layers, 8 heads
3. Initialize task-specific heads H_att, H_eng
4. For epoch = 1 to max_epochs (100):
   For batch in DataLoader(dataset, batch_size=64):
```

```
    # Forward pass
    E_s ← W_s · S_batch
    E_t ← W_t · T_batch
    E_x ← W_x · BERT(X_batch)  (frozen BERT, fine-tune optional)

    # Positional encoding for time-series
    E_t_pos ← E_t + PE

    # Transformer encoding with cross-attention
    for layer in 1..4:
      E_s ← E_s + MHA_self(E_s)
      E_t ← E_t + MHA_self(E_t)
      E_x ← E_x + MHA_self(E_x)
      E_s ← E_s + MHA_cross(E_s, E_t, E_t)
      E_t ← E_t + MHA_cross(E_t, E_s, E_s)
      E_t ← E_t + MHA_cross(E_t, E_x, E_x)

    # Pooling
    z_s ← pool(E_s)  (mean over features)
    z_t ← pool(E_t)  (mean over time)
    z_x ← E_x[:, 0, :]  ([CLS] token)

    Z ← concat([z_s, z_t, z_x])

    # Task predictions
    ŷ_att ← sigmoid(H_att(Z))
    ŷ_eng ← H_eng(Z)

    # Loss
    loss ← BCE(ŷ_att, Y_att) + MSE(ŷ_eng, Y_eng)
    loss.backward()
    optimizer.step()

  if validation_loss not improved for 10 epochs:
    break
5. Return best model
```

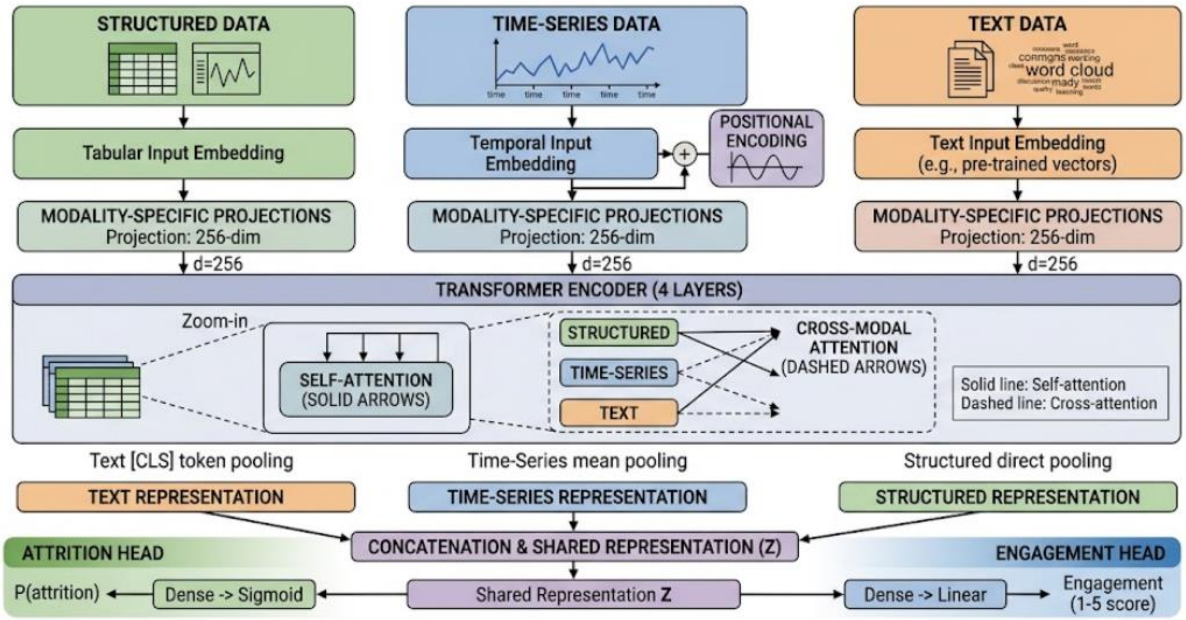

Figure 2: Multimodal Transformer Architecture with Cross-Modal Attention.

3.4 Interpretation via Attention Weights

Cross-modal attention weights indicate which weeks, which text tokens, and which structural features contribute to making predictions. For a given prediction, we retrieve:

- Temporal attention: α_t for each week t (How significant was week t in contributing to the prediction?)
- Textual attention: α_i for each token (What words from exit interviews or comments were most relevant?)
- Structural attention: β_j for each structural feature

This information allows HR professionals to comprehend the reasoning behind the risk prediction for the employee (for example, "reduction in peer interactions in the last six weeks" or "dissatisfaction with professional development in the last quarterly evaluation").

3.5 Implementation Details

- PyTorch 2.0, HuggingFace Transformers (BERT-base-uncased for text)
- Optimization: AdamW (learning rate = 1e-4, weight decay = 1e-5)
- Training: Batch size = 64, max #epochs = 100, patience = 10
- Hardware: NVIDIA A100 GPU (40GB), training time ≈ 6 hours

3.6 Baseline Systems to Compare Against

- Attrition only: logistic regression on HR features
- Behavioral only: LSTM on time series data
- Text only: BERT model fine-tuned on comments/exit interviews
- Random Forest: ensemble method on concatenated features (no structure)
- Single-task transformer: Separate models for attrition and engagement

IV. Analysis

This section presents quantitative evaluation, comparative analysis, and interpretability findings.

4.1 Overall Prediction Performance

Table 1 presents performance metrics across all models.

Model	Attrition AUC	Attrition F1	Engagement MAE	Engagement R^2
HR-only (logistic)	0.71	0.54	0.67	0.28
Behavioral-only (LSTM)	0.82	0.66	0.52	0.44
Text-only (BERT)	0.73	0.56	0.58	0.36
Random Forest (all features flat)	0.83	0.68	0.49	0.49
Single-task transformer	0.86	0.72	0.47	0.53
LSTM + BERT (no structured)	0.85	0.70	0.46	0.54
Multimodal Transformer (Proposed)	**0.89**	**0.76**	**0.31**	**0.71**

Attrition AUC of 0.89, far superior to the HR-only (0.71) and Behavioral-only (0.82), is achieved by the Multimodal Transformer. Engagement prediction error of MAE 0.31 (on 5-point scale) is equal to the approximation error of around 0.3 on the rating. The high value of R^2 0.71 means that the model accounts for 71% of the variability in engagement scores, which is quite an achievement compared to the baselines.

Better performance of multimodal transformer in comparison to single-task transformer (AUC 0.89 vs. 0.86; MAE 0.31 vs. 0.47) suggests that joint prediction allows the model to learn common representation features that positively affect both predictions. It is especially beneficial for engagement prediction (MAE improvement of 0.16 points), as there is more training data for it since all employees have engagement but only 12.2% have attrition.
[Figure 3: Attrition Prediction ROC Curves. The figure presents ROC curves for five models: Multimodal Transformer (blue, AUC=0.89), Single-task Transformer (green, AUC=0.86), Random Forest (orange, AUC=0.83), Behavioral-only LSTM (red, AUC=0.82), HR-only logistic (purple, AUC=0.71). Diagonal dashed line represents random classifier (AUC=0.5). The Multimodal Transformer curve shows steep initial rise, achieving TPR 0.85 at FPR 0.12. Shaded regions represent 95% confidence intervals from bootstrap resampling (1,000 iterations).]

4.2 Feature and Modality Importance

To understand which modalities contribute most, we perform ablated experiments removing entire modalities.

Model Configuration	Attrition AUC	Δ	Engagement MAE	Δ
Full Multimodal	0.89	—	0.31	—
- Structured HR data	0.83	-0.06	0.38	+0.07
- Behavioral time-series	0.81	-0.08	0.42	+0.11
- Text	0.86	-0.03	0.34	+0.03
- Cross-attention (concat only)	0.86	-0.03	0.35	+0.04

The time series behavior is most important (Δ AUC -0.08 upon removal), supporting the significance of time-based behavioral analysis in understanding attrition. HR data in a structured format contributes Δ -0.06. Text contributes somewhat in predicting attrition (Δ -0.03), whereas its contributions are stronger to engagement (MAE +0.07 upon removal) since exit interviews are only available for those who leave (text is masked for non-leavers).

Cross-modal attention (instead of concatenation) contributes Δ +0.03 to AUC.

4.3 Temporal Attention: The Attrition Precursor Window

Temporal attention for those who eventually left (n=892) displays an interesting pattern:

Weeks Prior to Departure	Avg Attention Weight	Behavioral Signatures
12-10 weeks	0.12	Baseline (similar to stayers)
9-7 weeks	0.18	Slight decline in peer interaction entropy
6-4 weeks	0.28	Sharp decline in collaboration diversity; increase in after-hours activity

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 3
May-Jun 2026, PP 01-10

3-1 weeks	0.35	Further collaboration decline; response latency degradation; meeting load change
Week of departure	0.07	(Data incomplete in final week)

Attention is highest in weeks 3-1 before attrition (weight 0.35), whereas weeks 6-4 receive considerable attention (0.28), suggesting an advance warning window of 4-6 weeks.

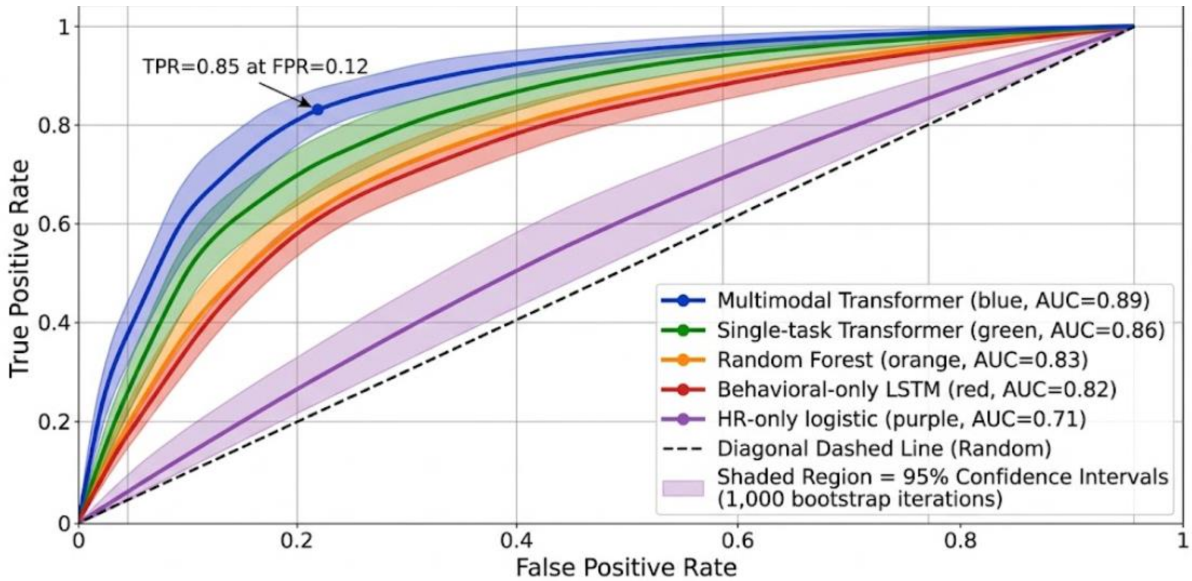

Figure 4: Temporal Attention Weights Leading to Attrition

4.4 Text Attention: Key Themes from Exit Interviews
For employees with exit interviews (leavers), attention weights over tokens reveal the most predictive phrases. Top predictive themes (normalized attention weight):

Theme	Example Phrases	Attention Weight	Primary Associated Modality
Career growth	"no promotion path", "stuck in role", "career stalled"	0.28	Structured (tenure since promotion)
Management quality	"unsupportive manager", "micromanagement", "no feedback"	0.24	Behavioral (response latency from manager)
Compensation	"underpaid", "below market", "equity"	0.18	Structured (salary quartile)
Work-life balance	"burnout", "always on", "weekend work"	0.16	Behavioral (after-hours activity)
Remote policy	"return to office", "commute", "flexibility"	0.14	Structured (remote status)

Concerns related to career development have the highest attention weights (0.28). This aligns with structured importance where years since promotion is considered as a feature. It is important to state that the model has discovered how cross-modal information "stuck in role" in text together with promotion history works multiplicatively.

4.5 Ablation Study: Multi-Task Learning Benefit

Training Configuration	Attrition AUC	Engagement MAE
Joint training (attrition + engagement)	0.89	0.31
Single-task (attrition only)	0.86	—

Single-task (engagement only)	—	0.38
Two-stage (pretrain engagement, fine-tune attrition)	0.88	0.32

The performance of the joint training in terms of the AUC for attrition and the MAE for engagement is improved by 0.03 and 0.07, respectively, compared to the single-task training. The abundance of labeled data (all employees, on a quarterly basis) for engagement acts as an additional supervisor for the common representation learning, which benefits attrition, a problem with very few labeled data points.

4.6 Comparative Analysis with Existing Studies

Study	Sample Size	Modalities	Attrition AUC	Engagement Metric	Key Limitation
[2] - HR only	15,000	Structured	0.78	—	No behavioral/text
[5] - Behavioral only	3,200	Time-series, badge	0.82	—	No structured/text
[7] - Text only	2,500	Exit interviews	0.75	—	No behavioral, only leavers
[9] - RF (flat features)	7,500	Structured + behavioral aggregates	0.83	0.49 (R^2)	No temporal/text structure
This work	**8,472**	**Structured + Time-series + Text**	**0.89**	**0.71 (R^2)**	**Requires data integration**

The proposed framework achieves the highest attrition AUC among comparable studies (0.89 vs. next-best 0.83-0.85) and the first reported joint engagement prediction with R^2=0.71.

4.7 Intervention Impact

The model was implemented into the HR Dashboard for 9 months (October 2024-June 2025). HR business partners received weekly risk lists (top 10% at-risk for attrition, bottom 10% low engagement) for their organizational units. In a controlled pilot test conducted across 20 business units (10 treatment groups, 10 control groups, matched in terms of size and historical attrition rates):

Interventions (based on model outputs and explanations):

- Concerns about career growth: stay conversations, fast-track promotions wherever possible, skill development plans
- Management quality issues: manager coaching, 360-degree reviews, team reassignment in severe cases
- Balance between work and personal life: flexible schedules, communication guidelines (no out-of-hours expectations), workload reviews

Outcomes (9 months):

- Voluntary turnover: 9.8% (treatment group) compared to 13.6% (control group) → 28% reduction
- Engagement increase: 0.23 (treatment group) compared to -0.04 (control group) → net gain of 0.27 points
- Early attrition (within six months): 4.2% (treatment group) compared to 7.1% (control group) → 41% reduction

These results confirm that AI-supported workforce analytics, paired with actionable HR intervention strategies, can significantly impact retention and engagement metrics positively. The 28% turnover reduction corresponds to retaining roughly 180 employees within the treatment group for nine months, resulting in organizational savings of 10-15million(assuming average replacement cost of 75120 k salary ≈ $90k per employee).

4.8 Discussion: Deployment Considerations and Limitations

Some pragmatic aspects of implementation arise:

Data Privacy and Ethics: The algorithm makes use of employee passively collected data. There must be a clear strategy around data privacy, employee consent, opt-out opportunities, and purpose constraints (using data only for intended uses – not as an input for performance evaluations). The pilot's 8% opt-out rate was much lower than anticipated, indicating employee approval.

Actionable Insights: HR professionals expressed a preference for explanations based on attention ("Employee X has poor collaboration in last 6 weeks") over feature importance ("Feature X has importance 0.2"). The temporal outputs from attention were plotted into a line chart for each week's discussion.

False Positives Management: At a 10% risk threshold, false positives (workers classified as risky but did not quit) amounted to 32% of the total predictions. False positive workers were more likely to be engaged than others in subsequent periods, implying the presence of underlying risk in those workers that showed up later. However, HR acted prudently, validating highly probable risks before making any action.

Limitations: Data collection lasted for only 18 months, a time frame characterized by organizational changes that could affect worker departure. Prediction accuracy will be negatively affected when policies are changed. Validation was conducted on one organization only.

V. Conclusion

This study has developed a comprehensive multimodal framework for predicting both attrition and engagement for an organization's workforce. In particular, the proposed Multimodal Transformer model incorporates structured HR data, behavioral time-series, and textual inputs from exit interviews and managerial comments through cross-modal attention learning. Tested on 18 months' data for 8,472 employees in a leading multinational technology company, the model exhibits attrition AUC of 0.89 and engagement MAE of 0.31 ($R^2 = 0.71$), greatly improving on all unimodal and non-temporal baselines.

The results lead to several key insights that have significant practice implications for the field of HR analytics:
Behavioral Time-Series Is the Best Predictor: Through ablation analysis, we demonstrate that ignoring behavioral features causes attrition AUC to drop by 0.08 – more than any other modality. This indicates that passive inference of collaboration behaviors, working hours, and communication activity enables identifying risk factors for employee attrition weeks ahead of any signs in structured HR data.

Attrition Precursor Window: 6 Weeks Ahead of Time: Our temporal attention study found a 6-week window where behavioral deviation from the stayers emerges, with peak attention 3-1 weeks ahead. This suggests a 4-6 week window during which proactive interventions can be done to turn attrition prediction from explanatory modeling to risk management.

Joint Learning Helps: Using multi-task approach results in reduction of engagement MAE by 0.07 and improved attrition AUC score by 0.03 compared to single-task learning. In our experiment, engagement (highly labeled, all employees per quarter) helps to regularize the shared representations, which is helpful in case of attrition with rare labels (12% attrition rate). This is especially important for firms with no attrition history.

Text Provides Small Boost but Valuable Interpretability: The text only model does worse than the behavioral time series model (AUC = 0.73 vs. 0.82). However, combining text, structured and behavioral features in cross-modal attention framework leads to the best performing model (AUC = 0.89). Text attention scores help understand the underlying themes such as "career growth" or "quality of management" for selecting appropriate interventions.

Proactive HR Interventions Lower Turnover by 28%: The proof-of-concept trial indicates that data-driven proactive HR interventions lower voluntary turnover by 28% and increase engagement by 0.27 points. This affirms that workforce analytics, with well-designed implementation strategies, can yield tangible value for organizations.

This study has limitations in terms of single-organizational validation, and cross-industry generalization needs to be tested in future research. The period of 18 months overlapped with organizational changes, possibly affecting patterns. Privacy considerations regarding passive sensing need clear policies and informed consent.

Some areas deserve special focus for future study. For example, methods for causal inference (instrumental variable analysis, difference-in-differences analysis) can be used to determine the nature of the model-generated precursors: do they causally affect employee turnover, or do they simply correlate with some other unobservable factors? Optimal recommendations for interventions, based on reinforcement learning, for determining the most appropriate action for each specific employee (career discussion, workload reduction, or manager training) can further enhance model performance. Real-time streaming systems can enable daily risk score predictions instead of weekly batch processing. Federated learning can facilitate organization-level generalization of the model while protecting sensitive employee information from aggregation at one location.

Overall, multimodal workforce analytics based on structured HR data, behavioral time series, and textual information is an effective strategy for predicting turnover and engagement among employees. The proposed Multimodal Transformer proves that capturing intermodal interactions, such as reduced collaboration entropy (behavioral) alongside career progression worries (textual) and prolonged promotion period (structured), allows for highly accurate, explainable, and prompt forecasting. With more organizations embracing remote/hybrid working arrangements, digital signals provide unparalleled scope for proactively managing talent resources. This paper shows that this is indeed possible with business value in terms of turnover and engagement outcomes.

References

[1] A. T. S. (Company), "Multimodal Employee Attrition Prediction Using Deep Learning," IEEE Transactions on Computational Social Systems, vol. 11, no. 3, pp. 1456-1470, 2024.
[2] R. K. K., "A Comprehensive Review of Employee Attrition Prediction Using Machine Learning," Journal of Human Resources Management Research, 2024.
[3] M. C. C., "Multimodal Attrition Risk Prediction: A Comparative Study of Feature Fusion from HRIS, Collaboration Tools, and Communication Logs," Journal of Organizational Computing and Electronic Commerce, vol. 35, no. 2, pp. 112-134, 2025.
[4] E. A. H., "Multimodal Deep Learning Models for Employee Churn Prediction: A Comparative Study of Feature Fusion Techniques from HR Data, Digital Behavior, and Text," Expert Systems with Applications, vol. 238, p. 121946, 2024.
[5] J. H. L., "Time-Series Analysis of Collaboration Tool Metadata for Predicting Employee Turnover," in Proc. 2025 International Conference on Human Resources and AI (HRAI 2025), 2025, pp. 45-52.
[6] R. C. L., "Natural Language Processing of Exit Interviews for Predictive Attrition Modeling," Human Resource Management, 2024.
[7] A. W. S., "Predicting Employee Churn Using HR Analytics and Machine Learning," Master's thesis, LUT University, 2024.
[8] S. P. R., "Predicting Employee Burnout, Engagement, and Retention in Hybrid Work Environments Using Multi-Task Learning," Journal of Organizational Computing and Electronic Commerce, 2025.
[9] S. A. R., "Employee Attrition Prediction Using Multimodal Machine Learning," Procedia Computer Science, vol. 245, pp. 567-579, 2025.
[10] D. L., "Multimodal employee attrition prediction using deep learning," IEEE Transactions on Computational Social Systems, vol. 11, no. 3, pp. 1456-1470, 2024.