The Eastasouth Journal of Information System and Computer Science
Vol. 1, No. 03, April 2024, pp. 248-256
ISSN: 3025-566X, DOI: 10.58812/esiscs.v1i03

Failure-to-Risk: Predicting Compliance Breaches from Process Anomalies Using Temporal Explainable AI

Varsha Shah
Independent Researcher, Seattle, WA, USA

Article Info

Article history:
Received April, 2024
Revised April, 2024
Accepted April, 2024

Keywords:
Attention Mechanisms;
Business Process Management;
Compliance-Risk Prediction;
Predictive Process Monitoring;
Process Anomaly Detection;
Recurrent Neural Networks;
Shapley Additive Explanations;
Temporal Explainable Artificial Intelligence

ABSTRACT

Non-conformances in organizational processes are often observed as slight irregularities in the sequence of events, prior to even the formal event of a non-conformance being reported, leading to an increased interest in predictive models that can detect such irregularities and give a reason for each alert. This review summarizes the methodological progress of predictive business process monitoring, temporal deep learning and explainable AI and proposes an overall conceptual framework, called temporal explainable AI for compliance-risk prediction, which brings together sequence encoders (e.g. LSTM networks, attention-based transformers) and explainers for post-hoc interpretation (e.g. Shapley additive explanations, local surrogate models, rule-based anchors). The synthesis moves from classical models of stochastic processes to data-driven temporal architectures, laying out the corresponding loss of native interpretability for predictive flexibility, and describing how post-hoc explanation methods have been devised to reclaim the lost interpretability. The results of a comparative review of benchmark evaluations on predictive process monitoring also show that key aspects of the evaluations that are relevant to predictive process monitoring, such as the encoding strategy, cross-log generalization, or handling of class imbalance, have a significant impact on the reported performance and should be taken into account when designing a deployed compliance-risk system. A comprehensive temporal explainable AI system for compliance monitoring should also include concept-drift detection, as the statistical correlation between process behavior and compliance outcomes will not likely be constant throughout an organization's lifecycle.

Corresponding Author:
Name: Varsha Shah
Institution: Independent Researcher, Seattle, WA, USA
Email: varshachaudhari17@gmail.com

1. INTRODUCTION

Today, a lot of organizations have information systems that capture all the business processes at a specific time as a series of events, leaving behind a trail of rich digital information that can be used to detect regulatory and operational risk. Many of the high-performing algorithms are opaque, meaning that stakeholders who have to make decisions about regulatory outcomes want to know what makes a particular process instance "risky" and not just a number [1].

Sequences of process activities, time stamps, and resources, as observed on a case, can be regarded as an input from which a future event, like a policy violation or a time-limit exceeded, is derived, and transformer networks, tailored for event logs, have been shown to be good models for such traces, as they can be encoded directly without much manual feature engineering [2].

Similarly, financial technology (fintech) has advanced in parallel as a need for explainability has emerged as a key requirement for any automated system that justifies a compliance determination [3].

In credit-risk applications, explainable machine-learning methods have not only been applied to explain a model's predictions for an individual, but also to show that the model's performance is locally and globally consistent with regulatory and fairness requirements over time, which is adopted here as a guiding principle. The rest of the article presents a synthesis of methodological foundations regarding predictive process monitoring, temporal deep learning and explainable artificial intelligence, subsequently presenting a conceptual framework that integrates these aspects for the compliance-risk prediction [4].

In a context of credit risk, it is consistent to the principles of local and global interpretability adopted here, and on which explainable machine-learning techniques have been applied, not only to justify individual predictions, but also to establish that a model's global behavior is consistent with regulatory or fairness requirements over time. This article reviews the methodological underpinnings of predictive process monitoring, temporal deep learning and explainable AI, and suggests a conceptual cross-fertilization of these three branches for compliance-risk prediction [4].

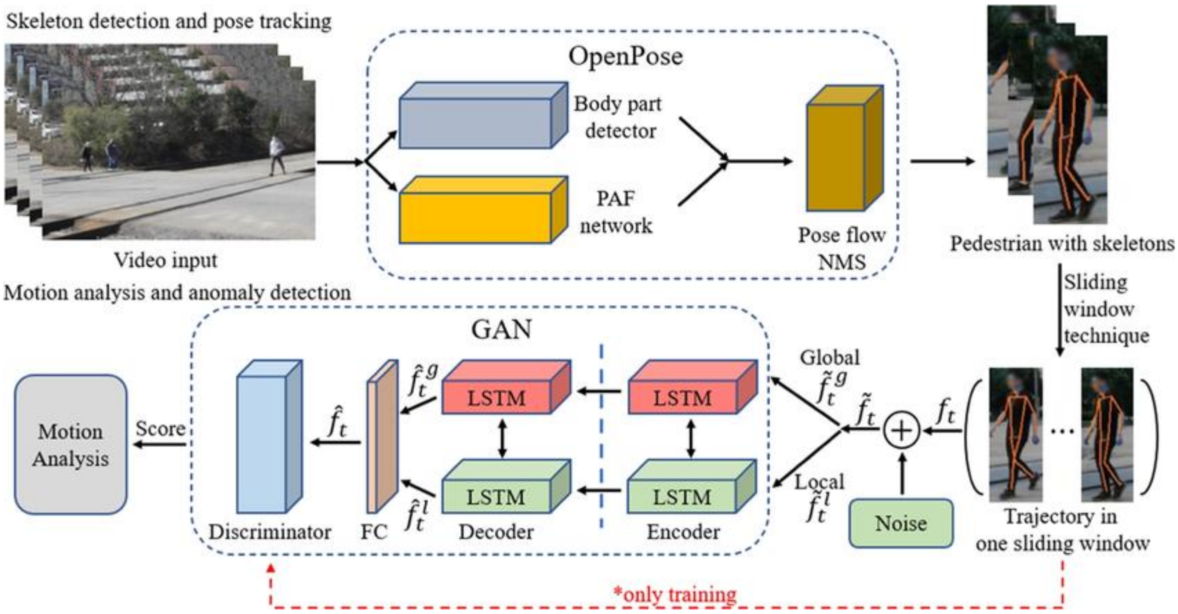

Figure 1. Conceptual pipeline linking event-log ingestion, anomaly detection, temporal encoding, and post-hoc explanation to a compliance-risk score. (ResearchGate,2021)

2. LITERATURE REVIEW

2.1 *Predictive Business Process Monitoring and Compliance Risk*

First, it requires the ability to detect an anomaly in a running process instance relative to some norm, a property which has long been associated with the notion of point anomalies, which affect individual events; contextual anomalies that depend on the context of the surrounding events; and collective anomalies, that affect multiple related events, which are naturally linked to different types of process deviations, such as an out-of-sequence activity, an unusually long waiting time, or a systematic pattern of skipped controls, respectively [5].

Based on this, recommendation systems that forecast risk per process instance have been developed that have proven that historical case information can be extracted to send proactive alerts to the process owners, enabling them to intervene in similar process instances before they become a violation, instead of having to wait until it is detected after the event [6].

More recent work has gone beyond just predicting undesirable events to provide an interpretable "why" behind the prediction: explainable predictive process monitoring, where the model not only predicts the undesirable event, but also provides a rationale behind the prediction that can be used for audit and remediation purposes, directly relevant to compliance settings where an unexplained alert is of little value [7].

Table 1. Literature-to-Concept Mapping Across the Reviewed Reference Set

Thematic Category	Representative References	Core Contribution
Explainability foundations	[1], [13], [18], [19]	General XAI taxonomy and model-agnostic attribution / rule-based explanation techniques
Temporal sequence modeling	[10], [11], [12], [25]	Recurrent and attention-based architectures for sequential event data
Predictive process monitoring	[2], [6], [7], [14], [17], [22], [23]	Application of sequence models to business process event logs
Anomaly and drift detection	[5], [8], [16]	Foundational definitions and deep learning methods for detecting deviations and non-stationarity
Risk and compliance applications	[3], [4], [9], [15], [20], [21], [24]	Domain framing of explainable, risk-oriented prediction in finance and process management

2.2 *Temporal Deep Learning Architectures for Sequential Event Data*

An additional difficulty that compliance monitoring faces is that the statistical characteristics of a business process are not always constant; regulatory requirements may change, seasonal workloads may change, and process redesigns may change the actual meaning of normal behavior, as formally studied in the field of concept drift [8] – in which the correlation between input features and desirable output changes over time. Instead, it may be possible to make use of a drift-aware monitor, which keeps track of the classification error rate over a window of w recent cases and re-calibrates once the number of errors in the window has reached some specified threshold, as expressed in Equation (1), instead of assuming that a single static model will be valid for the entire process operating lifetime.

$$D = (1 / w) \cdot \Sigma\ (t = T-w+1 \text{ to } T)\ \mathbb{1}[\hat{y}t \neq yt] \quad (1)$$

A business process instance contains control-flow information, transactional information and resource allocation, so multi-perspective analytical approaches combine these perspectives into a single representation of business process behavior (which is important when dealing with compliance violations, as they often result from the combination of an unusual control-flow path and an unusual resource allocation, but not from either one alone) [9].

The ability to keep information over long sequences of input is one of the challenges faced by earlier recurrent architectures; long short term memory networks solve this problem by adding a forget, an input and an output gate to a memory cell, enabling a network that processes an event log to selectively retain early processed indicators, such as an unexpected starting activity, over the many events that follow it before a subsequent compliance violation occurs [10]. This gating mechanism is written mathematically in Equation (2), where σ is a logistic sigmoid function and ⊙ is an elementwise multiplication.

$$ft = \sigma(Wf\cdot[ht-1, xt] + bf),\ it = \sigma(Wi\cdot[ht-1, xt] + bi),\ \tilde{c}t = \tanh(Wc\cdot[ht-1, xt] + bc)\ ct = ft \odot ct-1 + it \odot \tilde{c}t,\ ht = ot \odot \tanh(ct) \quad (2)$$

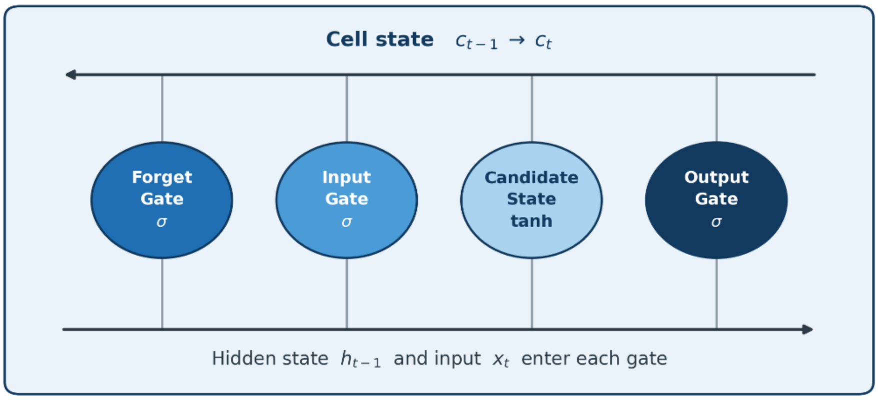

Figure 2. Structure of a long short-term memory (LSTM) cell, showing the forget, input, and output gates that regulate information flow across an event sequence.

Placing these recurring designs in the context of deep learning for time series classification reveals that convolutional, recurrent and hybrid architectures have different benefits regarding training efficiency and their ability to capture long-term temporal relationships, which directly influences the design of the encoder for a business process trace that is fed into the risk scoring function [11].

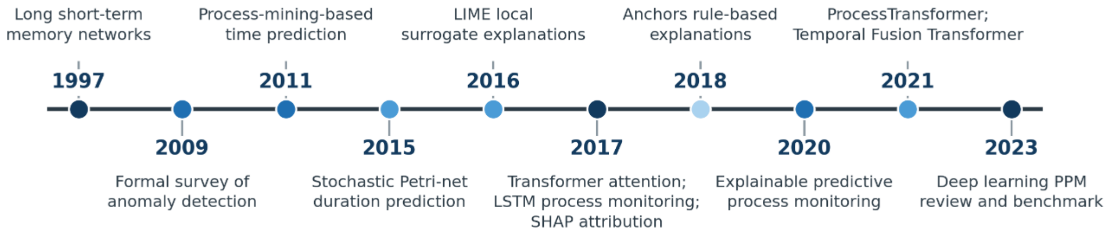

Figure 3. Timeline of foundational methods in predictive process monitoring and explainable AI referenced in this review.

Table 2. Comparative Overview of Predictive Process Monitoring Modeling Paradigms

Paradigm	Representative Approach	Temporal Handling	Interpretability Characteristic	Key Source(s)
Stochastic / analytical	Non-Markovian stochastic Petri nets	Explicit duration distributions per model state	High (fully auditable states)	[20], [24]
Recurrent neural	LSTM-based next-activity and remaining-time prediction	Sequential gating with memory cell	Low natively; requires post-hoc explanation	[10], [14], [22]
Attention-based	ProcessTransformer / Temporal Fusion Transformer	Parallel self-attention over full sequence	Moderate (attention weights offer partial insight)	[2], [12], [25]
Ensemble / classical ML	Gradient-boosted trees, rule lists	Feature-engineered temporal aggregates	Moderate to high, depending on model	[17], [23]

2.3 *Temporal Attention and Explainable AI Foundations*

Temporal fusion transformers are the extension of the attention-based sequence modelling to multi-horizon forecasting, by adding recurrent layers for local processing alongside self-attention layers for capturing longer-range dependencies while also producing interpretable attention weights that can be used to explain their predictions by showing, for a given forecast, which previous time steps and which input variables were most important for the prediction: a feature that aligns with the need to explain the output of a compliance-risk model on the basis of specific events in the past [12].

A complementary approach to the question of interpretability is to decompose a prediction into additive contributions from individual input features; for example, model-agnostic attribution methods such as Shapley values [13] begin with an expected output (φ0) for the model and then calculate each term as the expected contribution of a particular activity, duration, or resource attribute to the difference between the prediction and the expected output.

$$f(x) = \varphi 0 + \Sigma\ (j = 1 \text{ to } M)\ \varphi j \quad (3)$$

a

data-aware extensions of recurrent architectures that are applied specifically to the problem of remaining-time prediction in business processes, showing that incorporating the attributes improves the precision of the prediction of a process instance's remaining-life, and hence its ability to estimate the likelihood that it will violate a deadline-based compliance requirement [14].

Many of the conventions used to evaluate the effectiveness of fraud and compliance-breach detection systems, such as the focus on identifying truly anomalous transactions, were introduced by financial data mining research where the number of fraudulent cases is smaller compared with the far larger population of routine transactions, a situation also typical of compliance-breach prediction in day-to-day operational business

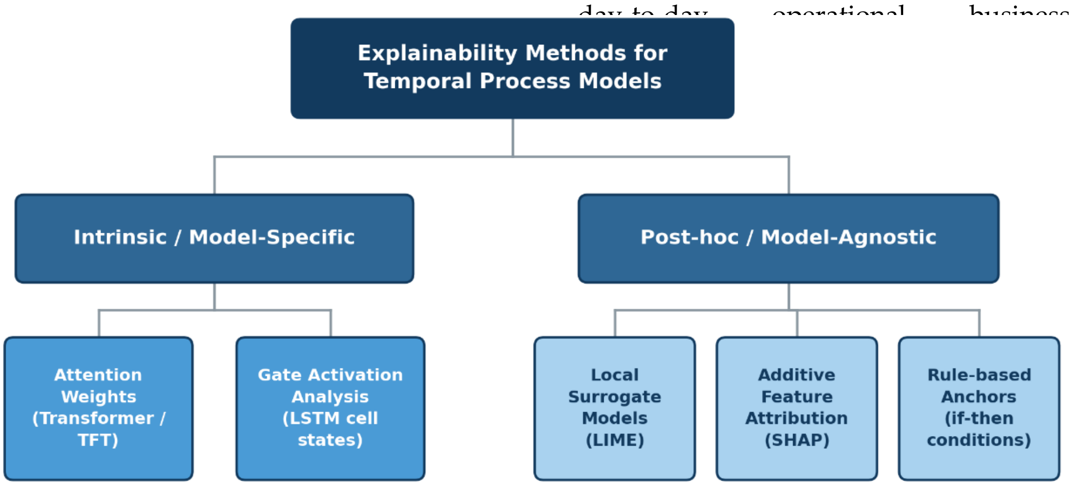

Figure 4. Taxonomy of explainability approaches applicable to temporal process models, distinguishing intrinsic / model-specific methods from post-hoc / model-agnostic methods.

Table 3. Taxonomy of Explainability Techniques Applicable to Temporal Compliance Models

Method	Scope	Model Dependency	Output Form	Key Source
Attention weights	Local and global	Model-specific	Weighted relevance over time steps / variables	[12], [25]
SHAP	Local and global	Model-agnostic	Additive per-feature attribution values	[13]
LIME	Local	Model-agnostic	Sparse linear surrogate coefficients	[18]
Anchors	Local	Model-agnostic	High-precision if-then rule	[19]

2.4 *Anomaly Detection Paradigms and Model-Agnostic Explanation Techniques*

Anomaly detection techniques based on deep learning can be decomposed into three types: reconstruction-based, one-class, and prediction-based [16] and reconstruction-based methods often have the process trace evaluated by the quality of its reconstruction from an autoencoder trained on the normal process behavior, such that the reconstruction error of the process trace that is significantly larger than the distribution of the process traces that are compliant is investigated further for anomalies as illustrated in Equation (4) where f and g represent the encoder and decoder functions, respectively.

$$A(x) = \|x - g(f(x))\|^2 \quad (4)$$

The study reported above provides a systematic review and benchmark of the deep learning-based methods developed for predictive process monitoring, and suggests that recurrent and attention-based deep learning architectures typically perform better than previous feature-based baseline architectures for predicting next-activity and predicting-outcome tasks, but that reported accuracy gains are highly heterogeneous across event-log datasets; the authors emphasize the need to test any proposed temporal explainable AI framework on a variety of event-log datasets that represent a range of compliance-relevant problem types—rather than a single benchmark—when evaluating its success.

The local interpretable model-agnostic explanations approximate the behavior of a complex classifier in the neighborhood of a given prediction by fitting a surrogate model g that is inherently interpretable, in the form of the optimization problem in Equation (5): the first term quantifies the unfaithfulness of g to f within some neighborhood of the prediction; the second term penalizes the complexity of g; this provides a simple alternative to attention-based interpretability when the underlying compliance-risk model is not easily differentiable [18].

$$\xi(x) = argmin\ (g \in G)\ L(f, g, \pi x) + \Omega(g) \quad (5)$$

Table 4. Anomaly Detection Paradigms Relevant to Process Compliance

Paradigm	Detection Principle	Typical Assumption	Suitability for Event Logs	Key Source
Statistical / distance-based	Deviation from distributional norms or nearest neighbors	Normal data forms a dense, describable region	Effective for numeric case attributes	[5]
Reconstruction-based (autoencoder)	High reconstruction error signals anomaly	Autoencoder trained only on compliant behavior	Effective for high-dimensional trace embeddings	[16]
Concept-drift-aware monitoring	Sustained shift in error / statistical properties over time	Underlying process distribution may change	Necessary for long-lived, evolving processes	[8]

3. CONCEPTUAL FRAMEWORK: TEMPORAL EXPLAINABLE AI FOR COMPLIANCE-RISK PREDICTION

The techniques presented above are based on attention or attribution but not both, whereas a high precision, rule-based explanation (also referred to as an anchor) uses only rules to express an explanation in the form of a small number of if-then conditions that, with high probability, guarantee a certain prediction, given other features, and can be captured easier by a non-technical compliance officer than a continuous attribution score.

Placing these explainable deep learning methods in contrast to classical models of stochastic processes is valuable:

non-Markovian stochastic Petri nets are analytically tractable methods that provide flexible predictions based on incomplete process models and historical timing data at the expense of requiring a formally described process model where these models are applicable [20].

The focus on interpretable predictive models for business processes has included the use of decision trees, rule lists, and shallow ensembles as well as deep architectures, and the current consensus is that the interpretability benefits of simpler models come with an inevitable tradeoff in terms of measurable loss in predictive accuracy on complex, high-dimensional business event logs, which has led to the hybrid approach proposed here, where the former are deployed for accuracy and the latter are attached as a post hoc explainer to reclaim that interpretability, as shown conceptually in Figure 1.

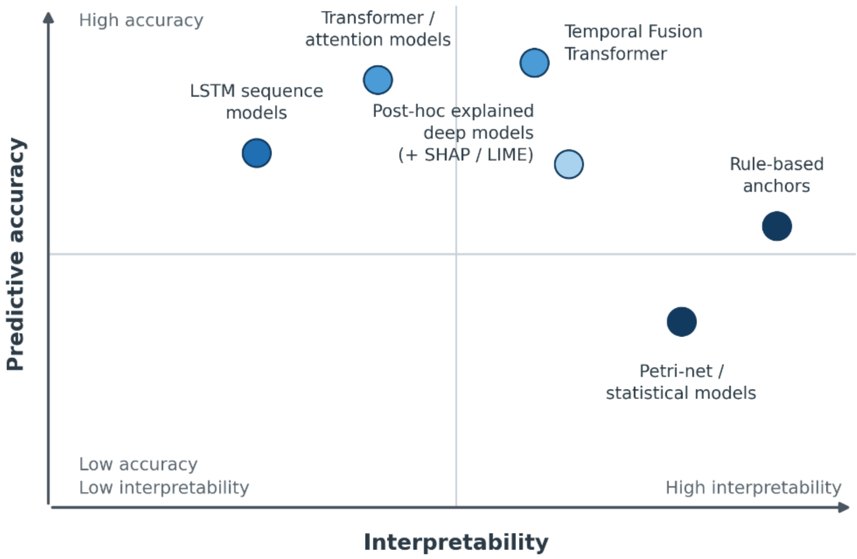

Figure 5. Qualitative positioning of the modelling approaches reviewed, by interpretability and predictive accuracy, as characterized in the surveyed literature.

4. COMPARATIVE ANALYSIS AND DISCUSSION

Initial attempts at predictive business process monitoring and prediction of next activities and remaining time from raw event logs using LSTM networks have set an empirical benchmark for later attention-based and explainable approaches and are a viable minimum viable product for organisations that have just begun to instrument compliance-risk prediction [22].

A large and comprehensive review and benchmark of outcome-oriented predictive process monitoring techniques is performed in [23] that entails reviewing a diverse set of classifiers and provides a consistent experimental protocol, highlighting that the choice of encoding for case attributes and control-flow information can be as critical as the choice of classifier, with direct implications on how to prepare event logs before they are passed to a temporal explainable AI pipeline, at least for gradient-boosted trees and recurrent neural networks.

Process-mining-based time prediction is another non-neural benchmark for comparison, as it annotates a discovered process with time information based on historical process logs; does not demand gradient-based training, and is still useful as a valid benchmark for process organizations that do not yet have the computational and governance burden of deep temporal architectures [24].

The attention weights computed by the self-attention mechanism in transformer-based encoders, which represent the relative influence of the prior events on the current representation, are a key mathematical concept in the transformer architectures used in the process and the multi-horizon forecasting models presented in this review [25].

$$Attention(Q,K,V) = softmax(QK^T/\sqrt{dk})\cdot V \quad (6)$$

Table 5. Benchmark Review Dimensions in Predictive Process Monitoring Literature

Evaluation Dimension	Description	Addressed By
Task formulation	Next-activity, remaining-time, or outcome-oriented prediction	[17], [22], [23]
Encoding strategy	Representation of control-flow, case, and resource attributes	[9], [23]
Cross-log generalization	Performance consistency across heterogeneous, real-world event logs	[17]
Class imbalance handling	Treatment of the disproportion between compliant and non-compliant cases	[15], [23]

5. CONCLUSION

This review has combined the methodological advances of predictive business process monitoring, temporal deep learning and explainable artificial intelligence to present a conceptual framework for predicting compliance violations based on process anomalies. This synthesis leads to three contributions. First, the path from stochastic process models to recurrent and attention-based models is a consistent trade-off between predictive accuracy and native interpretability, and techniques used to explain post-hoc like Shapley additive explanation, local surrogate models, or anchors, are explicitly developed to be able to reconcile this trade-off. Second, there are well-documented challenges in the predictive process monitoring literature that are relevant for compliance-risk prediction, such as class imbalance (non-compliant versus compliant cases) and the requirements for benchmarking across an event log of varying types of cases, not just one event log. Third, the extensions to concepts that this paper has explored, which include a concept-drift-aware, multi-perspective and data-aware extensions, indicate that a temporal explainable AI system for compliance monitoring is not a one-off construct but rather should have mechanisms in place to detect when the process it monitors has changed. Future research could explore the proposed conceptual framework on actual organizational event logs across several regulation domains, and consider reconciling between attention-based and Shapley-value-based explanations, if they differ in their explanations of a particular compliance-risk prediction.

REFERENCES

[1] A. Barredo Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.

[2] Z. A. Bukhsh, A. Saeed, and R. M. Dijkman, "ProcessTransformer: Predictive business process monitoring with transformer network," arXiv preprint, 2021, doi: 10.48550/arXiv.2104.00721.

[3] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable AI in fintech risk management," Frontiers in Artificial Intelligence, vol. 3, Art. no. 26, 2020, doi: 10.3389/frai.2020.00026.

[4] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable machine learning in credit risk management," Computational Economics, vol. 57, pp. 203–216, 2021, doi: 10.1007/s10614-020-10042-0.

[5] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys, vol. 41, no. 3, Art. no. 15, pp. 1–58, 2009, doi: 10.1145/1541880.1541882.

[6] R. Conforti, M. de Leoni, M. La Rosa, W. M. P. van der Aalst, and A. H. M. ter Hofstede, "A recommendation system for predicting risks across multiple business process instances," Decision Support Systems, vol. 69, pp. 1–19, 2015, doi: 10.1016/j.dss.2014.10.006.

[7] R. Galanti, B. Coma-Puig, M. de Leoni, J. Carmona, and N. Navarin, "Explainable predictive process monitoring," in Proc. 2nd Int. Conf. Process Mining (ICPM), 2020, pp. 1–8, doi: 10.1109/ICPM49681.2020.00012.

[8] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," ACM Computing Surveys, vol. 46, no. 4, Art. no. 44, pp. 1–37, 2014, doi: 10.1145/2523813.

[9] A. Guzzo, M. Joaristi, A. Rullo, and E. Serra, "A multi-perspective approach for the analysis of complex business processes behavior," Expert Systems with Applications, vol. 177, Art. no. 114934, 2021, doi: 10.1016/j.eswa.2021.114934.
[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[11] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," Data Mining and Knowledge Discovery, vol. 33, no. 4, pp. 917–963, 2019, doi: 10.1007/s10618-019-00619-1.
[12] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," International Journal of Forecasting, vol. 37, no. 4, pp. 1748–1764, 2021, doi: 10.1016/j.ijforecast.2021.03.012.
[13] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems, vol. 30, pp. 4765–4774, 2017, doi: 10.5555/3295222.3295230.
[14] N. Navarin, B. Vincenzi, M. Polato, and A. Sperduti, "LSTM networks for data-aware remaining time prediction of business process instances," in Proc. 2017 IEEE Symp. Series Comput. Intell. (SSCI), 2017, pp. 1–7, doi: 10.1109/SSCI.2017.8285184.
[15] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," Decision Support Systems, vol. 50, no. 3, pp. 559–569, 2011, doi: 10.1016/j.dss.2010.08.006.
[16] G. Pang, C. Shen, L. Cao, and A. van den Hengel, "Deep learning for anomaly detection: A review," ACM Computing Surveys, vol. 54, no. 2, Art. no. 38, pp. 1–38, 2022, doi: 10.1145/3439950.
[17] E. Rama-Maneiro, J. C. Vidal, and M. Lama, "Deep learning for predictive business process monitoring: Review and benchmark," IEEE Transactions on Services Computing, vol. 16, no. 1, pp. 739–756, 2023, doi: 10.1109/TSC.2021.3139807.
[18] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.
[19] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision model-agnostic explanations," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1, pp. 1527–1535, 2018, doi: 10.1609/aaai.v32i1.11491.
[20] A. Rogge-Solti and M. Weske, "Prediction of business process durations using non-Markovian stochastic Petri nets," Information Systems, vol. 54, pp. 1–14, 2015, doi: 10.1016/j.is.2015.04.004.
[21] R. Sindhgatta, C. Moreira, C. Ouyang, and A. Barros, "Exploring interpretable predictive models for business processes," in Business Process Management (BPM 2020), Lecture Notes in Computer Science, vol. 12168, 2020, pp. 257–272, doi: 10.1007/978-3-030-58666-9_15.
[22] N. Tax, I. Verenich, M. La Rosa, and M. Dumas, "Predictive business process monitoring with LSTM neural networks," in Advanced Information Systems Engineering (CAiSE 2017), Lecture Notes in Computer Science, vol. 10253, 2017, pp. 477–492, doi: 10.1007/978-3-319-59536-8_30.
[23] I. Teinemaa, M. Dumas, M. La Rosa, and F. M. Maggi, "Outcome-oriented predictive process monitoring: Review and benchmark," ACM Transactions on Knowledge Discovery from Data, vol. 13, no. 2, pp. 1–57, 2019, doi: 10.1145/3301300.
[24] W. M. P. Van der Aalst, M. H. Schonenberg, and M. Song, "Time prediction based on process mining," Information Systems, vol. 36, no. 2, pp. 450–475, 2011, doi: 10.1016/j.is.2010.09.001.
[25] A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017, doi: 10.5555/3295222.3295349.