PMiner: Process Mining using Deep Autoencoder for Anomaly Detection and Reconstruction of Business Processes

Original Scientific Paper

Veluru Chinnaiah*
Department of CSE,
Vijaya Engineering College,Khammam,Telangana, India
vtchinna2k12@gmail.com

Vadlamani Veerabhadram
Department of C. S. E
CVR College of Engineering,
Hyderabad, Telangana, India
drbhadram@gmail.com

*Corresponding author

Ravi Aavula
Department of CSE - DS,
Anurag University, Hyderabad
aavularavi@gmail.com

Srinivas Aluvala
Department of Computer Science and Artificial Intelligence, SR University
srinu.aluvala@gmail.com

Abstract *– We proposed a deep learning-based process mining framework known as* ***PMiner*** *for automatic detection of anomalies in business processes. Since there are thousands of business processes in real-time applications such as e-commerce, in the presence of concurrency, they are prone to exhibit anomalies. Such anomalies if not detected and rectified, cause severe damage to businesses in the long run. Our Artificial Intelligence (AI) enabled framework* ***PMiner*** *takes business process event longs as input and detects anomalies using a deep autoencoder. The framework exploits a deep autoencoder technique which is well-known for Its ability to discriminate anomalies. We proposed an algorithm known as Intelligent Business Process Anomaly Detector (IBPAD) to realize the framework. This algorithm learns from historical data and performs encoding and decoding procedures to detect business process anomalies automatically. Our empirical results using the BPI Challenge dataset, released by the IEEE Task Force on Process Mining, revealed that* ***PMiner*** *outperforms state-of-the-art methods in detecting business process anomalies. This framework helps businesses to identify process anomalies and rectify them in time to leverage business continuity prospects.*

Keywords: *Process Mining, Artificial Intelligence, Deep Autoencoder, Long Short Term Memory, Deep Learning*

Received: October 21, 2023; Received in revised form: February 9, 2024; Accepted: February 9, 2024

1. INTRODUCTION

Enterprise business applications are very complex and involve several hundreds of business processes. Often millions of users across the globe use such applications. Each business process involved in the application can be accessed by thousands of users simultaneously. In other words, there is concurrent access to business processes. It may lead to anomalous behaviour of business processes in terms of sequence of events or temporal dimension. Detection of anomalies in business processes is a tedious and complex phenomenon [1]. To enable the discovery of business processes, business process event logs are generated. Process mining is the science of dealing with business processes and analysing them to discover potential faults in the execution of business processes [2]. Complex business processes should be understood from multiple perspectives towards discovering actionable knowledge [3]. Process mining research involves diversified activities aimed at monitoring, tracking and rectifying business processes.

Many researchers focused on process mining since it is crucial for enterprise-level businesses. Association rule learning is one of the techniques used in [1] and [4] for finding business anomalies. Machine learning approaches are widely used for process mining as discussed in [5] and [6]. Advanced neural network models or deep learning techniques are also used by researchers to leverage business processes. This kind of research includes repairing missing activities [4], process prediction [7], anomaly detection [8, 9] and outcome prediction [6]. Hybrid learning approaches are also found important for process mining as discussed in [4] and [10]. Business process anomaly classification is found significant as explored in [11] and [12]. From the literature, it is observed that process mining research focuses on different aspects. However, an integrated

approach with process discovery, anomaly detection and enhancement still requires further research. Our contributions to this paper are as follows.

1. We proposed an Artificial Intelligence (AI) enabled framework known as as ***PMiner*** which takes business process events longs as input and detects anomalies using a learning-based approach. It also has provisions for rectifying anomalies to improve the quality of business processes.
2. We proposed an algorithm known as Intelligent Business Process Anomaly Detector (IBPAD) to realize the framework. This algorithm learns from historical data and performs encoding and decoding procedures to detect business process anomalies automatically.
3. We evaluated our framework using the BPI Challenge dataset, released by the IEEE Task Force on Process Mining, which revealed that ***PMiner*** outperforms state-of-the-art methods in detecting business process anomalies. This framework helps businesses to identify process anomalies and rectify them in time to leverage business continuity prospects.

The remainder of the paper is structured as follows. Section 2 reviews existing research on process anomaly detection. Section 3 presents the proposed framework for the automatic detection of process anomalies and rectifying them. Section 4 presents the results of our empirical study. Section 5 discusses important findings in our research along with limitations. Section 6 concludes our work besides providing scope for future research.

2. RELATED WORK

This section reviews existing methods of process mining involving anomaly detection and rectification. The literature review covers research from 2013 to 2023. The rationale behind choosing older references is that they do have credible process mining research. Sungkono et al. [1] observed that ERP systems manage business processes, generating extensive logs. This study integrates process mining, fuzzy decision-making, and association rule learning to detect anomalies, enhancing fraud detection accuracy at low confidence levels. Kovalchuk et al. [2] found that deep learning, specifically LSTM models, enhances process mining for business operations. This approach combines accuracy and explainability, generating informative graphs. Stefanini et al. [3] stated that process Mining is a valuable technique for business process analysis, though its managerial potential remains underexplored. This review identifies research gaps and proposes a research agenda for its application in various business contexts. Chen et al. [13] observed that process mining bridges process modelling and data mining. To propose an LSTM-based model to repair missing activity labels in event logs, outperforming existing methods. Future work includes expanding and optimizing the approach.

Koninck et al. [14] introduced representation-learning techniques for business processes, enabling low-dimensional vectors for activities, traces, logs, and models. Applications include trace clustering and process model comparison. Future research avenues include interpretability and incorporating additional data dimensions—Joaristi et al. [15] utilized event logs for business process analysis. Existing encoding methods focus on control flow, leaving out other aspects. Deep-TRace2Vec, a deep learning approach, produces superior trace representations considering multiple perspectives. In Future, the work includes anomaly detection and transformer neural networks. Dewandono et al. [4] proposed a hybrid method combining association rule learning and process mining to improve fraud detection accuracy with fewer false discoveries compared to process mining alone. Vasumathi and Vijayakamal [16] showed that enterprise applications with Service Oriented Architecture (SOA) became complex. A framework using auto encoders improves these aspects, especially with Probabilistic Auto Encoder based Anomaly Detection (PAE-AD). Empirical results support its efficiency. Future work includes deep learning integration.

Fettke et al. [7] used process mining to reconstruct business processes from digital traces. A systematic review examines 32 methods to identify strengths, weaknesses, and research gaps. Unified benchmarks could enhance future process prediction approaches. According to Dumas et al. [17] complex business systems generate event logs that can be analysed for predictive business constraint monitoring, allowing early intervention. Implemented in ProM, validated using cancer treatment data. Further enhancements could involve different similarity measures and classification techniques for more significant accuracy. Charles et al. [18] found that organizations face challenges in detecting process abnormalities. A novel approach using conformance analysis identifies abnormalities by comparing successful and failed process instances. Fitness scores predict anomalies. Alexander et al. [19] observed that detecting subtle changes and anomalies in business processes is crucial. A neural network-based system can filter noisy event logs and detect anomalies without prior knowledge. In Future, this work includes investigating frequent anomalies and different noise levels. Neural networks are applicable and can capture underlying process patterns in event logs.

Franczyk et al. [11] proposed a semi-supervised deep learning classification model that effectively identifies anomalies in business process event sequences. It considers time dependencies and outperforms existing approaches in accuracy. In Future, we need to improve time-related anomaly detection and integrate the model into real-time environments. Flammini et al. [20] improved process mining with IoT log analytics and machine learning to detect and fix IoT anomalies, enhancing resilience. Research should address proto-

cols and error predictability. Consistent Event Logs are key to Self-Healing in IoT-based CPS. Hemmer et al. [21] used process mining to detect IoT system misbehaviours and attacks, even with heterogeneous platforms and protocols. It employs data pre-processing and clustering techniques for predictive security. Experiments demonstrate its effectiveness. Future work involves automated countermeasures and deep learning integration. Capurro et al. [22] said that process mining in healthcare analyses processes using data from information systems. A literature review examines 74 relevant papers, providing insights and guidance for future applications in healthcare.

Cristina Nicoleta [23] discussed Industry 4.0 reliability and safety, suggesting a method for real-time robotic process verification with IIoT and Celonis. It enhances quality control and cuts errors, costs, and downtime. While focusing on a synthetic robotic arm, it offers a blueprint for boosting real-time industrial automation. Vanhoof et al. [24] focused on corporate fraud, particularly internal transaction fraud, which is costly. Process mining helps detect fraud by analysing event logs. A case study confirms its benefits in mitigating internal transaction fraud, especially in auditing and compliance checking. Pauwels and Calders et al. [25] automated modelling of behaviour captured in complex log files, enabling anomaly detection and concept drift identification using extended Dynamic Bayesian Networks. Luettgen et al. [5] proposed auto encoder-based approach for detecting and interpreting anomalies in business processes, achieving an F1 score of 0.87. Gyunam et al. [26] opined that process mining extracts insights but lacks actionable improvements. This framework connects monitoring with automated actions for process enhancement, successfully tested on real systems.

Okubo and Kaiya [27] Introduced a method for enhancing security in the DevOps lifecycle, focusing on threat analysis, attack detection, vulnerability extraction, and countermeasure assessment. Tested in a development case, it proves effective. Clemente et al. [8] proposed a 5G-oriented cyber defence architecture that employs deep learning for efficient cyber threat detection and self-adaptation to network traffic fluctuations. Experiments demonstrate its effectiveness. In Future, the work includes optimizing deep learning models and real-data training. Benedi et al. [28] presented emotive process mining algorithms for analysing human behaviour patterns in ambient assisted living environments. Fathalla et al. [29] introduced a deep reinforcement learning approach for business process anomaly detection, using limited labelled data and exploring unlabelled data. The model outperforms existing methods. Lagraa [29] discussed an approach using process mining to investigate and track malicious activities in authentication events, improving defence systems against such events. Guha and Samanta [10] presented a hybrid model for anomaly detection (AD) in title insurance using autoencoders (AE) and one-class support vector machines (OSVM). This approach shows promise but requires improvements in training and data-generative techniques.

Ashok Kumar et al. [30] focused on Conformance Checking (CC) which assesses the alignment between process models and real execution. Process Mining aids analysis, validation, and improvement. Challenges include data volume, control flow focus, and tool efficiency, suggesting room for future enhancements. Kratsch et al. [6] Predictive process monitoring anticipates business process behaviour. Deep learning outperforms classical machine learning, especially with high variant-to-instance ratios and imbalanced variables. Future research should consider broader log types and develop decision models. Luettgen et al. [12] explored BINet, a neural network for real-time multi-perspective anomaly detection in business process event logs. It outperforms other methods on synthetic and real-life datasets. BINet is adaptable for autonomous operation and can handle concept drift. In future, the work may discuss issues of forgetting in repeated event sequences. Folino and Pontieri [31] stated that process mining research is extending to less structured logs from non-process-aware systems. However, interpreting deep neural networks remains challenging. Research in Explainable DL aims to address this, and informed PM methods are being developed to utilize expert guidance. From the literature, it is observed that process mining research focuses on different aspects. However, an integrated approach with process discovery, anomaly detection and enhancement still requires further research.

3. PROPOSED FRAMEWORK

This section presents a proposed framework and the underlying methodology for the automatic detection of business process anomalies and solving the problem.

3.1. PROBLEM DEFINITION

Provided a set of business processes in the form of event logs, developing a process mining framework using deep autoencoder for automatic detection and rectification of anomalies is the challenging problem considered.

3.2. OUR FRAMEWORK

We proposed a deep learning-based process mining framework known as PMiner for the automatic detection of anomalies in business processes. Since there are thousands of business processes in real-time applications such as e-commerce, in the presence of concurrency, they are prone to exhibit anomalies. PMiner with its underlying mechanisms helps in the detection of anomalies and solves them automatically. PMiner is illustrated in terms of its anomaly detection in Fig. 1 and the reconstruction process in Fig. 2. We used the BPI

Challenge 2020 dataset collected from [32]. This dataset provides real-life event logs for research. However, the data was anonymized to preserve privacy. This section, later, illustrates an excerpt from the dataset while discussing the proposed methodology. Notations used in the proposed system are provided in Table 1.

Notation	Meaning
g_{φ}	Denotes encoder
f_{θ}	Denotes decoder
x^i	Original input
$f_{\theta}(g_{\varphi}(x^i))$	Reconstructed input
$P(x)$	Probability of input x
$P(x\|y)$	Denotes conditional probability
$P(y\|x)$	Denotes posterior probability
$P(y)$	Denotes prior probability
$P(x\|y)/P(y)$	Denotes likelihood ratio

Table 1. Notations used in the proposed system

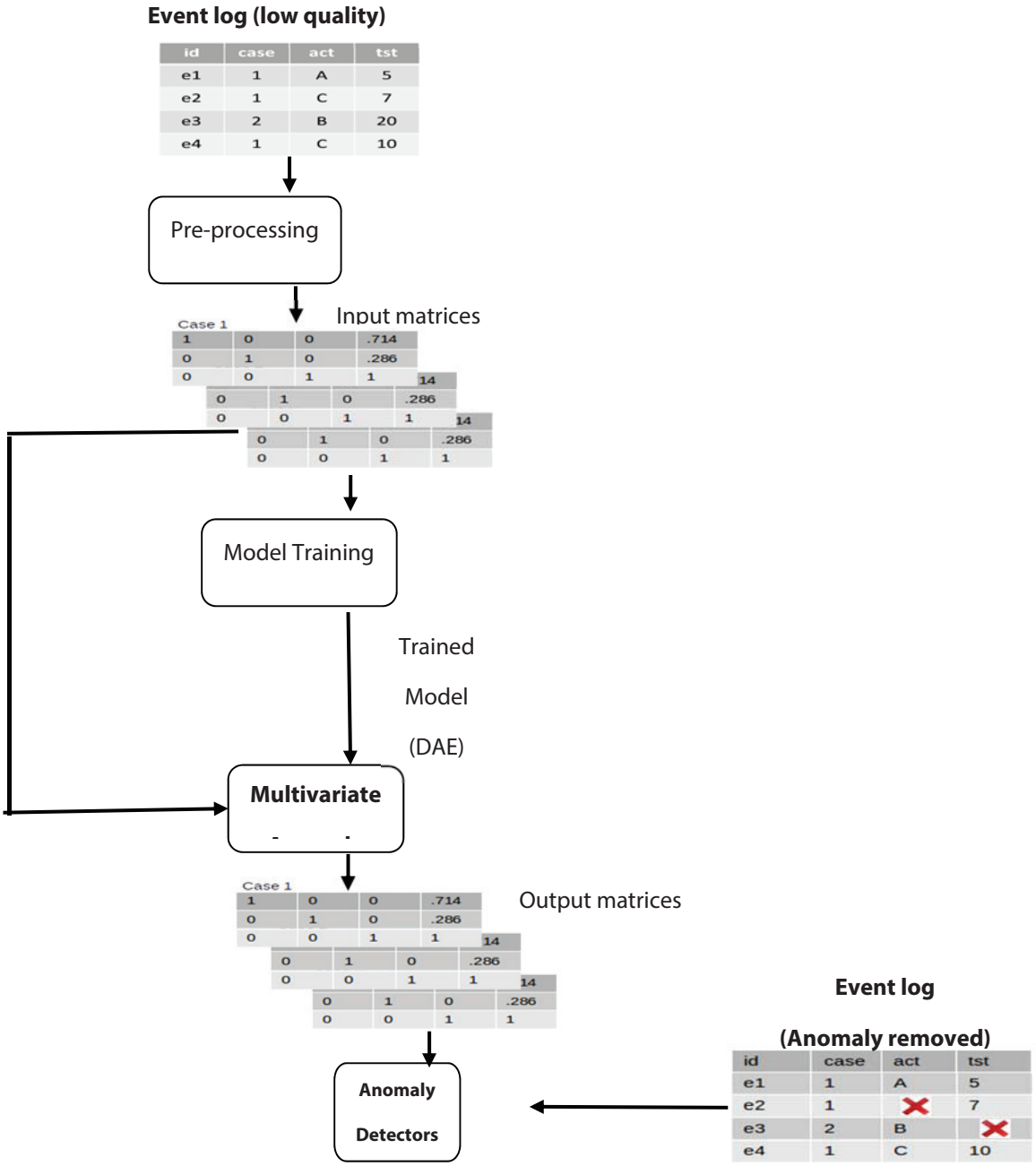

Fig. 1. PMiner framework reflecting process anomaly detection process

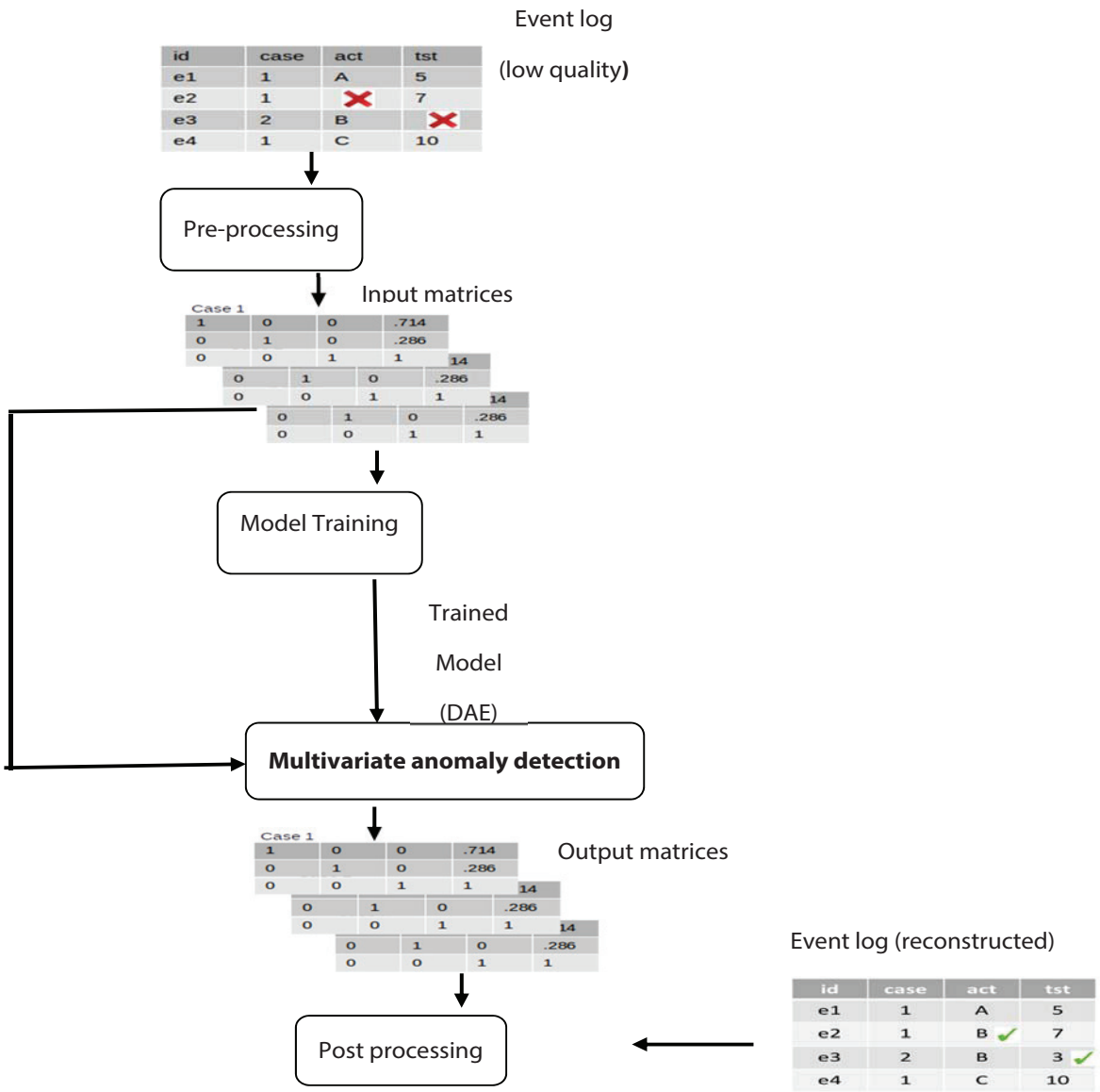

Fig. 2. PMiner framework reflecting process anomaly rectification process

PMiner takes event log data as input. The event log is a text file containing log entries reflecting a set of cases represented as $L \in$. Each case contains several events and attributes. The presence of a value and absence of value for a given attribute are denoted as $\#_a(c)$ and $\#_a(c) = \perp$ respectively. A sequence of events in the given trace or case is denoted as $\#_{trace}(c) \in \varepsilon^*$.

An event in the log entry is an activity involved in a process. In a given case there are several events denoted as e $\in \varepsilon$. The activity attribute associated with data is $ddetedas\#_{act.}(e) \in A$. Similarly, $\#_{time}(e) \in T$ denotes the $ttimestampattributeOther$ attributes such as cost, resource and transaction are denoted as $\#cost(e)$, $\#_{resourse}(e)$ and $\#transe(e)$ respectively. As shown in Fig. 1, the given dataset is subjected to pre-processing. Table 2 shows an excerpt from the event log.

The data presented in Table 1 is subjected to attribute standardization where event ID and case attributes contain the identity of the event and case respectively.

The rest of the two columns do have discrete and continuous values. The normalization process has resulted in Table 3.

Id	Case	Act	Test
e1	1	A	5
e2	1	B	7
e3	2	B	3
e4	1	C	10

Table 2. An excerpt from the event log dataset

Id	Case	C_A	C_B	C_C	C_{tst}
e1	1	1	0	0	-0.42
e2	1	0	1	0	0.25
e3	2	0	1	0	-1.09
e4	1	0	0	1	1.26

Table 2. An excerpt from the event log dataset

After completion of processing, input matrices are generated. These matrices are used to train deep autoencoders as part of the encoding process. In the decoding process, the deep autoencoder generates output matrices. These outputs enable the framework to derive two kinds of anomaly detectors. They are generated based on activity and time. The selection criterion for these two is that the anomaly is generally based on inconsistency in activity or time in which events occur. This is the rationale for generating those two types of anomaly detectors. Detection of these two kinds of anomalies is very important for owners of businesses that make use of an enterprise application that relies on several business processes. These anomaly detectors are used by the framework to detect anomalies and remove them as illustrated in Fig. 1.

The anomaly rectification process of PMiner takes the output of the process illustrated in Figure 1. This output containing log entries with events where anomalies are removed is subjected to pre-processing. As in the anomaly detection phase, pre-processing generates input matrices and a deep autoencoder model is trained with those matrices. Then the trained model is used to generate output matrices that help in the reconstruction of log entries in the form of post-processing. Fig. 3 shows the learning process resulting in labelling through reconstruction error and finally detecting anomalies.

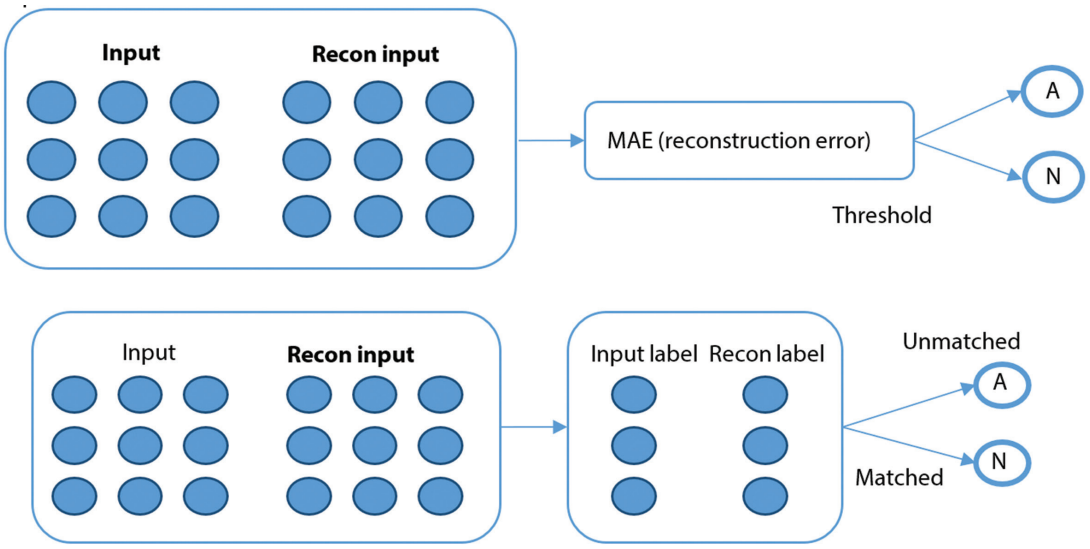

Fig. 3. Outlines the learning process involved in PMiner

In each phase of PMiner, there are deep encoding and decoding procedures involved as illustrated in Figure 4.

The deep autoencoder maps inputs to a distribution, in terms of two vectors such as mean and standard deviation, instead of fixed vector.

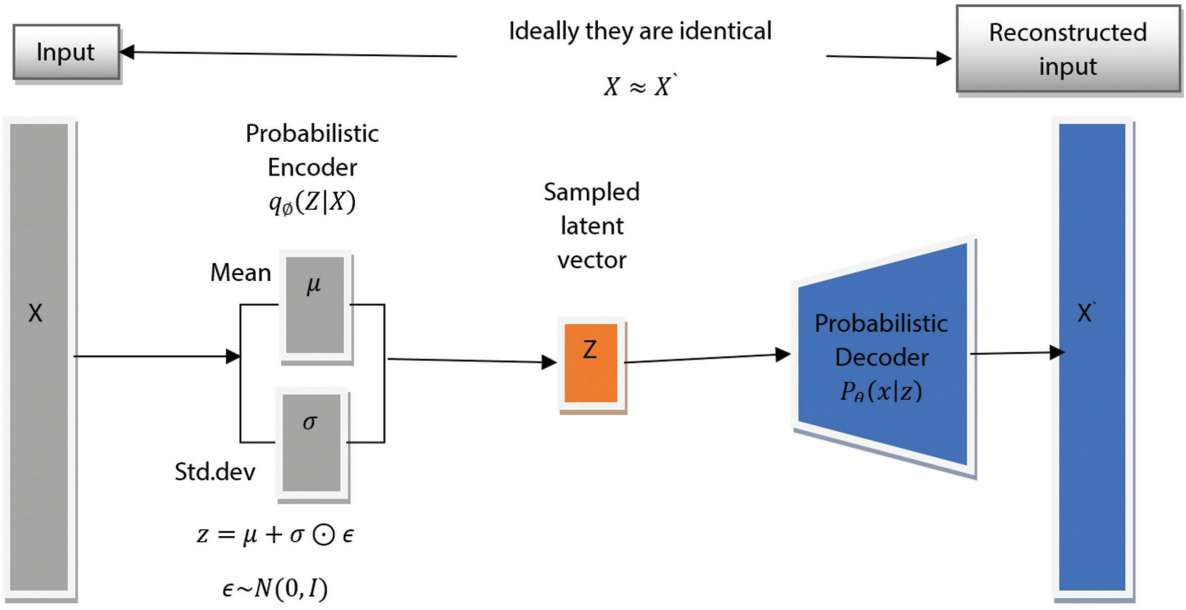

Fig. 4. Deep autoencoder used in the PMiner framework

The encoder and decoder functionalities in the autoencoder help in realizing anomalies in the business processes. The input X is mapped to mean vector μ and standard deviation vector σ. The encoding process results in the compressed nature of sampled latent vector z. The loss function associated with the autoencoder has two terms such as reconstruction loss and regularizer as expressed in Eq. 1.

$$L(\theta,\varphi)=-E_{z\sim q_\theta}[P_\theta(x|z)]+D_{KL}(q_\varphi(z|x)||p_\theta(z)) \quad (1)$$

The autoencoder functions based on probability theory. Given a random variable x, its probability is defined as $P(x)$ and its conditional probability is denoted as $P(x|y)$. Therefore, the probability theory can be expressed as in Eq. 2.

$$P(y|x)=\frac{P(x|y)P(y)}{P(x)} \quad (2)$$

This theory is based on the well-known Baye's theorem where the likelihood ratio is denoted as $p(x|y)$ / $p(x)$, prior probability is denoted as p(y) while posterior probability is denoted as $P(y|x)$. Then theorem of total probability is expressed as in Eq. 3.

$$P(x)=\sum_{i-1}^{n}P(x|y)P(y) \quad (3)$$

Given input variable x, the expected value associated with the random variable is weighted as per the probability of the event. Therefore, $E(x)$ of a random variable is computed as in Eq. 4.

$$E(x)=\sum x_i p(x=x_i) \quad (4)$$

3.3. ALGORITHM DESIGN

We proposed an algorithm known as Intelligent Business Process Anomaly Detector (IBPAD) to realize the framework. This algorithm learns from historical data and performs encoding and decoding procedures to detect business process anomalies automatically.

Algorithm 3: Intelligent Business Process Anomaly Detector (IBPAD)

Input: Event logs $L=\{e1, e2, \dots en\}$ for training

Output: $L_{(x,\hat{x})}$//reconstruction error

$\varphi, \theta \leftarrow$ network parameter initialization

repeat

$X^M \leftarrow$ obtain random points containing data points

$\epsilon \leftarrow$ nnoisebasedrandom samples $p(\epsilon)$

; g $\leftarrow \nabla_{\theta,\varphi}\tilde{L}^M)(\theta,\varphi; X^M,\epsilon)$//gradients

$\varphi, \theta \leftarrow$ parameter update

until parameter convergence (φ, θ)

$\varphi, \theta \leftarrow$ trained parameters

$\alpha \leftarrow$ threshold as per training data

repeat

for i=1 to N do

Compute $L(x,\hat{x})$

$L_{(\varphi,\theta;xi)}=\sum_i \|x_i-g_\theta(f_\varphi(x_i))\|^2$

if $L(x,\hat{x})>\alpha$ then

x_i is considered anomaly

else

x_i has no anomaly

end if

end for

Algorithm 1. Intelligent Business Process Anomaly Detector (IBPAD)

Algorithm 1 takes event log entries as input and detects anomalies through deep autoencoder based approach. It has training process where the algorithm gains knowledge which is then used in the anomaly detection process. Provided $L=\{e_1, e_2, \dots e_n\}$ as input, the algorithm eventually results in $L(x,\hat{x})$. Since event logs reflect activities of a business process that occur in temporal order, the proposed methodology and underlying algorithm learn from the huge data associated with business processes and finds anomalies. Once anomalies are detected, it is possible to rectify them from the knowledge gained in the process of detecting abnormality. The proposed system considers two kinds of anomalies such as time related and also activity related anomalies.

4. EXPERIMENTAL RESULTS

We implemented the proposed framework PMiner using Python language and process mining library. Anaconda distribution is used for building prototype. Environment used for the implementation is a PC with i3-1215U processor, 8GB RAM and Windows 11 operating system. BPI challenge 2020 dataset [32] is used in our empirical study. The dataset is freely available for usage by researchers. This section presents experimental results along with performance evaluation.

4.1. EXPLORATORY DATA ANALYSIS

This section presents data distribution dynamics in the data collected from [32]. The data is analysed in terms of anomalous data and normal data.

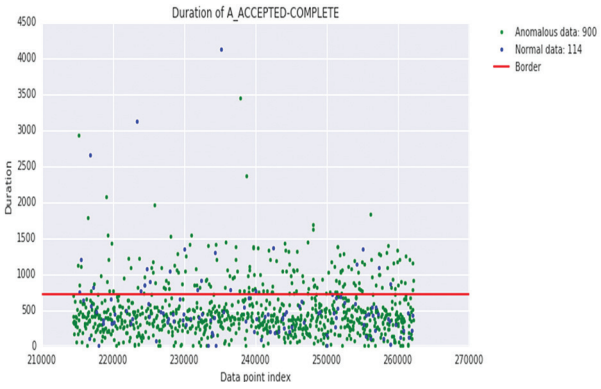

Fig. 5. Data distribution dynamics of A_ACCEPTED-COMPLETE attribute

As presented in Fig. 5, data point index against duration are visualized reflecting number of normal data points (114) and number of anomalous data points (900) distributed in the dataset.

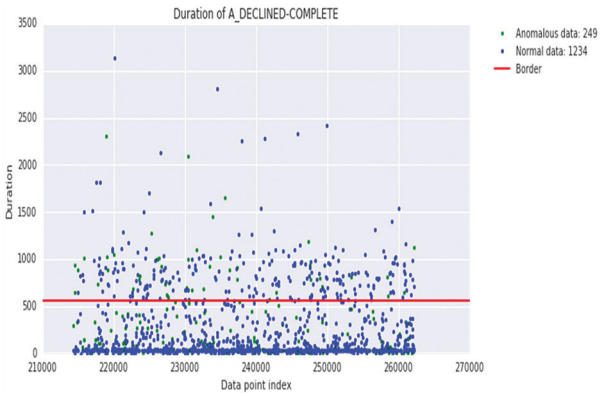

Fig. 6. Data distribution dynamics of A_DECLINED-COMPLETE attribute

As presented in Fig. 6, the data point indexes against duration are visualized reflecting number of normal data points (1234) and number of anomalous data points (249) distributed in the dataset.

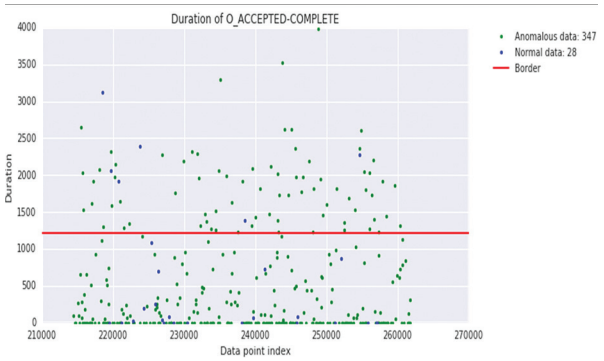

Fig. 7. Data distribution dynamics of O_DECLINED-COMPLETE attribute

As presented in Fig. 7, data point index against duration are visualized reflecting number of normal data points (28) and number of anomalous data points (347) distributed in the dataset.

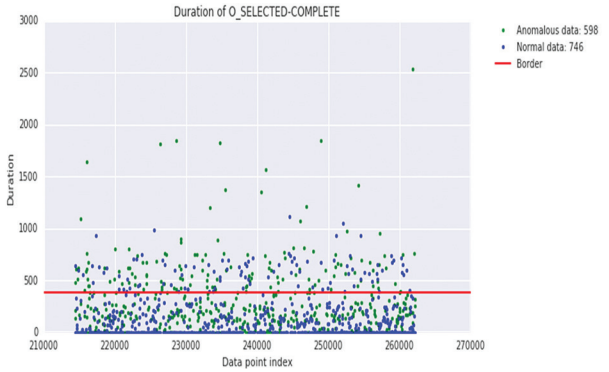

Fig. 8. Data distribution dynamics of O_SELECTED-COMPLETE attribute

As presented in Fig. 8, data point index against duration are visualized reflecting number of normal data points (746) and number of anomalous data points (598) distributed in the dataset.

4.2. TIME BASED ANOMALY DETECTION

This section presents time based anomaly detection results using the proposed PMiner framework. It covers reconstruction error, confusion matrix and AUC.

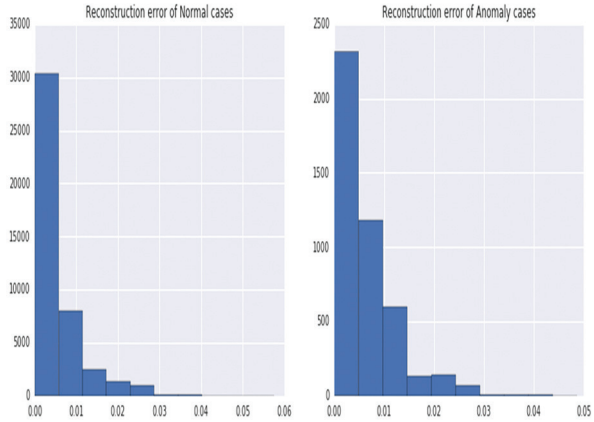

Fig. 9. Reconstruction error for normal and anomaly classes pertaining to time based anomalies

As presented in Fig. 9, it shows reconstruction error for normal class and also anomaly class. The proposed methodology has tested the entire dataset and the confusion matrix reflecting its detection process is presented in Fig. 10.

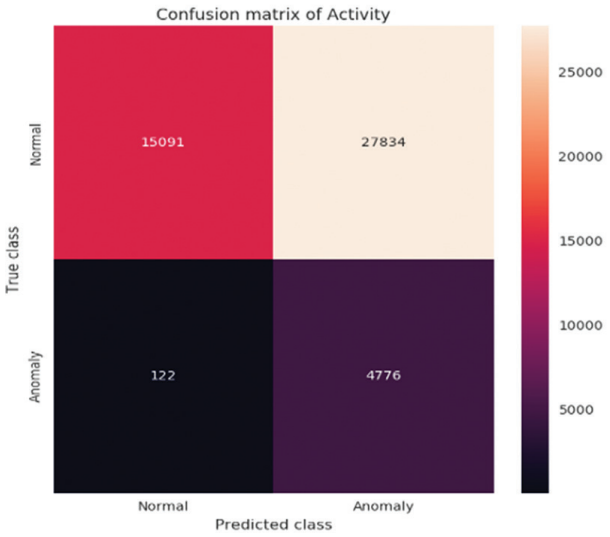

Fig. 10. Confusion matrix for time based anomaly detection

The confusion matrix visualizes the summary of results containing ground truth and also prediction results. It shows 4776 true positives, 15091 true negatives, 122 false positives and 27834 false negatives. Fig. 11 shows the AUC curve reflecting the performance of the proposed system.

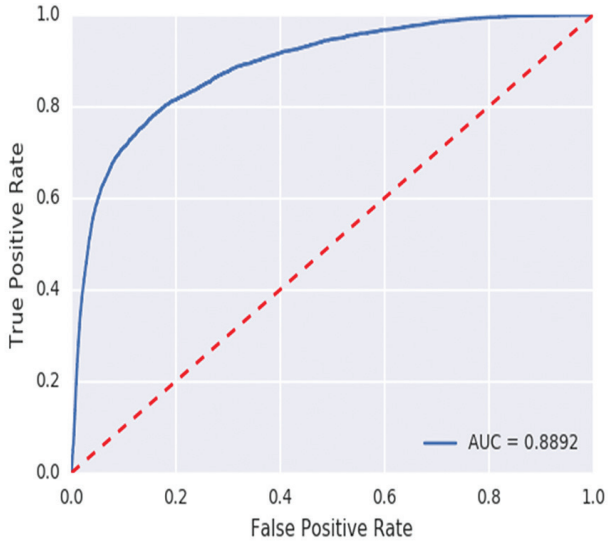

Fig. 11. AUC performance of the proposed system for time based anomaly detection

Area Under Curve (AUC) measure is used to assess the performance of the proposed system. AUC curve is computed as in Eq. 5.

$$AUC = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \quad (6)$$

AUC of the proposed system for time based anomaly detection is 0.8892. Higher in AUC indicates better performance.

4.3. ACTIVITY BASED ANOMALY DETECTION

This section presents activity-based anomaly detection results using the proposed PMiner framework. It covers reconstruction error, confusion matrix and AUC.

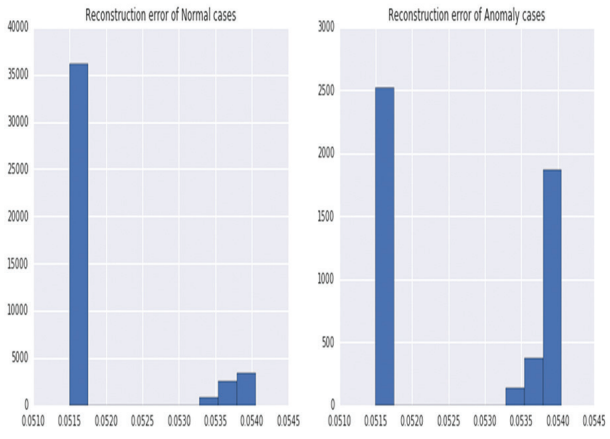

Fig. 12. Reconstruction error for normal and anomaly classes pertaining to activity based anomalies

As presented in Fig. 12, it shows reconstruction error for normal class and also anomaly class. The proposed methodology has tested the entire dataset and the confusion matrix reflecting its detection process is presented in Fig. 13.

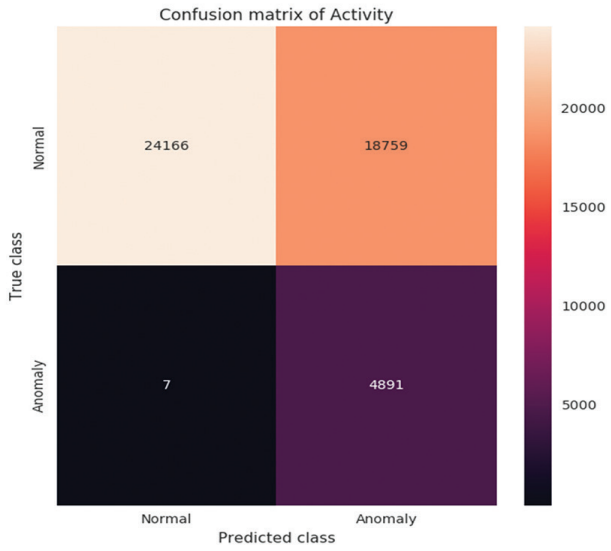

Fig. 13. Confusion matrix for activity based anomaly detection

The confusion matrix visualizes the summary of results containing ground truth and also prediction results. It shows 4891 true positives, 24166 true negatives, 7 false positives and 18759 false negatives. Fig. 14 shows AUC curve reflecting the performance of the proposed system.

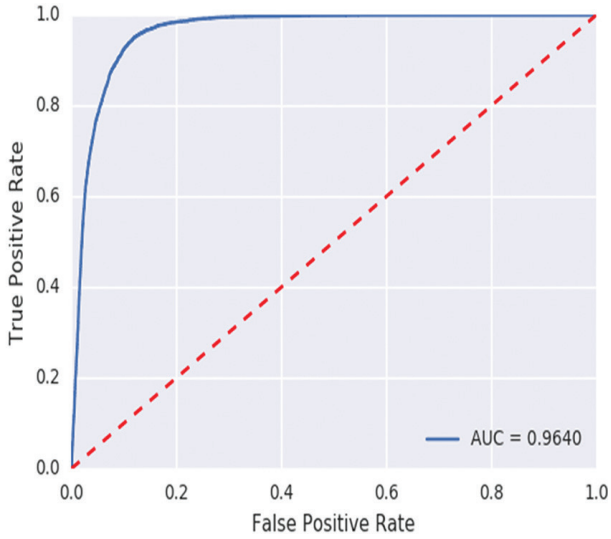

Fig. 14. AUC performance of the proposed system for activity-based anomaly detection

Area Under Curve (AUC) measure is used to assess the performance of the proposed system. With activity-based anomaly detection, the AUC of the proposed model is 0.9640. The activity-based anomaly detection performance is found to be better than that of time based anomaly detection.

4.4. PERFORMANCE COMPARISON

The proposed model is compared against simple autoencoder that does not make use of probability theory.

Table 4. Shows performance comparison among models

Anomaly Detection Model	Precision	Recall	F-Measure
Simple Autoencoder	0.9	0.76	0.824096
Proposed (DAE)	0.95	0.89	0.919022

As presented in Table 4, the performance of the proposed model is compared against simple autoencoder with the proposed framework.

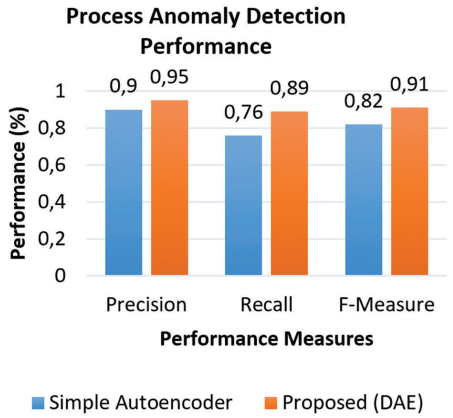

Fig 15. Performance comparison of process anomaly detection

As presented in Fig. 15, the performance of the proposed framework PMiner is compared against deep autoencoder (proposed) and simple autoencoder. It is observed that PMiner is capable of detecting anomalies and rectifying them. However, it could work better with the proposed deep autoencoder which is based on probabilistic theory. The precision achieved by a simple autoencoder with PMiner framework is 90%, recall 76% and F1-Score 82%. The PMiner framework with deep autoencoder could achieve 95% precision, 89% recall and 91% F1-Score. Therefore, the proposed PMiner framework along with the proposed algorithm based on deep autoencoder achieved the highest performance in process anomaly detection and rectification.

5. DISCUSSION

This section discusses important questions like why process mining? how does the proposed method achieve process anomaly detection and rectification? and what is the implication of this research for future endeavours? Enterprise applications in the real world, in the contemporary era, are running businesses through distributed applications. Such applications have several thousands of business processes. Due to the high complexity of the business processes and the concurrency nature of the processes in multi-user environments, there is ever possibility of anomalies in the execution of processes. Such execution dynamics are generally saved into log files known as process logs. If there is an anomaly which is not detected can lead to potential errors in the application. This, in turn, leads to a deterioration of customer satisfaction besides attracting legal issues. Therefore, it is indispensable to monitor process log entries to detect any sort of anomalies and rectify them. Therefore, process mining plays an important role in improving business process consistency. The proposed framework named PMiner in this paper is very useful for this purpose as it can automatically detect business process anomalies and rectify them. The research in this paper has implications that lead to further research endeavours in future.

5.1. LIMITATIONS

Though the proposed framework is capable of detecting and rectifying business process anomalies, it has several limitations. First, it is evaluated with the BPI Challenge 2020 dataset. Though this dataset is close to real-time processes in businesses, the proposed framework has not yet been evaluated by deploying in real enterprise premises with live data. Second, the dataset used for evaluation is relatively smaller in size (7.20 MB) and belongs to a specific domain. Therefore, it is important to evaluate our framework further with data from multiple domains and also with large data. Third, business process log entries grow dynamically. Therefore, it is desired to consider big data environment and computing frameworks to deal with streaming data.These limitations can be overcome by using live streaming of process event logs of enterprises, increasing the data for implicit training of autoencoder and usage of MapReduce kind of parallel processing framework.

6. CONCLUSION AND FUTURE WORK

A process mining framework known as ***PMiner*** is proposed for automatic detection of anomalies in business processes. The framework is designed to take real life business process event logs as input and detect anomalies using a deep autoencoder as it has potential to discriminate anomalies. An algorithm named IBPAD is proposed to realize the framework. This algorithm is able to process business process event logs with the proposed deep autoencoder, detect anomalies and rectify the same. BPI Challenge dataset released by IEEE Task Force on Process Mining is used for the empirical study. The proposed algorithm could achieve highest F1-Score 91% outperforming its existing autoencoder counterpart. In future, we intend to improve our framework to evaluate it with real enterprise application's live streaming business process event logs.

7. REFERENCES

[1] S. Riyanarto, S. Fernandes, S. K. Rossa, "Anomaly detection in business processes using process mining and fuzzy association rule learning", Journal of Big Data, Vol. 7, No. 1, 2020, pp. 1-19.

[2] H. K. Muzzammil, K. Yevgeniya, G. M. Medhat, "A Graph-Based Approach to Interpreting Recurrent

Neural Networks in Process Mining", IEEE Access, Vol. 8, 2020, pp. 172923-172938.

[3] P. Zerbino, A. Stefanini, D. Aloini, "Process Science in Action: A Literature Review on Process Mining in Business Management", Technological Forecasting and Social Change, Vol. 172, 2021, pp. 1-20.

[4] R. Sarno, R. D. Dewandono, T. Ahmad, M. FaridNaufa, "Hybrid Association Rule Learning and Process Mining for Fraud Detection", IAENG International Journal of Computer Science, Vol. 42, No. 2, 2015, pp. 1-14.

[5] N. Timo, L. Stefan, S. Alexander, M. Max, "Analyzing business process anomalies using autoencoders", Machine Learning, Vol. 107, 2018, pp.1-19.

[6] K. Wolfgang, M. Jonas, R. Maximilian, S. Johannes, "Machine Learning in Business Process Monitoring: A Comparison of Deep Learning and Classical Approaches Used for Outcome Prediction", Business & Information Systems Engineering, Vol. 63, 2020, pp. 261-276.

[7] D. A. Neu, J. Lahann, P. Fettke, "A systematic literature review on state-of-the-art deep learning methods for process prediction", Artificial Intelligence Review, Vol. 55, 2021, pp. 801-827.

[8] M. L. Fernandez, G. A. L. Perales, C. F. J. Garcia, P. M. Gil, P. G. Martinez, "A Self-Adaptive Deep Learning-Based System for Anomaly Detection in 5G Networks", IEEE Access, Vol. 6, 2018, pp. 7700-7712.

[9] E. A. Elaziz, R. Fathalla, M. Shaheen, "Deep reinforcement learning for data-efcient weakly supervised business process anomaly detection", Journal of Big Data, Vol. 10, 2023, p. 33.

[10] G. Abhijit, S. Debabrata, "Hybrid Approach to Document Anomaly Detection: An Application to Facilitate RPA in Title Insurance", International Journal of Automation and Computing, Vol. 18, No. 1, 2020, pp. 1-18.

[11] B. Franczyk, "Semi-Supervised Anomaly Detection in Business Process Event Data using Self-Attention based Classification", International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Vol. 192, 2021, pp. 39-48.

[12] N. Timo, L. Stefan, S. Alexander, M. Max, "BINet: Multi-perspective business process anomaly classification", Information Systems, Vol. 103, 2022, p. 101458.

[13] Y. Lu, Q. Chen, S. K. Poon, "A Deep Learning Approach for Repairing Missing Activity Labels in Event Logs for Process Mining", Information, Vol. 13, No. 5, 2022, p. 234.

[14] W. Mathias, M. Marco, W. Ingo, V. Jan, "act2vec, trace2vec, log2vec, and model2vec: Representation Learning for Business Processes", Proceedings of the International Conference on Business Process Management, Sydney, NSW, Australia, 9-14 September 2018, pp. 305-321.

[15] A. Guzzo, M. Joaristi, A. Rullo, E. Serra, "A multi-perspective approach for the analysis of complex business processes behaviour", Expert Systems with Applications, Vol. 177, 2021, pp. 1-13.

[16] M. Vijayakamal, D. Vasumathi, "Unsupervised Learning Methods for Anomaly Detection and Log Quality Improvement Using Process Event Log", International Journal of Advanced Science and Technology, Vol. 29, No. 1, 2020, pp. 1109-1125.

[17] F. M. Maggi, C. D. Francescomarino, M. Dumas, C. Ghidini, "Predictive Monitoring of Business Processes", Lecture Notes in Computer Science, Springer, 2014, pp. 457-472.

[18] Z. Tariq, D. Charles, S. McClean, I. McChesney, Pau, "Anomaly Detection for Service-Oriented Business Processes Using Conformance Analysis", Algorithms, Vol. 15, No. 8, 2022, pp. 1-25.

[19] C. Toon, C. Michelangelo, M. Donato, "Unsupervised Anomaly Detection in Noisy Business Process Event Logs Using Denoising Autoencoders", Proceedings of the International Conference on Discovery Science, Bari, Italy, 19-21 October 2016, pp. 442-456.

[20] P. Singh, M. S. Azari, F. Vitale, F. Flamm "Using log analytics and process mining to enable self healing in the Internet of Things", Environment Systems and Decisions, Vol. 42, 2022, pp. 234-250.

[21] H. Adrien, B. Remi, C. Isabelle, "A Process Mining Approach for Supporting IoT Predictive Security", Proceedings of the IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, 20-24 April 2020, pp. 1-9.

[22] R. Eric, M. G. Jorge, S. Marcos, C. Daniel, "Process mining in healthcare: A literature review", Journal of Biomedical Informatics, Vol. 61, 2016, pp. 224-236.

[23] T. C. Nicoleta, "Process Mining on a Robotic Mechanism", Proceedings of the IEEE International Conference on Software Testing, Verification and Validation Workshops, Porto de Galinhas, Brazil, 12-16 April 2021, pp. 1-8.

[24] M. Jans, J. Martijn, V. Werf, N. Lybaert, K. Vanhoof, "A business process mining application for internal transaction fraud mitigation", Expert Systems with Applications, Vol. 38, No. 10, 2011, pp. 13351-13359.

[25] S. Pauwels, T. Calders, "An anomaly detection technique for business processes based on extended dynamic bayesian networks", Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, April 2019, pp. 494-501.

[26] G. Park, M. P. Wil, V. Aalst, "Action-oriented process mining: bridging the gap between insights and actions", Progress in Artificial Intelligence, 2022, pp. 1-22.

[27] T. Okuboa, H. Kaiya, "Efficient secure DevOps using process mining and Attack Defense Trees", Procedia Computer Science, Vol. 207, 2022, pp. 446-455.

[28] C. Fernández-Llatas, J.-M. Benedi, J. García-Gómez, V. Traver, "Process Mining for Individualized Behavior Modeling Using Wireless Tracking in Nursing Homes", Sensors, Vol. 13, No. 11, 2013, pp. 15434-15451.

[29] L. Sofiane, S. Radu, "Process mining-based approach for investigating malicious login events", Proceedings of the IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, 20-24 April 2020, pp.1-5.

[30] S. A. Kumar, R. Kamra, U. Shrivastava, "Conformance Checking Techniques of Process Mining: A Survey", Recent Trends in Intensive Computing, 2021, pp. 335-341.

[31] F. Folino, L. Pontieri, "AI-Empowered Process Mining for Complex Application Scenarios: Survey and Discussion", Journal on Data Semantics, Vol. 10, 2021, pp. 77-106.

[32] BPI Challenge 2020 dataset, https://www.tf-pm.org/competitions-awards/bpi-challenge/2020 (accessed: 2023)