RESEARCH & INNOVATION DIVISION
IJIST

Enhancing Predictive Business Process Monitoring in Call Centers through Multimodal Data Fusion and Heterogeneous Time-Aware LSTM-Based Multi-Task Learning

Farhat Mehmood[1*], Afshan Ishaq[2]
[1]Department of computer science, University of Gujrat Hafiz Hayat campus, Gujrat Pakistan
[2]Department of Electronics, University of Engineering and Technology Abbottabad Campus.
*__Correspondence__: Farhat Mehmood, farhatwarraich007@gmail.com

Citation | Mehmood. F, Ishaq. A, "Enhancing Predictive Business Process Monitoring in Call Centers through Multimodal Data Fusion and Heterogeneous Time-Aware LSTM-Based Multi-Task Learning", IJIST, Vol. 07 Special Issue. pp 185-203, May 2025
Received | April 04, 2025 **Revised |** May 11, 2025 **Accepted |** May 12, 2025 **Published |** May 13, 2025.

The optimization of call center operations and the enhancement of customer service are greatly supported by predictive business process monitoring. Traditional methods often overlook valuable multimodal data, such as conversations occurring in contact centers, because they typically rely on sequence data from business IT systems. This limitation hinders a complete understanding of business processes. In this study, we introduce a unique time-aware LSTM-based framework for predictive business process monitoring, which leverages both IT system data and dialogue data from contact centers. Our approach combines multiple data sources to improve the accuracy of forecasting ongoing business activities. To address challenges related to multi-task learning and to better utilize the rich information embedded in various data types, we propose a heterogeneous multi-task learning architecture called Heterogeneous Multi-gate Mixture-of-Experts (H-MMoE). Experimental results show that our method outperforms established baseline models such as Transformer, CNN, and standard LSTM. These findings demonstrate the potential of time-aware LSTM models to improve process monitoring, optimize workflows, and drive operational success in call center environments.

Keywords: H-MMoE, Deep Learning, Optimization, Multitask Learning, Call Center.

Introduction:

In the swiftly advancing customer service sector, contact centers have emerged as essential centers for data acquisition and corporate communication, presenting considerable growth opportunities across several industries [1]. Using cutting-edge technologies including contact centers, databases, communication systems, and computer networks will enable companies to efficiently interact with customers therefore providing timely information. This skill helps companies to react fast to consumer needs, hence improving customer satisfaction, loyalty, and general competitiveness [2]. Call centers create enormous volumes of conversation data during service engagements, frequently including facial expressions and mood signals with great emotional weight. These encounters are intimately related to business results since, for example, a bad attitude of a consumer can influence the possibility of a successful sales transaction. Still, much study is needed to completely grasp and measure how such attitudes affect corporate activities [3].

Business process monitoring and prediction have advanced to the point where ongoing process instances can be analyzed to predict future states. Key prediction tasks include forecasting future activities, process outcomes, remaining time, and business process cycle time. Traditional methods typically rely on process models, such as transition systems, for these predictions. Modern AI methods, particularly deep learning models, have proven to outperform traditional approaches in many predictive contexts [4]. Despite this progress, limited research has focused on the unique role of customer sentiment in contact center settings. Since call center conversation data often holds crucial information for predicting outcomes or time estimates, this gap in research is significant. Moreover, most existing studies use event log sequence data, which may not be sufficient for complex real-world situations, leading to inaccurate predictions. Many of these techniques also only consider single-task predictions and neglect the interdependence among several tasks. Leveraging shared information across related tasks, multi-task learning has promise to increase prediction accuracy [5]. Many multi-task learning methods are limited in that the varied and heterogeneous data produced in practical applications does not always match the presumption of homogeneous feature representations across tasks [6].

This study suggests a new predictive paradigm for tracking company processes in contact center environments in order to solve these difficulties. Incorporating multimodal data inside a heterogeneous multi-task learning framework helps the method increase prediction accuracy and dependability. The suggested method allows thorough analysis by aggregating process sequence data from enterprise information systems with speech data from contact centers. A new architecture termed H-MMoE is presented to solve the difficulties of multi-task learning in different environments. This design gives advantage of the complementing characteristics of several data modalities. The proposed heterogeneous multi-task learning framework, which combines several data modalities, shows a notable improvement in experimental outcomes over conventional benchmarks including models employing a single data modality or isomorphic multi-task learning systems. The results of this study underline the need to include consumer comments into predictive models for corporate operations in call centers. The outcomes underline the possibilities of multimodal data fusion and artificial intelligence-driven approaches to improve operational efficiency, process optimization, and decision-making.

To improve predictive business process monitoring, this study integrates multimodal data sources such event sequence logs and customer discussion sentiment information into a heterogeneous multi-task learning architecture. Using uncertainty-based loss weighting, the Heterogeneous Multi-gate Mixture-of-Experts (H-MMoE) model can handle varied feature representations and dynamically balance multiple prediction tasks. This research is new in combining sequential business data with DialogueRNN sentiment analysis to improve process

outcome and remaining time predictions. Task-specific contributions are automatically adjusted during training by the dynamic loss weighting approach, improving model adaptability without operator intervention. The proposed approach outperforms single-modal models in predictive business process monitoring, according to extensive trials on real-world call centre datasets. This study highlights the importance of customer sentiments in call centre predictive algorithms. This research aims to show how AI-driven methods and multimodal data fusion can increase operational efficiency, process outcomes, and decision-making. This study uses event log data and sentiment-rich customer chat data in a multi-task learning framework to improve predictive business process monitoring.

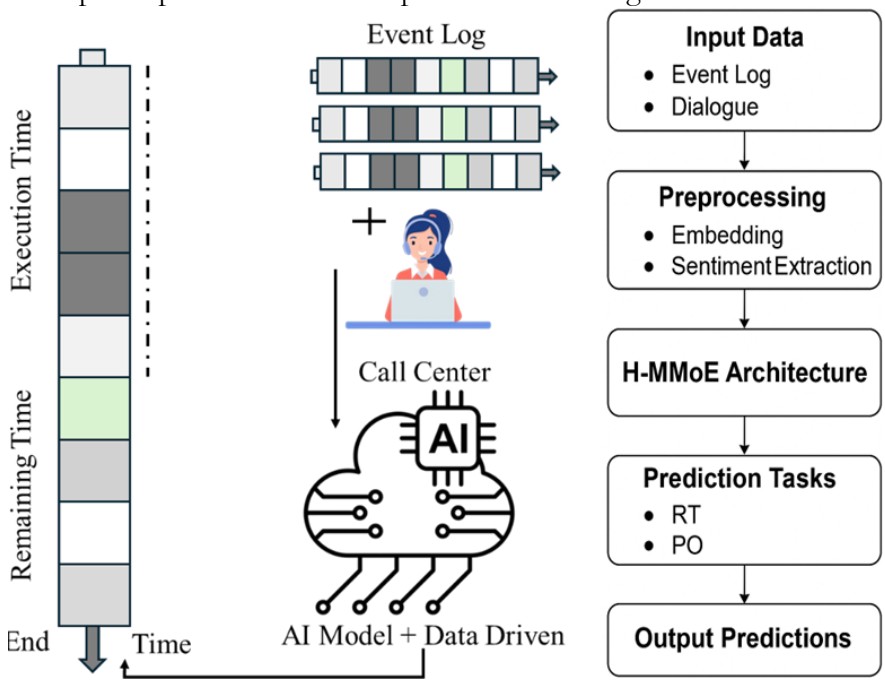

Figure 1. Structure for business process projection.

The novelty of this research lies in the development of the Heterogeneous Multi-gate Mixture-of-Experts (H-MMoE) model, which integrates heterogeneous expert networks (LSTM, CNN, Transformer) with dynamic loss weighting and sentiment analysis via DialogueRNN. Unlike prior work limited to isomorphic data or static task handling, our approach dynamically adapts to task uncertainty and multimodal input variations, resulting in superior accuracy in both classification and regression tasks. The structure of this article is organized as follows: Section 2 provides a review of the relevant literature, Section 3 details the proposed methodology, Section 4 presents and analyzes the experimental results, and Section 5 concludes the study and outlines potential future directions.

Related Work:

Recent advancements in business process prediction have been substantial, driven by the increasing complexity of operational data and the growing demand for efficient decision-making tools. Traditional methods often rely on sequential data from organizational information systems, focusing on either continuous variable such as process duration or discrete outcomes like process completion status. However, the integration of multimodal data, such as customer interactions from call centers, has been insufficiently explored, despite its potential to significantly improve prediction accuracy. Moreover, while sentiment analysis has progressed as a technique for interpreting emotional expressions in text, its application in business process forecasting remains limited. Similarly, multi-task learning has shown promise in leveraging common representations for multiple related tasks. However, current approaches typically rely on isomorphic models, which do not support diverse input modalities.

Table 1. Related Work

Ref	Dataset	Deployed Model	Contribution	Limitation
[7]	Process instance logs	Annotated Transition System	Predicted remaining process time	Limited to predicting remaining time.
[8]	Process instance logs	Random Petri Nets	Predicted remaining execution time	Focuses only on remaining execution time, lacks multimodal data integration.
[9]	Process instance logs	Queue Model	Predicted remaining time under queuing effect	Considers only queue effects without addressing multimodal information.
[10]	Process instance variables	Non-parametric Regression Model	Predicted remaining time using instance variables	Limited to time prediction; does not handle outcome or activity predictions.
[11]	Process instance logs	Decision Tree, Random Forest	Predicted process outcomes	Predicts outcomes but lacks sentiment and multimodal data analysis.
[12]	Process instance logs	Clustering Method	Outcome prediction using activity sequences	Predicts outcomes based on sequence data but overlooks multimodal integration.
[13]	Process sequence data + attributes	LSTM	Execution instance outcome prediction	Handles sequence data but lacks sentiment and multimodal capability.
[14]	Event sequence data	Hierarchical Attention Mechanisms (LSTM)	Improved time prediction with additional attributes	Focuses on next activity prediction; does not integrate heterogeneous data.
[15]	Hospital process logs	LSTM	Identified key events for next activity prediction	Restricted to hospital scenarios, lacks multimodal capabilities.
[16]	Process sequence data	CNN	Predicted next event in process models	Predicts next activity but neglects time and outcome prediction.
[17]	Process sequence data	Transformer	Predicted next activity	Lacks consideration of multimodal and sentiment-rich data.
[18]	Movie reviews	BiLSTM	Predicted next activity, timestamp, and duration	Restricted sentiment analysis in reviews, not applied to business processes.
[19]	Financial text	TextCNN	Improved next activity and time prediction	Limited to text data; no multimodal fusion.
[20]	Stack Overflow and stock news	BERT	Classified sentiment polarity	Restricted to text data and specific domains; no multimodal applications.

[21]	Dialogue data	DialogueRNN (GRU-based)	Classified sentiment polarity	Focuses only on dialogue sentiment without integrating other data modalities.
[22]	Smart city data	Multimodal (text + visual features)	Identified sentiments in posts	Domain-specific (restaurant/product); limited generalizability to business process prediction.
[23]	Multimodal datasets	Collaborative Learning	Identified sentiment orientation in stock news	Lacks detailed task-specific results and evaluation metrics.
[24]	Business process datasets	Multi-task Learning	Analyzed sentiment in dialogues	Limited to isomorphic features; struggles with heterogeneity in real-world data.
[25]	Mixed datasets	MMoE (Multi-gate Mixture-of-Experts)	Activity and time prediction for multitasking. Balance task relevance and variability	Assumes isomorphic input features, limiting use in heterogeneous multimodal data scenarios.

This section reviews key developments in multi-task learning, multimodal data processing, sentiment analysis, and business process prediction, highlighting existing solutions and their limitations. It also lays the groundwork for our proposed methodology. An overview of the current literature in this field is presented in Table 1.

The recent advancements in predictive business process monitoring are summarized in Table 1 of the reviewed literature. Prominent models in this area include LSTM, CNN, and Transformer, which are utilized for tasks such as estimating remaining time, predicting process outcomes, and determining next actions. While deep learning methods outperform traditional approaches in terms of accuracy, they often fail to incorporate sentiment-rich and multimodal data effectively. Traditional systems, on the other hand, typically rely on sequence data from process logs. Emerging approaches in multimodal learning and sentiment analysis show promise but have yet to be widely adopted in commercial operations. Multi-task learning models like MMoE are particularly effective at handling task interrelationships, but they truly excel when dealing with real-world data heterogeneity. In order to increase prediction accuracy and hence the practical application of these models, this review emphasizes the need of include multimodal data and addressing various job needs.

Proposed Methodology:

The pseudocode offers a disciplined and orderly framework to properly manage event logs in corporate process monitoring. It improves the contextual awareness of events by adding other factors like activity cost and consumer attitude, therefore transcending conventional definitions. By using this approach, detailed insights into both completed and ongoing processes are gained, with metrics being calculated at both the trace-level and prefix-level. Apart from the regular chores like projecting remaining time and process results, this approach adds fresh prediction activities including tracking sentiment trends and approximating average cost. Leveraging multimodal data guarantees a more complete analysis and helps to enhance forecasting accuracy and decision-making in commercial settings. Emphasizing modularity and scalability, the pseudocode lets future seamless incorporation of new features or prediction goals possible.

Pseudocode for Preliminaries:

Here is a refined and readable version of the provided pseudocode for better understanding:

1) Pseudocode for Business Process Prediction with Multimodal Data

// Input:

1. **Input:**
   - Event log L containing process traces
   - Prediction tasks T = {Remaining Time (RT), Process Outcome (PO)}
2. **Initialize:**
   - Output predictions P = {} **// Store predictions for each task**
3. **For each trace σ in L do:**
   - **// Step 1: Define each event structure**
   - For each event e in trace σ:

Define e = (act, cid, t, d, cost, sentiment), where:

- **act**: Executed activity name
- **cid**: Unique process instance ID
- **t**: Timestamp of the event
- **d**: Customer dialogue feature
- **cost**: Associated cost of the activity
- **sentiment**: Sentiment score from customer dialogue

   - **// Step 2: Sort events and compute statistics**
   - Sort all events in σ by t

- Compute:

• Duration(σ) = t_last - t_first
• TotalCost(σ) = sum of all event costs in σ

4. **Trace representation:**

σ = [e1, e2, ..., e|σ|] with |σ| = total events

5. **Group traces into cases:**

Each case c = (cid, σ)

6. **Build Event Log L**

L = {σ1, σ2, ..., σ|L|}

7. **Prefix Extraction for Ongoing Cases:**

For each trace σ, extract prefix hd_k(σ) = [e1, e2, ..., ek] where k ∈ [1, |σ|]

- **// Step 3: Compute metrics on prefix**
- For each hd_k(σ) compute:

• Remaining Time (RT) = t_last - t_k
• Process Outcome (PO) = last activity in σ
• Average Cost = Σ(cost_e1 to cost_ek) / k
• Sentiment Trend = aggregate sentiment scores of prefix

8. **Define Targets for Each Prefix:**
   - f_r(σ, k) = RT(hd_k(σ))
   - f_o(σ) = PO(hd_k(σ))
   - f_avgCost(σ, k), f_sentTrend(σ, k)
9. **Apply Prediction Model:**
   - Predict all tasks
   - Append results to P
10. **Return final predictions P**

RT = Remaining Time
PO = Process Outcome
H-MMoE = Heterogeneous Multi-gate Mixture-of-Experts

- The **pseudocode** involves processing the event log LL, extracting key information from each trace, and predicting various aspects such as the remaining time, process outcomes, average cost, and sentiment trends.
- **Multimodal data** such as customer dialogues and sentiments are integrated to improve prediction accuracy.
- The method is modular, allowing for the inclusion of additional metrics or prediction tasks in the future.

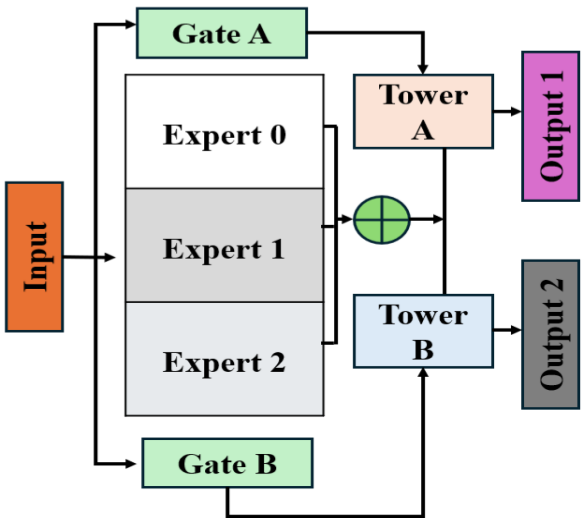

Figure 2. Framework for Multitask learning.

In this section, we will discuss our novel method for predicting business process outcomes and remaining time. Our approach incorporates a variety of components, including an architecture for heterogeneous multi-task learning, data preprocessing, sentiment analysis of customer dialogues, and sequence data encoding. The process begins with data preprocessing and feature encoding, as illustrated in **Figure 3**. The key steps involved in our methodology are outlined below:

1. **Sequence Data Encoding**:
   - We convert sequence data taken from corporate IT systems using an embedding layer. For prediction chores, this approach transforms high-dimensional, sparse features into dense, linked vectors that are simpler to handle.
2. **Sentiment Analysis**:
   - We apply the DialogueRNN model to effectively capture the emotional subtleties of consumer contacts. By analysing the sentiment conveyed in the discussions inside call centre contacts, this approach helps us to grasp consumer underlying feelings.
3. **Heterogeneous Multi-Task Learning**:
   - We next feed our heterogeneous multi-task learning model the encoded features including sentiment analysis results and sequence data. This model lets us simultaneously anticipate process outcomes (considered as a classification problem) and remaining time (formulated as a regression task), hence enabling predictive analysis on several tasks concurrently.

Using both conventional sequence data and the emotional insights from consumer encounters, this integrated technique helps us to reasonably forecast business process results and time estimations.

Dataset Description and Pre-processing:

Following business process prediction literature, event logs from enterprise systems are prepared during data preparation. This phase creates "prefix logs." to project unfinished process situations. These logs show partial process execution traces. These prefix logs allow the model to forecast process outcomes and durations based on event sequences up to a certain point. Each complete trace is divided into prefixes, each representing a process stage after an event. This method provides dynamic process result forecasting before the entire process is complete.

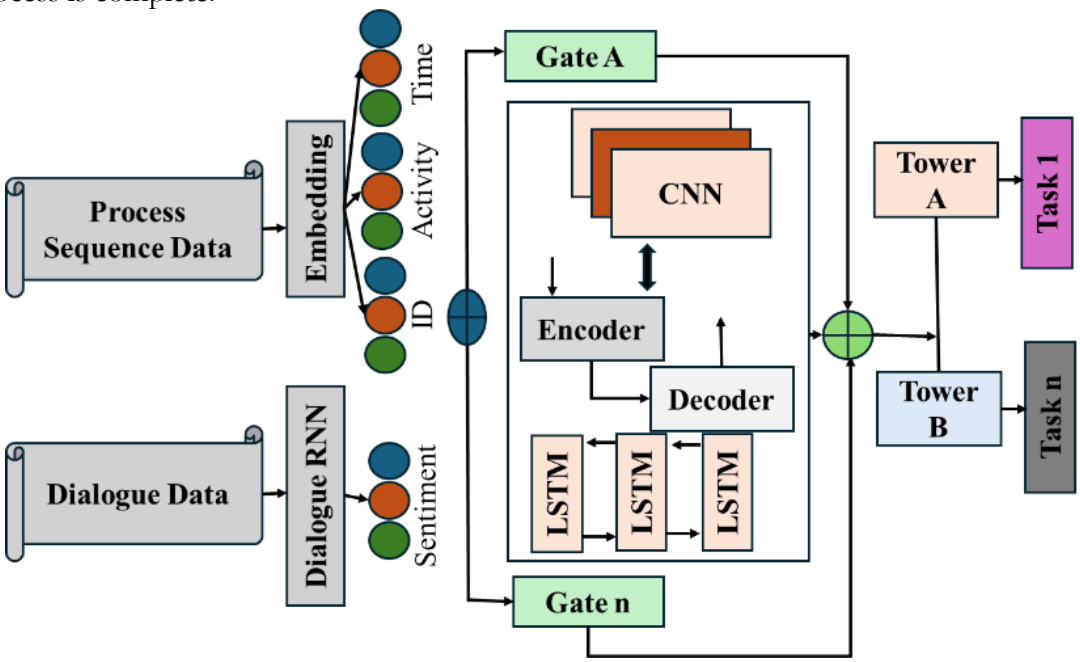

Figure 3. Proposed Methodology Framework.

A trace has three prefixes after the first, second, and third occurrences. This prefix log, which has three sections, is used to train and evaluate the predictive model. Before creating the prefix log, event data is partitioned into training and testing sets to avoid prefix overlap. This divide ensures a thorough model evaluation and reduces data leakage. The prefix log is created by processing all event log instances. Events are recalled progressively for each instance, creating incomplete traces. The outcome (final activity) and remaining time (the difference between the current and final event timestamps) are determined for each prefix. Prefixes are systematically compiled into the prefix log for analysis. This careful approach ensures that each sample is evaluated independently, preventing any interference between forecasts and ensuring the integrity of the predictive model.

Feature Representation and Encoding:

The prefixes in the preprocessed log need to be encoded into feature vectors of a specific size before the deep learning model could be trained. In the log, the attributes are classified as either categorical (qualitative) or numeric (quantitative), and each of these different types of attributes requires a different set of encoding techniques. In order to translate input activities into sparse, high-dimensional feature vectors, one-hot encoding is in the beginning applied to category attributes. Further encoding of these vectors is performed with the help of an embedding layer to solve the sparsity and loss of relational information. Normalization of numeric timestamp attributes is accomplished using min–max scaling in order to guarantee uniformity across all values.

The activity encoding procedure employs the skip-gram algorithm, with training depending on the simultaneous occurrence of actions in traces. For a trace $L= [x1, x2 ,xn]$ of length n, the likelihood of noticing m elements surrounding the t-th action is calculated as;

$$P\left(x_{t-m},..,x_{t-1},x_{t+1},x_{t+m|}x_t\right) = \prod_{-m\le j\le m, j\neq 0} P\left(x_{t+j|}x_t\right) \quad \text{Eq (1)}$$

During training, this research aims to increase the likelihood of events surrounding specific behavior as much as possible. This is expressed as.

$$e(x) = -\sum_{-m\le j\le m, j\neq 0} \text{Log}P\left(x_{t+j|}x_t\right) \quad \text{Eq (2)}$$

The approach efficiently represents predicted tasks' activities and temporal linkages using one-hot encoding, embedding layers, and skip-gram modeling.

Analysis of Sentiments:

Capturing sentiment in call center discussions is challenging due to the interactive, brief, and contextually ambiguous nature of the conversations. To address these challenges, the DialogueRNN model is specifically designed for dialogue contexts. GRU networks underpin its Emotion Representation modules, Global State, and Party State. The Party State module tracks speaker and listener emotions. Listener and Speaker GRUs are crucial components. These elements analyze the conversation's mood and emotions. The Global State module integrates contextual information to help comprehend conversation flow. Encoding both the present and prior dialogue utterances gives a complete picture of the conversation's progression. The Emotion Representation module classifies conversation sentiment by analysing the speaker's words and context. This module pulls important emotional insights from speech, allowing the model to reliably represent sentimental changes throughout the exchange.

Dynamic Loss Weighting for Heterogeneous Multi-task Models:

Designed to improve multi-task learning for commercial process prediction, the H-MMoE (Heterogeneous Mixture of Experts) architecture. It enhances prediction efficiency and accuracy over several tasks by combining dynamic loss weighting, heterogeneous expert networks, and transformer-based gating. Important elements of this construction consist in:

Heterogeneous Expert Networks:

- **LSTM (Long Short-Term Memory)**: Perfect for sequential data it captures temporal dependencies in event traces. Using input, forget, and output gates, LSTMs build models of the connections between events across time.
- **CNN (Convolutional Neural Networks)**: Uses spatial representations derived from process data to extract local information. CNNs concentrate in spotting trends in localised data areas.
- **Transformer Networks**: Use attention processes to calculate event linkages regardless of distance. Transformers capture long-range data dependencies well.

Transformer-based Gating:

Based on multi-head attention, the gating mechanism effectively distributes information tailored for each expert network. The model becomes more flexible and able of managing complicated input interactions by dynamically changing weights for expert outputs. This lets the model excel in predictions tailored for tasks.

Dynamic Loss Weighting (UWL):

Using data distribution and performance measures for every subtask, the Uncertainty Weighting Loss (UWL) method adjusts task-specific loss weights during training. UWL dynamically changes the weights unlike stationary loss weighting techniques, hence enhancing training efficiency and lowering the need for hand-held hyperparameter adjustment. By more skillful handling of task trade-offs, this also maximizes the balance between chores. These methods used together guarantee strong generalization, diverse feature extraction, and improved prediction performance in the H-MMoE architecture. The dynamic character of the gating and loss mechanisms adds even more to the adaptability and flexibility of the model in challenging, real-world corporate process prediction activities.

The loss function is expressed as:

$$\text{Loss}\,(W, \sigma 1, \sigma 2) = \frac{1}{2\sigma_1^2} L1(W) + \frac{1}{2\sigma_2^2} L2(W) + \log\sigma 1 + \log\sigma 2 \qquad \text{Eq (3)}$$

The task-specific losses are where L 1(W) and L 2(W); learnable parameters, or $\boldsymbol{\theta}$ 1 and $\boldsymbol{\theta}$ 2, reflect the uncertainty associated with each task. Regularizing to avoid overfitting are the logarithmic terms. Through constant weight updating during training, the UWL method helps the model to adjust to shifting data distributions. This lowers computing resource use and helps the model to attain improved performance. Combining heterogeneous experts, adaptive gating, and dynamic loss weighting guarantees that the H-MMoE model can efficiently manage the challenges of multi-task learning, hence enhancing both representation and prediction capability.

Decoding of Output:

This study compares activity and time prediction tasks. With the SoftMax activation function at the output layer, activity prediction is a multi-class classification challenge. This loss function uses categorical cross-entropy to minimize the difference between expected and actual labels. Time prediction is a regression task with a fully connected output layer and the Mean Absolute Error (MAE) loss function to quantify the difference between anticipated and actual results. These specialized methods do both prediction objectives accurately and efficiently.

$$\text{Loss}(x) = -\sum_{i=1}^{c} yi * \log fi(x) \qquad \text{Eq}(4)$$

The actual label for the i-th category by yi, the total number of categories by c, the projected probability for the i-th category by $fi(x)$, all define the input trace. This method reduces the discrepancy between the expected and real labels, thereby guaranteeing accurate classification.

Dataset:

The BPIC_2016 dataset [20] was used in this research, and its statistics are shown in Table 2. This dataset contains customer behavior data collected by a Dutch government agency over eight months. It includes information from sources such as message data, call center logs, website click data, and complaint records. To prepare the dataset, cases with missing information were removed, and features such as the day of the week, time, event duration, and elapsed time were engineered. Each conversation was transformed into a "question event" and recorded under the message attribute. This dialogue data was then integrated with the process sequence data using unique case identifiers.

Table 2. Dataset Statistics

Events	Dataset Entity
Attributes of Dialogue	Messages
Events Dialogue	Question
Activities	29
Mean cases duration in days	8.3
Median cases duration in day	0.979
Mean events per case	7.749
Events	221977
Trace variants	21948
Cases	28645

Results and Discussion:

This study assessed the efficacy of the suggested approach by means of a real-world event log, therefore contrasting its performance with modern techniques. This section provides details about the dataset, experimental setup, evaluation metrics, baseline methods, and results for the two prediction tasks: predicting the outcome of the process and the remaining time.

Table 3 provides a summary of the dataset that was produced consequently, which was restructured according to timestamps and stored in XES format.

Table 3. Samples of Deployed Dataset

Case ID	Activity	Timestamp	Message
912	question	6/08/2023 13:10	What is the deadline for submitting the revenue issue form?
	question	6/08/2023 13:13	Can you tell me when I can expect to get my unemployment benefits?
	home	6/08/2023 11:10	-
	taken	6/08/2023 11:12	-
	mijn_cv	6/08/2023 11:13	-
	mijn_cv	6/08/2023 11:13	-
	mijn_beri	6/08/2023 11:25	-
	mijn_rech	6/08/2023 11:25	-
	home	6/08/2023 11:26	-
15771	question	6/08/2023 9:44	The page has a script that has been running for some time.
	question	6/08/2023 9:58	What is the deadline for submitting the revenue issue form?
	taken	6/08/2023 20:51	-
	home	6/08/2023 21:24	-
	taken	6/08/2023 21:25	-

Experimental Settings:

Implementation Framework: The proposed models were developed using TensorFlow and Keras, selected for their powerful deep learning capabilities. PM4Py, a specialized process mining library, was used to analyze event logs and manage internal sequence data representations. Dialogue preprocessing, including sentiment scoring, was carried out using the DialogueRNN framework, which effectively captures conversational context and emotional nuances.

Data Partitioning: An 80-20 temporal split was applied to divide the event log into training and test datasets. This method ensures that the test set contains instances that occur after the training set in time, thus preventing information leakage and better reflecting real-world scenarios. Two separate data samples were prepared for model training: one containing only information system sequence data, and another combining sequence data with dialogue data for enhanced multimodal analysis.

Evaluation Metric: Predictive model performance was assessed across a broad spectrum of criteria. While the mean absolute error (MAE) gauged the mistake in estimating remaining process time, accuracy and dependability of process outcome predictions were evaluated using precision (P), recall (R), and F1 Score. For both classification and regression tasks, this mix of measures guarantees a complete evaluation of the models.

Hyperparameter Optimization: To improve model performance, the hyperparameters indicated in Table 4 were optimized using a grid search method. Based on literature reviews and experimental data, key parameters like activation function, number of epochs, batch size, and learning rate were deliberately chosen. To identify the ideal values for every job, the grid search investigated many setups carefully. Since it guarantees accurate probability distributions for predictions and fits multi-class classification problems, the SoftMax activation function was selected.

Table 4. Hyperparameter of the deployed Proposed Model

Parameter	Value
Optimizer	Nadam
Layers (CNN, LSTM, Transformer)	2, 2, 1
Learning Rate	0.01
Epochs	50
Loss Functions	Precision, Recall, F1, MAE
Batch Size	128
Dropout	0.2
Activation Function	ReLU, SoftMax
Transformer heads	4

Comparative Analysis:

Table 5 presents a comparison of the performance between single-modal and multimodal data techniques across baseline models. The multimodal approach showed higher prediction accuracy, highlighting the advantage of incorporating dialogue data, unlike single-modal methods that were limited to specific tasks. The proposed model reduced the Mean Absolute Error (MAE) by up to 7.47% for remaining time prediction and improved the F1 score by 0.4% to 3.4% for process outcome prediction compared to baseline techniques.

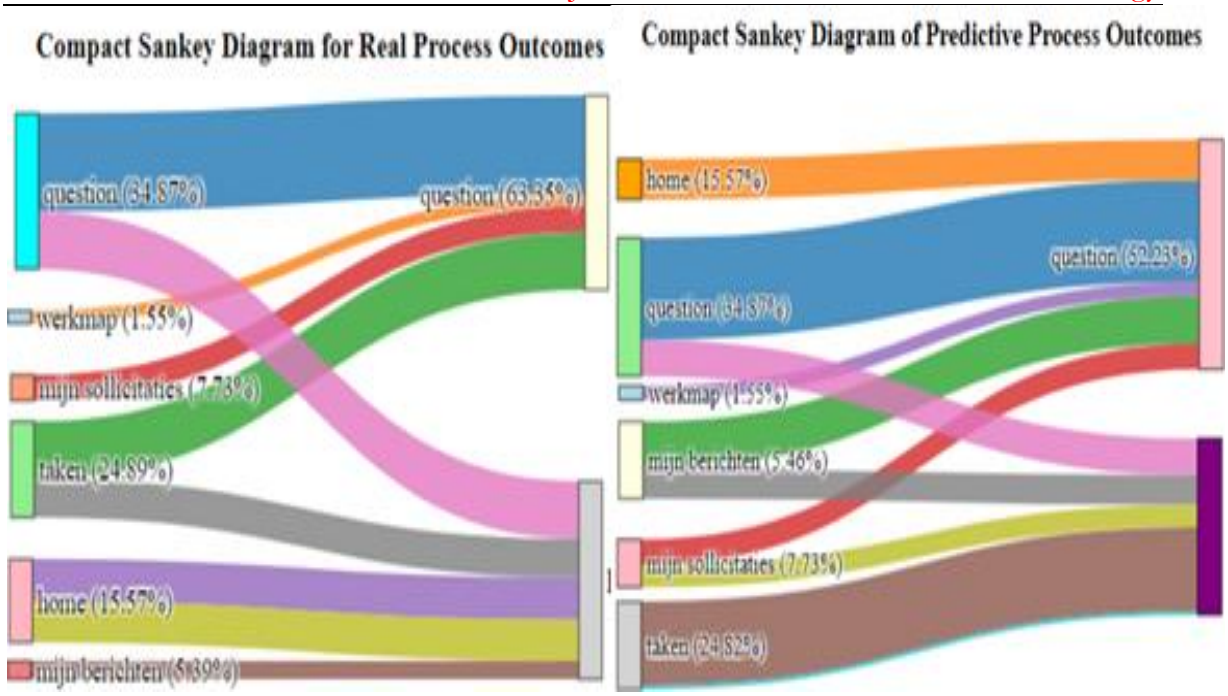

Figure 4. Visual Illustration of Proposed Model Results.

The use of LSTM and Transformer networks produced better results than CNN, highlighting the importance of capturing sequential dependencies in business process data. This study's findings demonstrate that heterogeneous multi-task learning is both robust and effective in handling a wide range of feature representations.

Table 5. Comparison with Existing Literature

Ref	Remaining Time (MAE)	Process Outcome Precision	Process Outcome Recall	F1
[26]	**Single modal**: 3.108	0.486	0.209	0.293
	Multimodal: -	-	-	-
[27]	**Single-modal**: -	0.492	0.309	0.379
	Multimodal: -	-	-	-
[28]	**Single model**: 1.553	0.718	0.716	0.717
	Multimodal: 1.553	0.719	0.754	0.736
[17]	**Single model**: 1.001	0.832	0.634	0.720
	Multimodal: 0.977	0.726	0.723	0.724
[16]	**Single model**: 1.652	0.702	0.695	0.699
	Multimodal: 1.588	0.743	0.723	0.733
[14]	**Single model**: 1.037	0.695	0.708	0.708
	Multimodal: 0.983	0.748	0.716	0.732
[4]	**Single model**: 1.044	0.725	0.708	0.716
	Multimodal: 0.966	0.738	0.745	0.742
[3]	**Single model**: 1.022	0.736	0.722	0.729
	Multimodal: 0.970	0.741	0.729	0.735

Table 5 demonstrates the superiority of our H-MMoE model, particularly in multimodal settings. For example:

- In Remaining Time prediction, our model achieved an MAE of 0.966, outperforming the baseline Transformer (1.001) and CNN (1.652) models.
- For Process Outcome prediction, H-MMoE attained a Precision of 0.738, Recall of 0.745, and F1-score of 0.742, whereas the best baseline had an F1-score of 0.720.

- These metrics reflect improvements of up to 7.47% in MAE and 3.4% in F1-score, demonstrating the value of combining heterogeneous data and dynamic task balancing.

Performance of H-MMoE:

The H-MMoE model outperformed its isomorphic MMoE counterparts. As shown in Figure 5, process outcome predictions saw F1 score improvements of 4.8% to 6.8%, while the remaining time prediction achieved MAE reductions ranging from 8.71% to 34.47%. These improvements are attributed to the dynamic loss weighting mechanism, Transformer-based gating units, and heterogeneous expert subnets, which adaptively filter and prioritize shared features. The results demonstrate that H-MMoE effectively addresses the limitations of isomorphic models by utilizing multimodal data and task-specific representations.

Ablation Study:

To evaluate the impact of individual components within H-MMoE, ablation experiments were conducted by removing key modules:

- **Without Dynamic Loss Weighting:** Replacing dynamic weighting with arithmetic means resulted in a 0.8% decrease in F1 and a 3.94% increase in MAE for process outcome and remaining time predictions, respectively.
- **Without Transformer Gate:** Using fully connected layers instead of Transformer gates led to a 1.6% decrease in F1 and a 2.48% increase in MAE.
- **Without Heterogeneous Experts:** Replacing heterogeneous subnets with isomorphic subnets caused a 1.9% decrease in F1 and a 2.59% increase in MAE.

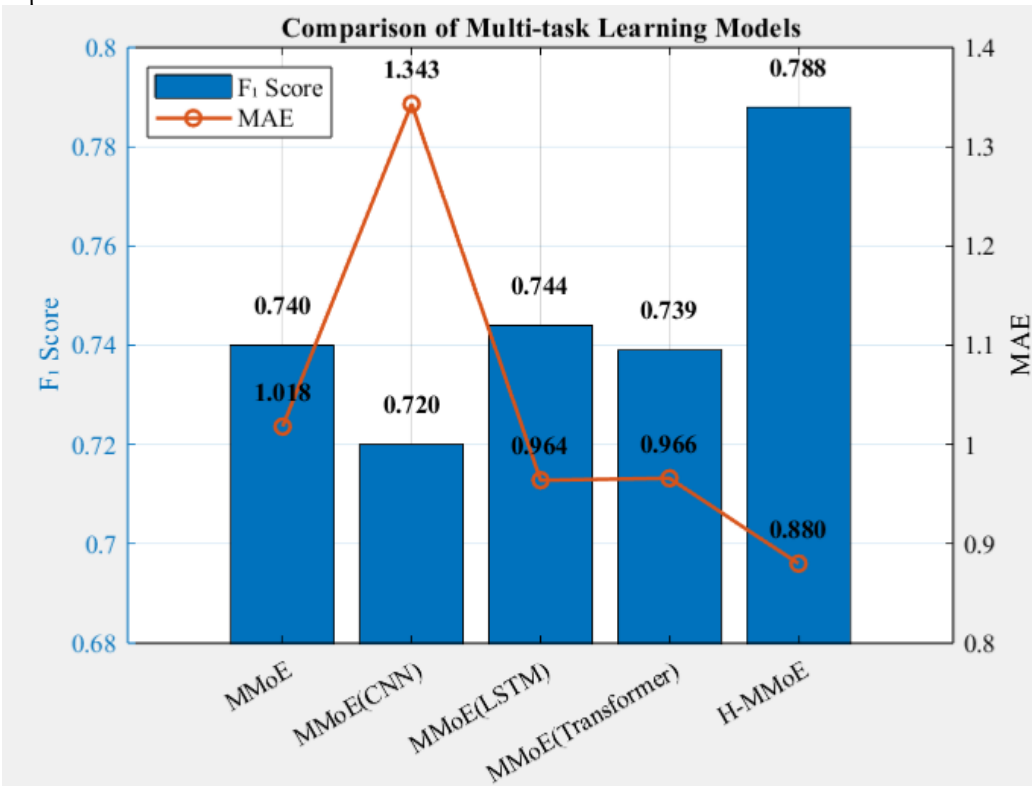

Figure 5. Multitask learning Results comparison.

Influence of DialogueRNN:

Figures 6 and 7 show the results of comparing the impact of DialogueRNN on sentiment analysis with that of BERT, BiLSTM, and TextCNN. With an F1 score of 0.735 and an MAE of 0.983, DialogueRNN outperformed all other models. Other models, including TextCNN, showed notable performance drops, with F1 decreasing by 3.3% and MAE increasing by 4.57%. In sentiment analysis for call center scenarios, DialogueRNN's superior

performance is evident, aided by its ability to track participant states and retain contextual discussion information.

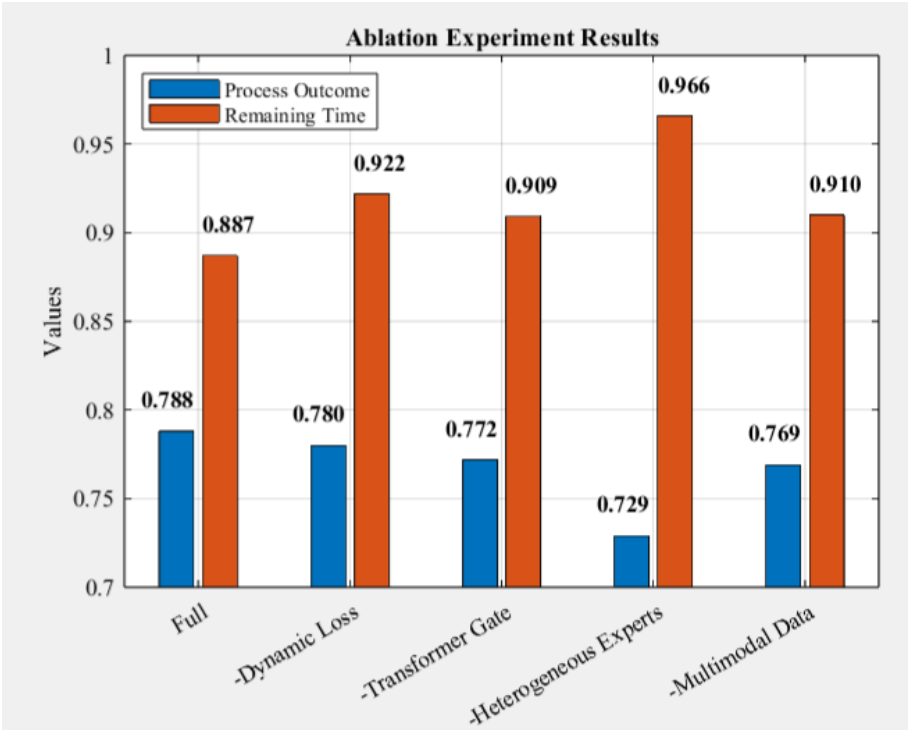

Figure 6. Ablation Experimental Results.

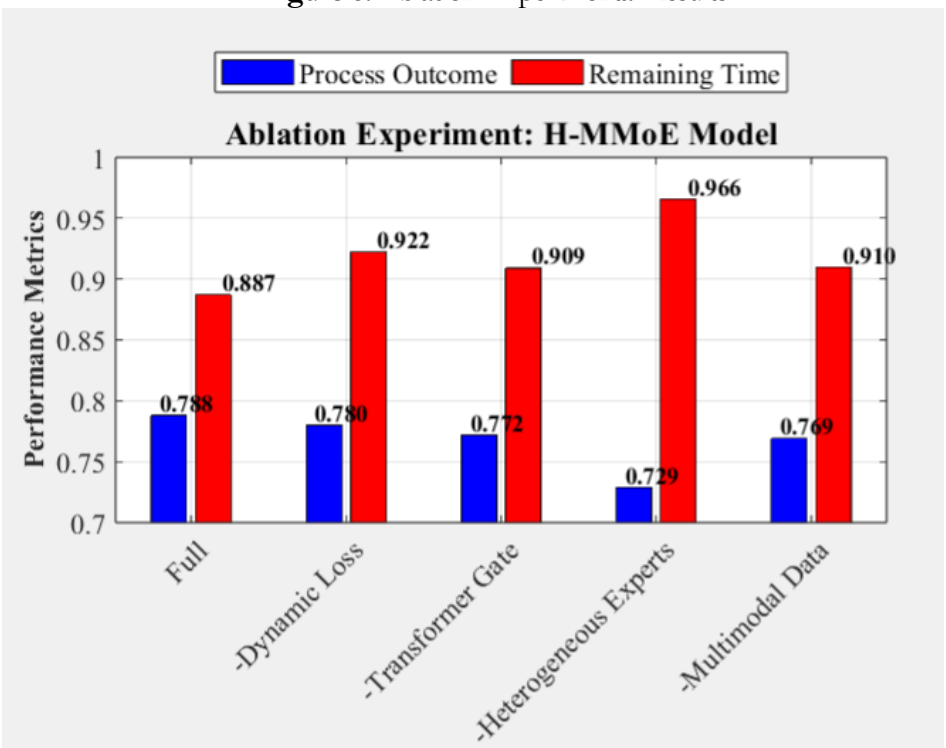

Figure 7. H-MMoE Model Ablation Experimental Result

Figure 6 shows the results of the ablation experiment comparing the performance of the full H-MMoE model with its variants (e.g., without dynamic loss, transformer gate, heterogeneous experts, and multimodal data) for predicting process outcomes and remaining time. It highlights the effectiveness of the complete model and demonstrates how each component contributed to improved predictions. The performance measures of the H-MMoE model in both its whole form

and ablated variations are shown in figure 7. Whereas the red bars reflect the MAE for estimating remaining time, the blue bars show the F1 scores for estimating process outcomes.

Discussion:

Comparative Approaches:

The developed H-MMoE model was compared to multiple state-of-the-art baseline business process prediction methodologies to assess its efficacy. These methods use specialized machine learning and deep learning techniques. Process models with temporal data were proposed in [26] to forecast process completion times. The author in [27] used probabilistic modeling to analyze the behavior of ongoing processes based on historical data. A researcher in [29] introduced a process monitoring technique based on stationarity assumptions, enabling predictions in stable process environments. In [16], running cases were transformed into spatial data representations, and Convolutional Neural Networks (CNNs) were applied for activity prediction. The researcher in [14], used Long Short-Term Memory (LSTM) networks to predict future activities and their durations by converting process traces into feature vectors. The author in [17] applied Transformer models to capture long-range dependencies in event logs, effectively modeling time-series data. In [4], a multi-task learning approach was developed using Transformer networks to predict event attributes and remaining time through shared feature representations. A researcher in [3], proposed a hierarchical Transformer model that combines information at different granularities using weighted representations, leading to improved predictions of next activities and remaining time. Finally, reference [28], employed BERT with transfer learning to build a multi-task prediction framework for forecasting activities and process outcomes. By benchmarking against these various methods, the proposed model demonstrates its superiority by integrating multimodal data and employing an advanced multi-task learning architecture.

The challenge of forecasting business process outcomes was analyzed using a Sankey diagram to evaluate the effectiveness of the proposed model. Figure 4 compares the actual and predicted results, highlighting two process outcomes: "question" and "home." In the diagram, nodes represent process activities, and the connecting lines show the flow and case counts between activities. The algorithm performed better for behaviors like "question," mostly because dialogue data improved prediction. However, "taken" predictions exhibited bigger errors because to the more diversified data distribution, illustrating the difficulty of this activity.

The experimental results indicate that the proposed H-MMoE model outperforms traditional models such as CNN, LSTM, and Transformer when applied to predictive business process monitoring. For example, in comparison to the Transformer-based baseline [17], which achieved an F1-score of 0.720 and MAE of 1.001, our H-MMoE achieved an F1-score of 0.742 and a reduced MAE of 0.966. This improvement is largely due to the integration of multimodal data and dynamic task weighting. Similarly, compared to [14] (LSTM-based), which reported an F1 of 0.708 and MAE of 1.037, our model improved results by ~3.4% in F1 and ~7% in MAE. These comparisons demonstrate the superiority of our architecture in capturing both temporal dependencies and sentiment context. The ablation study further validated that removing components like dynamic loss weighting or the heterogeneous experts led to noticeable performance degradation, proving their critical contribution. Unlike single-modal or static-task frameworks in [3], and [16], H-MMoE supports adaptive learning from multiple data streams, which is more aligned with real-world operational dynamics in call centers.

Conclusion:

A new predictive approach for call center business process monitoring utilising the H-MMoE model is presented in this study. We overcome the limits of single-modality approaches by merging corporate process sequence data with customer dialogue data. The suggested H-MMoE architecture outperformed LSTM, CNN, and Transformer in process result and duration prediction. These findings show that multimodal data and a field-specific framework improve prediction accuracy and understanding of complicated relationships. Dynamic loss weighting and Transformer-

based gating units helped H-MMoE handle task unpredictability and improve shared feature extraction. The model performed better with dialogue data, especially for sentiment-sensitive tasks like "question." The study also found forecasting issues with different data distributions, indicating further improvements. The findings suggest that AI-driven methods could transform business operations and decision-making. To improve prediction, future studies may include audio elements like silent periods and sound pressure changes. The platform could also contain real-time predictions and adaptive learning, providing unique workflow optimization potential in dynamic business situations. Integrating advanced deep learning frameworks from the medical domain highlights the versatility and potential of multimodal models to enhance predictive accuracy across diverse fields [30], [31], [32].

References:

1. M. Abbasi *et al.*, "A Review of AI and Machine Learning Contribution in Predictive Business Process Management (Process Enhancement and Process Improvement Approaches)," *Bus. Process Manag. J.*, vol. ahead-of-print, no. ahead-of-print, Jul. 2024, doi: 10.1108/BPMJ-07-2024-0555/FULL/XML.
2. L. X. Ruixuan Zheng, Yanping Baoa, Lihua Zhaob, "Prediction of steelmaking process variables using K-medoids and a time-aware LSTM network," *Heliyon*, vol. 10, no. 12, p. e32901, 2024.
3. S. B. Mohammad Al Olaimat, "TA-RNN: an attention-based time-aware recurrent neural network architecture for electronic health records," *Bioinformatics*, vol. 40, no. 1, pp. i169–i179, 2024, doi: https://doi.org/10.1093/bioinformatics/btae264.
4. Z. G. Ying Liu, Cai Xu, Long Chen, Meng Yan, Wei Zhao, "TABLE: Time-aware Balanced Multi-view Learning for stock ranking," *Knowledge-Based Syst.*, vol. 303, p. 112424, 2024, doi: https://doi.org/10.1016/j.knosys.2024.112424.
5. J. Cui, J. Ji, T. Zhang, Q. Ni, L. Cao, and Z. Chen, "A novel sparse Gaussian process regression with time-aware spatiotemporal kernel for remaining useful life prediction and uncertainty quantification of bearings," *Struct. Heal. Monit.*, 2024, doi: 10.1177/14759217241282876;PAGE:STRING:ARTICLE/CHAPTER.
6. L. Deng *et al.*, "Time-Aware Attention-Based Transformer (TAAT) for Cloud Computing System Failure Prediction," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 4906–4917, Aug. 2024, doi: 10.1145/3637528.3671547/SUPPL_FILE/TAAT-.
7. M. W. and P. X. X. Liu, Y. Liu, "LMGD: Log-Metric Combined Microservice Anomaly Detection Through Graph-Based Deep Learning," *IEEE Access*, vol. 12, pp. 186510–186519, 2024, doi: 10.1109/ACCESS.2024.3481676.
8. E.-P. L. Gary Ang, "Temporal Implicit Multimodal Networks for Investment and Risk Management," *ACM Trans. Intell. Syst. Technol.*, vol. 15, no. 2, pp. 1–25, 2024, doi: https://doi.org/10.1145/3643855.
9. H. Z. Chang Zong, "Stock Movement Prediction with Multimodal Stable Fusion via Gated Cross-Attention Mechanism," *arXiv:2406.06594*, 2024, doi: https://doi.org/10.48550/arXiv.2406.06594.
10. S. W. Wuzhida Bao, Yuting Cao, Yin Yang, Hangjun Che, Junjian Huang, "Data-driven stock forecasting models based on neural networks: A review," *Inf. Fusion*, vol. 113, p. 102616, 2025, doi: https://doi.org/10.1016/j.inffus.2024.102616.
11. Y. C. Xuan, Yang, Pang, Chengxin, Yu, Haimeng, Zeng, Xinhua, "An Enhanced Bidirectional Transformer Model with Temporal-Aware Self-Attention for Short-Term Load Forecasting," *IEEE Access*, vol. 12, pp. 75625–75639, 2024, doi: 10.1109/access.2024.3373801.
12. P. Wang, X. Zhang, Z. Cao, and Z. Chen, "MADMM: microservice system anomaly detection via multi-modal data and multi-feature extraction," *Neural Comput. Appl.*, vol. 36, no. 25, pp. 15739–15757, Sep. 2024, doi: 10.1007/S00521-024-09918-1/METRICS.

13. E. A. M. & M. S. M. Nurul Athirah Nasarudin, Fatma Al Jasmi, Richard O. Sinnott, Nazar Zaki, Hany Al Ashwal, "A review of deep learning models and online healthcare databases for electronic health records and their use for health prediction," *Artif. Intell. Rev.*, vol. 57, no. 249, 2024, doi: https://doi.org/10.1007/s10462-024-10876-2.
14. C. S.-R. Sergio Martínez-Agüero, Antonio G. Marques, Inmaculada Mora-Jiménez, Joaquín Alvárez-Rodríguez, "Multimodal Interpretable Data-Driven Models for Early Prediction of Antimicrobial Multidrug Resistance Using Multivariate Time-Series," *arXiv:2402.06295*, 2024, doi: https://doi.org/10.48550/arXiv.2402.06295.
15. B. A. M.-H. Ulises Manuel Ramirez-Alcocer, Edgar Tello-Leal, Gerardo Romero, "A Deep Learning Approach for Predictive Healthcare Process Monitoring," *Information*, vol. 14, no. 9, p. 508, 2023, doi: https://doi.org/10.3390/info14090508.
16. X. X. Weijian Ni, Gang Zhao, Tong Liu, Qingtian Zeng, "Predictive Business Process Monitoring Approach Based on Hierarchical Transformer," *Electronics*, vol. 12, no. 6, p. 1273, 2023, doi: https://doi.org/10.3390/electronics12061273.
17. Y. W. and D. Y. J. Wang, J. Huang, X. Ma, Z. Li, "MTLFormer: Multi-Task Learning Guided Transformer Network for Business Process Prediction," *IEEE Access*, vol. 11, pp. 76722–76738, 2023, doi: 10.1109/ACCESS.2023.3298305.
18. C. Liu, Y. Wang, L. Wen, J. Cheng, L. Cheng, and Q. Zeng, "Discovering Hierarchical Multi-Instance Business Processes from Event Logs," *IEEE Trans. Serv. Comput.*, vol. 17, no. 1, pp. 142–155, Jan. 2024, doi: 10.1109/TSC.2023.3335360.
19. S. Y. and M. Z. G. Duan, "A Hybrid Neural Network Model for Sentiment Analysis of Financial Texts Using Topic Extraction, Pre-Trained Model, and Enhanced Attention Mechanism Methods," *IEEE Access*, vol. 12, pp. 98207–98224, 2024, doi: 10.1109/ACCESS.2024.3429150.
20. F. A. D. Suraj, S. Dinesh, R. Balaji, P. Deepika, "Deciphering Product Review Sentiments Using BERT and TensorFlow," FMDB Transactions on Sustainable Computing Systems. Accessed: Apr. 29, 2025. [Online]. Available: https://www.fmdbpub.com/uploads/articles/170054179782534. FTSCS-55-2023.pdf
21. C. Y. Wen jun Gu, Yi hao Zhong, Shi zun Li, Chang song Wei, Li ting Dong, Zhuo yue Wang, "Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis," *ICCBDC '24 Proc. 2024 8th Int. Conf. Cloud Big Data Comput.*, pp. 67–72, 2024, doi: https://doi.org/10.1145/3694860.3694870.
22. S.-H. K. and H.-J. Y. A. -Q. Duong, N. -H. Ho, S. Pant, S. Kim, "Residual Relation-Aware Attention Deep Graph-Recurrent Model for Emotion Recognition in Conversation," *IEEE Access*, vol. 12, pp. 2349–2360, 2024, doi: 10.1109/ACCESS.2023.3348518.
23. Y. Zhao, P. Barnaghi, and H. Haddadi, "Multimodal Federated Learning on IoT Data," *Proc. - 7th ACM/IEEE Conf. Internet Things Des. Implementation, IoTDI 2022*, pp. 43–54, 2022, doi: 10.1109/IOTDI54339.2022.00011.
24. Y. Y. G. and A. A. I. U. Haq, M. Ahmed, M. Assam, "Unveiling the Future of Oral Squamous Cell Carcinoma Diagnosis: An Innovative Hybrid AI Approach for Accurate Histopathological Image Analysis," *IEEE Access*, vol. 11, pp. 118281–118290, 2023, doi: 10.1109/ACCESS.2023.3326152.
25. Y. Wang, Y. Wang, C. Shi, L. Cheng, H. Li, and X. Li, "An Edge 3D CNN Accelerator for Low-Power Activity Recognition," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 40, no. 5, pp. 918–930, May 2021, doi: 10.1109/TCAD.2020.3011042.
26. M. P. Negin Ashrafi, Armin Abdollahi, Greg Placencia, "Effect of a Process Mining based Pre-processing Step in Prediction of the Critical Health Outcomes," *arXiv:2407.02821*, 2024, doi: https://doi.org/10.48550/arXiv.2407.02821.
27. Jasmin Praful Bharadiya, "Machine Learning and AI in Business Intelligence: Trends and Opportunities," *Int. J. Comput.*, vol. 48, no. 1, pp. 123–134, 2023, [Online]. Available:

https://ijcjournal.org/index.php/InternationalJournalOfComputer/article/view/2087
28. H. F. Hang Chen, Xianwen Fang, "Multi-task prediction method of business process based on BERT and Transfer Learning," *Knowledge-Based Syst.*, vol. 254, p. 109603, 2022, doi: https://doi.org/10.1016/j.knosys.2022.109603.
29. N. Mehdiyev, J. Evermann, and P. Fettke, "A Novel Business Process Prediction Model Using a Deep Learning Method," *Bus. Inf. Syst. Eng.*, vol. 62, no. 2, pp. 143–157, Apr. 2020, doi: 10.1007/S12599-018-0551-3/METRICS.
30. IU Haq, IA Khan, G Husnain, SAF Jaffery, "Advancing breast cancer detection: Enhancing YOLOv5 network for accurate classification in mammogram images," IEEE Access, vol. 12, pp. 16474-16488, 2024.
31. HK Alkahtani, IU Haq, YY Ghadi, N Nnadi, M Alajmi, M Nurbapa, "Precision Diagnosis: An Automated Method for Detecting Congenital Heart Diseases in Children from Phonocardiogram Signals Employing Deep Neural Network," IEEE Access, vol. 12, pp. 106361-106373, 2024.
32. IU Haq, IA Khan, G Husnain, U Sadique, "An intelligent approach for blood cell detection employing Faster RCNN," Pakistan Journal of Engineering and Technology, vol. 6(1), pp. 1-6, 2023.