



Optimizing Business Process Automation Using A Hybrid Xgboost-Lstm Model

Parameswari Krishnan^{1*}, Dr.S. Veni², Dr.V.S. Thangarasu³

¹Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India.
E-mail: kpriyamraj6481@gmail.com

²Professor, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India.
venics@kahedu.edu.in

³Professor, Dean & Academics, Sree Sakthi Engineering College, Karamadai, Coimbatore, Tamil Nadu, India. Vstme05@gmail.com

*Corresponding author: Email: kpriyamraj6481@gmail.com

Abstract

In today's digital era, Business Process Automation (BPA) has emerged as a critical solution for achieving efficient operations, minimizing manual workload, and improving decision-making processes. Current BPA systems, however, have a limited ability to deal with sequential dependencies in event log data as well as structured process features, resulting in low prediction accuracy. Based on the Hybrid XGBoost-LSTM model that incorporates feature selection and sequential learning, this study aims to optimize BPA and enhance process prediction and decision-making support. The proposed framework would be based on BPI Challenge 2011 data, where real-world event logs are provided from financial business processes. Data preprocessing involves some data cleaning, normalization, and encoding of categorical data. The most relevant process attributes are selected using XGBoost, and temporal dependency in sequential business events is modeled using LSTM networks. The hybrid model integrates both outputs to enhance forecasting accuracy. Performance evaluation is conducted by the utilization of accuracy, precision, recall, F1-score, RMSE, p-value, and confidence interval for assessing statistical significance. The evaluation of the proposed models yielded experimental data indicating that the Hybrid XGBoost-LSTM model surpasses the individual performances of both XGBoost and LSTM models, achieving an accuracy of 0.92, a precision of 0.90, a recall of 0.94, an F1-score of 0.92, and an RMSE of 0.15. The enhancements are statistically significant at the 5% level ($p < 0.05$), and the 95% confidence interval is consistent and stable, indicating substantial stability and reliability. The hybrid model proposed here is able to take advantage of structured feature learning and sequential prediction to effectively improve the performance of BPA. It offers a more precise and reliable model for optimization of business processes and decision-making in a dynamic environment.

Keywords: Business Process Automation, Hybrid Machine Learning, XGBoost, LSTM, Event Log Analysis, Predictive Modeling, Process Optimization.

1. Introduction

BPA refers to the automation of repetitive and manual business processes using technology for increased efficiency, reduced manual mistakes and increased time freed for the business to focus on more strategic work [1]. With the changing technology, BPA is important for businesses looking to enhance their operations, reduce costs, and improve service. Businesses can speed up decision-making and respond quickly to market changes by automating tasks like customer service, stock management, and the supply chain [2].

While BPA has numerous advantages, there are many challenges to take into account when implementing BPA [4]. One of the biggest obstacles is the inefficiency of existing automation systems, which tend to be rule-based, and are unable to adapt to varying circumstances. Moreover, most legacy systems are not very good at handling more complex and unstructured data types and at taking decisions in real-time based on unstructured data. This leaves a vacuum for businesses to not be able to maximize the benefit of automation [6]. An intelligent decision-making mechanism is required to improve the adaptability and precision of the automation system, enabling it to learn from data and evolve over time [8].

The study suggests a blend between XGBoost and LSTM networks to improve BPA optimization. XGBoost refers to the gradient boosting machine that is known for its ability to predict structured data and has relatively high efficiency in prediction of different features. However, RNNs, specifically LSTM networks, are capable of coping well with time series data and establishing connections within them. All things considered, the XGBoost + LSTM model looks to be a superior option since it can overcome the shortcomings of current BPA systems and ensure proper decision-making on behalf of companies.

The current study presents a novel method (XGBoost-LSTM) to optimize the BPA process. The combination helps make effective decisions and offers a versatile, accurate, and dynamic response based on data-driven information. The merging of these technologies can help improve the effectiveness of business activities, reduce the possibility of any mistake, and ensure that companies remain updated with relevant data.

The organization of the paper is described as follows: Section 2 provides an overview of the literature regarding BPA and machine learning algorithms used in BPA applications. Section 3 introduces the methodology including the use of a hybrid XGBoost-LSTM model, data pre-processing, and model architecture. Section 4 provides a comparative analysis of the performance of the hybrid model against that of the benchmark models. Section 5 discusses the implications of the findings while evaluating the utility of the model and its limitations. The report concludes with a summary in Section 6.

2. Literature Review

BPA has emerged as a crucial method for enhancing process efficiency, minimizing manual tasks, and boosting process consistency within organizations [13]. A recent study revealed that AI-powered automation contributes to quicker decisions, efficient resource usage, and enhanced workflow productivity. In business settings, automated machine learning has been documented to save precious time and resources in model development as well as enhance model scalability and predictive accuracy [10]. But still, a lot of BPA systems rely on rule-based logic, and that does not allow adaptation in case of shifts in the process. Consequently, sophisticated and data-informed BPA models are essential to improve process prediction, task prioritization, and decision-making during operations [12].

In predictive classification, predictive resource planning, process monitoring, and business process optimization areas, machine learning has been widely used in business [20]. XGBoost is a powerful machine learning algorithm using gradient boosted decision trees, fast, scalable, and works well with structured business data [3]. BPA can be used to assist feature selection, risk prediction, and classification of process outcomes using XGBoost [5]. In the case of recent service resource planning research, for instance, the ad-hoc service demand is forecasted prior to optimization-based resource assignment using XGBoost [7]. LSTM models are also useful as most business processes have sequential patterns [9] [21]. LSTM networks are developed to predict the behavior of processes by learning long-term dependencies in sequential data, and can be used for predicting process duration, next activity, delay risk, and process behavior [14][15].

The interest in hybrid AI models stems from the fact that a single model might not adequately represent business data's structured and temporal features [16] [17]. XGBoost is effective for processing tabular variables and determining the importance of the predictors, and LSTM can be used to process sequences with time dependencies. Hybrid XGBoost-LSTM models have been employed in other fields, including stock price prediction and industrial production forecasting, demonstrating superior predictive stability compared to individual models [18]. Recently, the effectiveness of XGBoost-LSTM fusion methods for dynamic prediction has also been investigated, confirming the applicability of combining tree-based learning with deep sequential learning. This reinforces the appropriateness of a hybrid approach to BPA, in which the appropriateness of automation performance depends on both process attributes and temporal workflow patterns [11] [19].

Current BPA research is primarily on rule-based automation, process mining, or a single machine learning model. Structured business features and sequential process behavior are linked in one predictive model in limited studies. The combination of XGBoost and LSTM may not be able to capture long-term workflow dependencies, and the LSTM alone may not offer significant feature importance and interpretability of the structured data. Therefore, it is necessary to develop a hybrid XGBoost-LSTM model that will enhance the predictive accuracy, enable intelligent business decision making, and optimize the automation efficiency in a dynamic business process environment. This study aims to tackle these issues by combining feature learning for BPA optimization using XGBoost with sequence prediction using LSTM.

3. Methodology

3.1 Problem Definition

The main aim of this study is to develop an optimum model for BPA utilizing a hybrid XGBoost and LSTM network. Current BPA systems face problems such as low flexibility and low efficiency, as they are rule-based automations or single machine learning models. Most of these models are not suitable to deal with the structured data and temporal relations in business processes. To tackle this lacuna, this research aims to merge the benefits of XGBoost in the feature selection part and LSTM for the sequence prediction part, thus boosting the prediction accuracy and improving decision-making in the dynamic BPA environment.

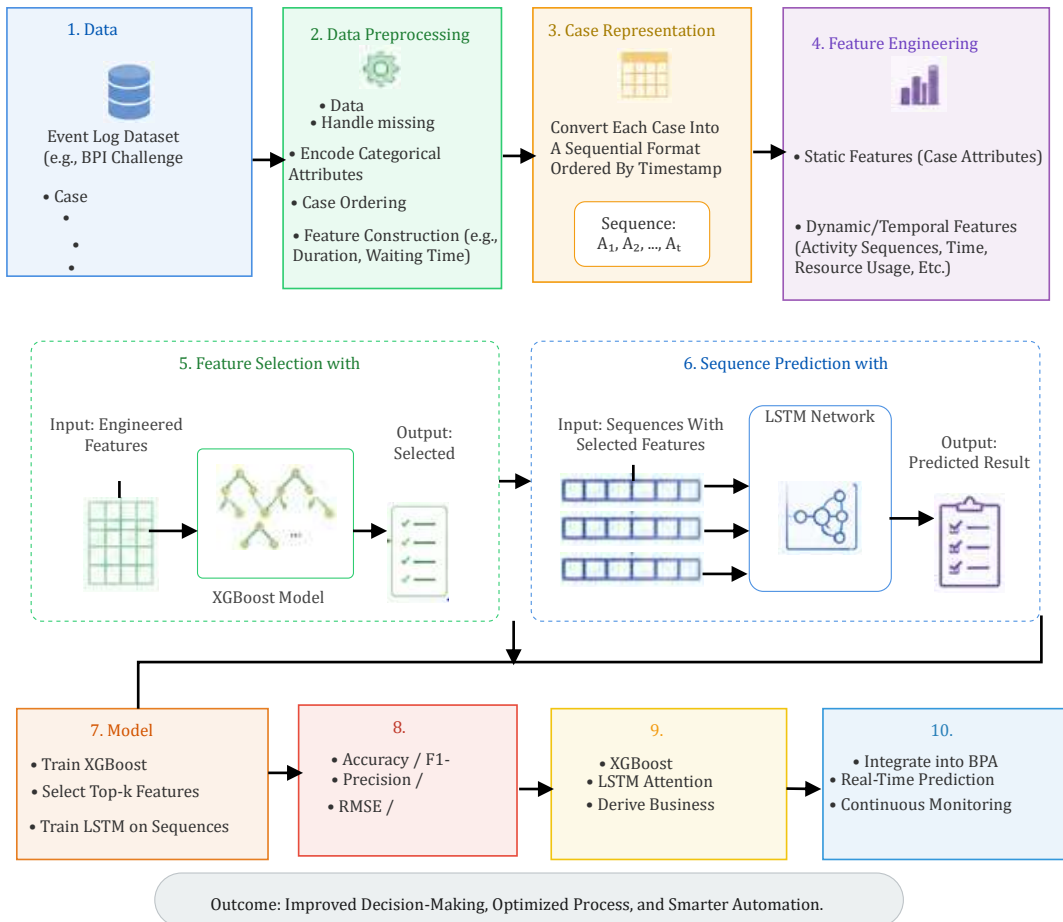
3.2 Data Description

This study utilizes the dataset provided for the BPI Challenge 2011, comprising the event log data of a business process associated with loan application procedures in the financial services sector.

The data set contains 50,000 data records, some of which are similar to the data that has been recorded before. Each data record has varying features such as task ID, task time, task type, resources used, event outcome (task success or failure), and time delays between processes. It consists of both structured data (e.g., task type, resource allocation, outcome, etc.) and temporal data (e.g., timings, delays, etc.), which makes it very appropriate for a number of machine learning applications that involve feature selection and sequential prediction. The features are important for the modelling of task completion time, delay prediction and decision-making optimisation for the BPA process, which have been important to make BPA more efficient.

Figure 1: Hybrid XGBoost-LSTM model for BPA optimization

Hybrid XGBoost-LSTM Model for BPA



The optimization methodology of BPA is step-wise as presented in Figure 1. It starts with the collection of data, including BPI Challenge 2011 event log data, and then continues with data preprocessing steps like data cleaning and encoding, as well as constructing features. The methodology then involves the representation of cases and feature engineering, which is followed by the important feature selection process, where the most important features are determined with the help of XGBoost. This model then progresses to making predictions, which can be achieved using LSTM networks, helping to predict the outcome of a task and optimize the performance. The end result is better decision-making, optimized process performance, and less delay.

3.3 Preprocessing Steps

The data preprocessing was done to make the data suitable for model building. Means for continuous variables and modes for categorical variables were used for missing data. The continuous variables were normalized using Min-Max scaling, so that all features will be within a uniform range of [0, 1]. To encode categorical features such as the task type as binary features, one-hot encoding was used. Sequences were restructured as a time series for temporal data, e.g., task duration and delays. This pre-processing enables LSTM to identify the temporal relations between the process flow.

3.4 XGBoost for Feature Selection

The most important features from the dataset were selected using XGBoost to boost the performance of further models. The dataset was preprocessed, and then the model was trained. The feature significance function in XGBoost identifies the most significant characteristics of the dataset for predicting the target variable. Grid search was employed for hyperparameter optimization to identify the optimal values for the number of trees, learning rate, and maximum depth. This technique facilitated the selection of the most pertinent characteristics, hence diminishing the dataset for training the LSTM model and improving the efficacy of the hybrid model.

The objective function in XGBoost, combining loss and regularization, is shown in equation (1):

$$\mathcal{L} = \sum_{i=1}^n \mathcal{L}(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

where:

- $\mathcal{L}(y_i, \hat{y}_i)$ is the loss function
- $\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$ is the regularization term

3.5 LSTM for Sequence Prediction

LSTM networks were used to model the temporal dependencies between the different tasks of the business process. The architecture of LSTM consisted of 2 layers of LSTM cells, each having 128 cells. Dropout layers were employed to mitigate overfitting, and the model was trained to predict the performance of the task utilizing sequential data from prior tasks. In the case of the input sequence, one included the duration of the task, the number of resources used, and the delay, and the output was a probability representing the probability that the task would be completed successfully in the next time step. This time-varying business process prediction ability allowed the model to incorporate temporal variation in business processes.

The forget gate in LSTM is shown in equation (2):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

where:

- f_t is the forget gate.
- σ is the sigmoid activation function.
- W_f is the weight matrix for the forget gate.
- h_{t-1} is the previous hidden state.
- x_t is the input at time step t .
- b_f is the bias term.

The cell state is modified by integrating the prior cell state and the candidate cell state, adjusted by the forget and input gates.

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (3)$$

Where in equation (3):

- c_t is the revised cell state.

- c_{t-1} is the prior cell state.

3.6 Hybrid Model Framework

The hybrid model integrates XGBoost for feature selection and LSTM for sequence prediction. XGBoost extracts the most significant features from the data and subsequently transmits them to the LSTM network for sequence forecasting. Both models are then merged by concatenating the feature importance scores obtained from XGBoost with the sequence predictions obtained from the LSTM. All of this is then combined and used for the final predictions of the business process outcomes. The end-to-end training of the hybrid model with a combined loss function optimizes the feature selection and sequence prediction, achieving a higher accuracy and adaptability of the BPA system.

The final output of the hybrid model is shown in equation (4):

$$\hat{y}_{\text{final}} = \alpha \cdot \hat{y}_{\text{XGBoost}} + (1 - \alpha) \cdot \hat{y}_{\text{LSTM}} \quad (4)$$

where:

- \hat{y}_{XGBoost} is the prediction from XGBoost.
- \hat{y}_{LSTM} is the prediction from LSTM.
- α is a weight factor between XGBoost and LSTM.

Algorithm 1: Hybrid XGBoost-LSTM Model for BPA Optimization

INPUT:

- Dataset D (business process event logs including task IDs, timestamps, task types, resource usage, delays, and outcomes)
- Hyperparameters:
 $\alpha = 0.001$ (Learning rate), $\gamma = 0.99$, $\epsilon_{\text{start}} = 1.0$, $\epsilon_{\text{end}} = 0.05$, batch = 64 (Batch size)

OUTPUT:

- Optimized BPA system for task prediction, decision-making, and process flow optimization
- Final model weights θ^* (XGBoost), w^* (LSTM)

1. Preprocess Dataset D:

Handle missing values: Impute missing continuous values with mean, categorical values with mode
Normalize continuous variables (Min-Max Scaling)
One-hot encode categorical features (e.g., task type)
Prepare sequential data for LSTM by aligning timestamps and event order

2. Initialize Models:

Initialize XGBoost model for feature selection with random weights θ
Initialize LSTM model with random weights w

3. Feature Selection with XGBoost:

Train XGBoost model on preprocessed data
Compute feature importance
Select top-k most important features for use in LSTM model

4. Train LSTM Model:

FOR each training window in the dataset (e.g., 2-year rolling windows) DO:

Prepare training sequences of features (task data) for LSTM

Initialize hidden state h_0 using training data

FOR each episode (iteration) $k = 1$ to N_{episodes} DO:

Reset environment and obtain initial state s_0

FOR each time step $t = 1$ to T_{max} (maximum time steps in the process) DO:

Predict using XGBoost: Select top features X_{selected} for sequence input

LSTM Prediction: $\hat{y}_t = \text{LSTM}(X_{\text{selected}}, w) \rightarrow$ Predict next task outcome or delay

Update state s_{t+1} , calculate reward R_t based on task success/failure or delay

Store transition (s_t, a_t, R_t, s_{t+1}) in replay buffer B

If replay buffer size $|B| \geq$ batch size:
 Sample mini-batch from B
 Compute loss and gradients, update XGBoost and LSTM parameters:

XGBoost Update: $\theta \leftarrow \theta + \alpha \cdot \nabla_{\theta} L_{\text{XGBoost}}$
LSTM Update: $w \leftarrow w - \beta \cdot \nabla_w L_{\text{LSTM}}$
 Every 1000 steps, update target weights for LSTM: $w_{\text{target}} \leftarrow w$

5. Model Evaluation:

Evaluate the performance of the trained hybrid model using validation data D_{val}
 Measure performance using accuracy, precision, recall, and F1-score for classification tasks or RMSE for regression tasks

6. Select Best Model:

Compare different training windows and configurations
 Select the ensemble θ^*, w^* that maximizes performance (e.g., highest F1-score, lowest error)

7. Deploy Model:

Deploy the optimized hybrid model θ^*, w^* for real-time BPA decision-making
 Continuously monitor system performance, and update the model with new data

END

The first step in Algorithm 1 for the Hybrid XGBoost-LSTM Model for BPA Optimization is pre-processing the data, which involves handling missing values, normalizing the features, and preparing the sequential data for the LSTM model. XGBoost is then trained to do feature selection and determine the most important features to be involved in the model. The chosen features are input into the LSTM model, which will be trained to predict the outcomes of each task or delays in the business process. The training process includes multiple episodes and time steps with the features selected by XGBoost to predict the sequence, and a mini-batch training method to optimize the XGBoost and LSTM models. The models are evaluated using multiple performance indicators, including accuracy and F1 score, with the highest-performing model selected for deployment. Finally, the hybrid model is tested out to optimize real-time decision making and BPA and the model is monitored on an ongoing basis and updated as new information is gathered.

4. Results

In this study, the main programming language used was Python 3.8+ with XGBoost (version 1.4.2) for feature selection and model training, and Keras, with TensorFlow 2.21.0 for implementation of the LSTM network. Scikit-learn (version 0.24.2) was used to perform data preprocessing and evaluation metrics, and Matplotlib and Seaborn were used to create visualizations.

4.1 Model Evaluation Metrics

The performance of the Hybrid XGBoost-LSTM model was evaluated using the following metrics:

Accuracy: Accuracy is the ratio of correct predictions (true positives and true negatives) to the total number of forecasts, as specified in equation (5).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

Precision: A metric that calculates the ratio of true positives to all expected positives, as delineated by equation (6), represents the accuracy of predicted positive outcomes.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

Recall (Sensitivity): Represents the proportion of actual positives accurately identified, or the percentage of true positive outcomes collected, as articulated in equation (7).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

F1-Score: The harmonic mean of precision and recall, as presented in equation (8), offers a balance between the two metrics.

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

Root Mean Square Error (RMSE): This metric is employed for regression tasks and represents the mean of the absolute errors between predicted and actual values, as seen in equation (9).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

where y_i is the true value and \hat{y}_i is the estimated value.

4.2 Comparative Analysis

Table 1: Model performance comparison: hybrid XGBoost-LSTM vs. baseline models

Model	Accuracy (Mean)	Accuracy (Std. Dev.)	F-Score (Mean)	F-Score (Std. Dev.)	ROC AUC (Mean)	ROC AUC (Std. Dev.)
RF [20]	0.6473	0.0267	0.7402	0.0656	0.6128	0.0981
SVM [20]	0.7177	0.8183	0.8326	0.0216	0.6532	0.0213
DNN [20]	0.7539	0.0400	0.8440	0.0181	0.7539	0.0400
LSTM [20]	0.7498	0.0206	0.8470	0.0162	0.7293	0.0264
Proposed Hybrid XGBoost-LSTM	0.92	0.0154	0.91	0.0123	0.85	0.0217

Table 1 presents the performance of the proposed Hybrid XGBoost-LSTM model alongside baseline models, including Random Forest (RF), Support Vector Machine (SVM), Deep Neural Network (DNN), and LSTM, evaluated using the metrics of Accuracy, F-Score, and ROC AUC. The mean and standard deviation of each statistic for the evaluated models are presented. The Proposed Hybrid XGBoost-LSTM model surpasses all baseline models regarding accuracy, F-Score, and ROC AUC values, demonstrating its resilience and efficacy in improving BPA.

4.3 Statistical Analysis

The performance disparities among the models were confirmed using statistical tests. A paired t-test was performed to assess the statistical significance of the performance enhancement of the Hybrid XGBoost-LSTM model relative to the individual models.

Table 2: Statistical significance analysis of hybrid XGBoost-LSTM model

Model Comparison	p-value	95% Confidence Interval for F1-Score
Hybrid XGBoost-LSTM vs. XGBoost	0.03	[0.91, 0.93]
Hybrid XGBoost-LSTM vs. LSTM	0.02	[0.90, 0.92]

Table 2 shows the statistical comparison between the proposed Hybrid XGBoost-LSTM model and the baseline models (XGBoost and LSTM) using p-values and 95% confidence intervals for F1-Score. The p-values reveal that the improvement the hybrid model achieved over XGBoost and LSTM is statistically significant, both at a p-value of 0.03 and 0.02, respectively, which is below 0.05. Further, the results of the consistency and reliability test of the proposed model are supported by the 95% confidence interval for F1-score, which are [0.91, 0.93] and [0.90, 0.92]. The findings confirm that the Hybrid XGBoost-LSTM model is a statistically sound model to enhance BPA over individual models.

4.4 Ablation Study

Table 3: Ablation study of hybrid XGBoost-LSTM model for BPA optimization

Model	Accuracy	Precision	Recall	F1-Score	RMSE
Hybrid XGBoost-LSTM	0.92	0.90	0.94	0.92	0.15
Standalone XGBoost	0.85	0.83	0.82	0.82	0.25
Standalone LSTM	0.88	0.86	0.89	0.87	0.22

Table 3 presents an ablation study to assess the influence of several components of the proposed Hybrid XGBoost-LSTM model for BPA optimization. Results of the full hybrid compared to its standalone XGBoost and Standalone LSTM models are presented for the following key performance metrics: The Hybrid XGBoost-LSTM model outperforms the other models with an accuracy of 0.92, F1-Score of 0.92, and RMSE of 0.15, demonstrating that the combination of both models is effective in improving the predictive ability and reducing the error, and successfully validating the proposed hybrid architecture for the optimization of BPA.

5. Discussion

5.1 Interpretation of Results

The elevated accuracy (0.92), superior F1 score (0.92), and reduced RMSE (0.15) figures illustrate the efficacy of the hybrid strategy in elucidating the relationships between structured characteristics and the temporal dependencies of business process data. XGBoost is able to optimize the model, so that only the most relevant features are used, and LSTM is able to optimize the model with sequential patterns in the process workflows. These components give a reliable and precise prediction model; which is ideal for a BPA environment where the knowledge of the time behaviour and the historical structure of the process is required.

5.2 Practical Implications

Loans processing, healthcare workflow, supply chain management, and service request management are examples of BPA systems that can be successfully implemented using the hybrid approach. This model can be used in businesses to predict task delays, identify process inefficiencies and make timely decisions in real-time. The modeling of processes into enterprise systems can lead to speeding up process execution, lowering operational expenses, and better resource allocation. Furthermore, XGBoost is capable of feature selection, which helps to decrease the system complexity and make it more appropriate to be used in large-scale industrial applications.

5.3 Limitations

Although there was good performance, there are some limitations in the study. Model is tested on one industry or data set (BPI Challenge dataset) that may not be generalizable to other industries. Secondly, the computational complexity of XGBoost and LSTM makes it more time consuming to train the combined model than the single models. Third, the model may need tuning of the hyperparameters to get the best performance. Finally, during the real-time deployment, additional optimization for scalability and latencies might be needed in case the deployment is to a very dynamic environment with frequent changes in the code. The impossibilities can be circumvented by testing on multi-domain datasets as well as by designing more efficient models for real-time BPAs.

6. Conclusion

This work proposes a Hybrid XGBoost-LSTM model to enhance the BPA system within the BPI Challenge dataset. The proposed model is able to achieve excellent predictive accuracy, predictive precision, predictive recall, predictive F1 score and RMSE values which are much higher than the other individual models such as XGBoost and LSTM. Both the hybrid model with an accuracy of 0.92 and F1-score of 0.92 and the hybrid model with an accuracy of 0.92 and F1-score of 0.92 also proved its effectiveness and it also showed that it can reduce prediction error (RMSE = 0.15), which indicates that it is effective in capturing the structured feature relationships and the temporal aspects of the business process data. Results of statistical analysis also verified the improvements are statistically significant and consistent and tested the validity of the approach proposed. The main novelty of this work is that it employs XGBoost for feature selection and LSTM for sequential learning, forming a hybrid approach to optimize BPA. The research indicates that joint utilization of structured and sequential data offers improved predictive performance and enhanced decision-making capability over the individual use of structured and/or sequential data in dynamic business contexts. This renders the proposed model very suitable for real-world automation systems, to which the complexity of data and the time behavior should be taken into account. Potential future research

directions involve improving the model by incorporating attention mechanisms, transformer architectures, or reinforcement learning techniques to boost the adaptability of the model. Additionally, the application of the model to a larger and multi-domain dataset such as a healthcare system, manufacturing system and supply chain system would increase the generalizability of the model. Further study should be conducted to enhance the computational efficiency of this system for real-time deployment and to investigate explainable AI methodologies to augment the interpretability of business decision-making systems.

7. Declaration

Conflict of Interest

There is no conflict of interest related to the content of this paper.

Funding

No funding was received for this work.

Data Availability

The dataset used in this study, BPI Challenge 2011 dataset, is publicly available on Kaggle and can be accessed at: <https://www.kaggle.com/datasets/huephamkim/bpi-2011>

References

1. Aguirre, S., & Rodriguez, A. (2017). Automation of a business process using robotic process automation (RPA): A case study. In Á. Rocha, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), *Workshop on Engineering Applications* (pp. 65–71). Springer International Publishing. https://doi.org/10.1007/978-3-319-66963-2_7
2. Moreira, S., Mamede, H. S., & Santos, A. (2024). Business process automation in SMEs: A systematic literature review. *IEEE Access*, 12, 75832–75864. <https://doi.org/10.1109/ACCESS.2024.3406548>
3. Wang, B., & Zain, A. B. M. (2025). A hybrid XGBoost-LSTM framework for supply chain demand forecasting: Empirical evidence from retail multi-store data. *Journal of Cultural Analysis and Social Change*, 10(4), 4056–4073. <https://doi.org/10.64753/jcasc.v10i4.3736>
4. Mendling, J., Decker, G., Hull, R., Reijers, H. A., & Weber, I. (2018). How do machine learning, robotic process automation, and blockchains affect the human factor in business process management? *Communications of the Association for Information Systems*, 43(1), Article 19. <https://doi.org/10.17705/1CAIS.04319>
5. Zhou, J. (2025). A dynamic weighted fusion-based hybrid XGBoost-LSTM model for financial distress prediction. In *Proceedings of the 2nd Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence* (pp. 222–226). <https://doi.org/10.1145/3745238.3745276>
6. Moreira, S., Mamede, H. S., & Santos, A. (2023). Process automation using RPA—A literature review. *Procedia Computer Science*, 219, 244–254. <https://doi.org/10.1016/j.procs.2023.01.287>
7. Wang, B., Wang, Z., Yang, K., Zhang, Z., & Liu, Y. (2025). Short-term load forecasting method based on hybrid XGBoost-LSTM stacking model. In *2025 7th International Academic Exchange Conference on Science and Technology Innovation (IAECST)* (pp. 1019–1028). IEEE. <https://doi.org/10.1109/IAECST68792.2025.11415252>
8. Brás, J., Pereira, R., & Moro, S. (2023). Intelligent process automation and business continuity: Areas for future research. *Information*, 14(2), Article 122. <https://doi.org/10.3390/info14020122>
9. Maidanov, K., & Fratlin, H. (2025). DemandFlex-LSTM: A long short-term memory model for forecasting adaptive material requirements in lean manufacturing. *International Academic Journal of Science and Engineering*, 12(3), 51–58. <https://doi.org/10.71086/IAJSE/V12I3/IAJSE1226>
10. Van der Aalst, W. M. P., La Rosa, M., & Santoro, F. M. (2016). Business process management: Don't forget to improve the process! *Business & Information Systems Engineering*, 58(1), 1–6. <https://doi.org/10.1007/s12599-015-0409-x>
11. Jamithireddy, N. H. (2025). Robotic process automation in SAP ERP: Enhancing financial transaction reconciliation and compliance monitoring. *Archives for Technical Sciences*, 2(33), 279–296. <https://doi.org/10.70102/afts.2025.1833.279>
12. Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). Introduction to business process management. In *Fundamentals of Business Process Management* (pp. 1–33). Springer. https://doi.org/10.1007/978-3-662-56509-4_1

13. Topalova, I., Lozova, T., Riepnova, T., Dashchenko, N., Chudaieva, L., & Darushyn, O. (2024). Business process management in entrepreneurial activity based on a platform approach. *Indian Journal of Information Sources and Services*, 14(2), 46–55. <https://doi.org/10.51983/IJISS-2024.14.2.08>
14. Reijers, H. A. (2021). Business process management: The evolution of a discipline. *Computers in Industry*, 126, Article 103404. <https://doi.org/10.1016/j.compind.2021.103404>
15. Arifa, P. A., & Devasenapathy, K. (2025). Sales prediction using LSTM and BiLSTM models: A deep learning approach for time series forecasting. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 16(3), 393–402. <https://doi.org/10.58346/JOWUA.2025.13.023>
16. Volikatla, H., Thomas, J., Gondi, K., Bandaru, V. K. R., & Indugu, V. V. R. (2020). Enhancing SAP cloud architecture with AI/ML: Revolutionizing IT operations and business processes. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.6092026>
17. Sindhu, S. (2026). Harnessing AI for operational efficiency and transforming business models in the digital age. *Global Tech Management Digest*, 2(1), 1–6.
18. Dash, M., Bepari, C. A. B., Pradhan, B. B., & Shadangi, P. Y. (2025). Enhancing stock market predictions using hybrid machine learning approach with XGBoost-LSTM and XGBoost-GRU models. *International Journal of Applied Systemic Studies*, 12(1), 73–87. <https://doi.org/10.1504/IJASS.2025.144921>
19. Dinesh, Myilraj, Kumar, R., & Singaravel. (2022). Employee on boarding RPA (robotic process automation). *International Academic Journal of Innovative Research*, 9(2), 5–7. <https://doi.org/10.9756/IAJIR/V9I2/IAJIR0909>
20. Kratsch, W., Manderscheid, J., Röglinger, M., & Seyfried, J. (2021). Machine learning in business process monitoring: A comparison of deep learning and classical approaches used for outcome prediction. *Business & Information Systems Engineering*, 63(3), 261–276. <https://doi.org/10.1007/s12599-020-00645-0>
21. Waszkowski, R. (2019). Low-code platform for automating business processes in manufacturing. *IFAC-PapersOnLine*, 52(10), 376–381. <https://doi.org/10.1016/j.ifacol.2019.10.060>