Computer Networks 243 (2024) 110303

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Robust integration of blockchain and explainable federated learning for automated credit scoring

Zorka Jovanovic [a,*], Zhe Hou [a], Kamanashis Biswas [b], Vallipuram Muthukkumarasamy [a]

[a] Griffith University, Australia
[b] Australian Catholic University, Australia

ARTICLE INFO

Keywords:
Automated credit scoring
Blockchain
Explainable artificial intelligence
Decentralised federated learning

ABSTRACT

This article examines the integration of blockchain, eXplainable Artificial Intelligence (XAI), especially in the context of federated learning, for credit scoring in financial sectors to improve the credit assessment process. Research shows that integration of these cutting-edge technologies is in its infancy, specifically in the areas of embracing broader data, model verification, behavioural reliability and model explainability for intelligent credit assessment. The conventional credit risk assessment process utilises historical application data. However, reliable and dynamic transactional customer data are necessary for robust credit risk evaluation in practice. Therefore, this research proposes a framework for integrating blockchain and XAI to enable automated credit decisions. The main focus is on effectively integrating multi-party, privacy-preserving decentralised learning models with blockchain technology to provide reliability, transparency, and explainability. The proposed framework can be a foundation for integrating technological solutions while ensuring model verification, behavioural reliability, and model explainability for intelligent credit assessment.

1. Introduction

Credit Assessment (CA) is the process that assures the development of credit scorecards to assess the creditworthiness of the customers and loan applications following the policy of the lending institution. Mature banks are looking at making the process efficient, transparent and sustainable to reduce the model risk and provide adequate governance. Increasing competition and growing pressure for revenue generation are setting the requirements for the banks to explore further effective integration and technologies that will result in quicker turnaround time while managing the authenticity of the data source, transparency and privacy protection. It is necessary to employ an efficient, transparent, traceable, secure, and interpretable modelling process to ensure accurate credit risk assessment. This approach aims to minimise model risk, mitigate bias and imperfections, and deliver reliable and sufficient results.

Problem statement: Traditional financial institutions assess credit applications based on data available to them at the time of the credit application, such as customers' credit scores, existing debt and income. The risk associated with the customer's creditworthiness may not be appropriately identified as the customer data comes from a single source of information provided at the time of the application, such as historical spending patterns. Thus, it does not consider the broader dynamic transactional customer data associated with the customer's financial behaviour, such as dynamic payment behaviour, spending patterns, and financial health. To mitigate the data scarcity in small and medium-sized financial institutions and reduce information asymmetry between lenders and borrowers, a proposed solution in [1] involves leveraging blockchain technology to establish a credit data-sharing alliance.

The risk in the existing process is that the customer may be deemed creditworthy based on the limited dataset. In [2] presented that combining call-detail records with traditional data in credit scoring models significantly increases their performance. At the same time, they may have a high likelihood of defaulting on their credit obligations. Additionally, some customers may be unfairly denied due to limited credit history [3]. Ensuring the absence of model bias and discrimination is crucial throughout the scoring process [4]. Therefore, there is a risk to the existing credit assessment system's trustworthiness, efficiency and fairness [5].

The credit assessment process encounters several challenges that need to be addressed. Firstly, recognising the increasing challenges arising from liability concerns, sharing or broadcasting data across various organisations. Various data-sharing regulations, such as the General

[*] Corresponding author.
E-mail addresses: zorka.jovanovic@griffith.edu.au (Z. Jovanovic), z.hou@griffith.edu.au (Z. Hou), kamanashis.biswas@acu.edu.au (K. Biswas), v.muthu@griffith.edu.au (V. Muthukkumarasamy).

https://doi.org/10.1016/j.comnet.2024.110303
Received 23 October 2023; Received in revised form 11 February 2024; Accepted 1 March 2024
Available online 7 March 2024

Data Protection Regulation (GDPR), limit data sharing opportunities across different organisations. Hence, increasingly complex regulatory compliance and governance requirements must be met [6]. Additionally, verifying and validating the accuracy of customer-provided information is crucial, ensuring that it reflects their actual financial behaviour, including income, expenses, assets, and liabilities. Investigating the customer's history of late payments or inconsistent employment records is also important. Traditionally, credit risk assessment has relied on historical application data, but acquiring reliable and dynamic transactional data for model development has proven challenging. Another essential aspect is the need for secure, transparent, traceable, explainable, and robust modelling techniques to ensure an ethical credit assessment decision process. Evaluating the strength and adaptability of the scorecard model is another area of concern. Assessing its robustness and enabling dynamic updates is crucial to ensure its effectiveness over time. Finally, leveraging emerging technologies becomes necessary to construct an intelligent and reliable scorecard engine that upholds data privacy, security, and immutability while enhancing the overall customer experience.

Blockchain can facilitate providing a decentralised credit scoring solution, as it trains a single credit scoring model without sharing customer data, as Hassija et al. [7] suggested. Additionally, Federated Learning (FL) may serve as part of a privacy-preserving machine learning framework, allowing multiple parties to collaboratively train a single credit scoring model without sharing their customer data. However, the authors in [7] consider single-model training.

The researchers in [8] presented a credit scoring system that combines explainable federated learning and blockchain to tackle challenges related to credit model sharing and safeguarding data privacy. Their method elucidates the FL process, suggesting a decentralized Byzantine fault-tolerant stochastic gradient descent algorithm (D-SGD). From a mathematical perspective, the study integrates the Shapley value with DPOS (Delegated Proof of Stake) as a consensus protocol, enabling the algorithm to compute the contribution values of the involved parties during the execution of the federated algorithm.

Imteaj and Amini [9] introduced a model based on FL to anticipate financial distress among borrowers. This approach involves constructing a global machine-learning model that evolves from the local models of distributed agents. The model achieved prediction accuracy almost indistinguishable from that of a centralised model. However, there is no interpretability of the model and local model generation applying Stochastic Gradient Descent (SGD).

Cheng et al. [10] proposed SecureBoost, an FL boosting model, providing theoretical evidence that the model achieves accuracy on par with the non-federated boosting model. However, the model is not interpretable.

Our motivation is to thoroughly understand customers' creditworthiness and trustworthiness in model prediction and to address the needs for well-informed decision-making in the financial sector while protecting customer privacy. We emphasise collaborative modelling, privacy-preserving protection, and adherence to regulatory requirements to ensure the accuracy and reliability of credit assessments while respecting all involved parties' privacy settings. This approach provides various benefits, including enhanced collaboration among multiple entities, improved privacy protection through multiparty privacy-preserving measures, and the development of more accurate credit assessment models. Integrating advanced technologies such as blockchain, FL, and XAI also fosters technological innovation and creates trustworthiness and unbiased credit assessment models. Overall, these advancements aim to optimise credit risk management, reduce defaults, and strengthen trust in the financial system.

The significance of this research is as follows:

- A novel credit assessment process is required that leverages comprehensive data sources and applies them to advanced Artificial Intelligence (AI) algorithms to provide a more holistic view of the customer's creditworthiness. The real-world financial sectors require a broader range of data sources and models to ensure well-informed decision-making in the credit assessment process.
- Enabling multi-source data support in credit modelling promotes collaborative modelling among multiple parties while upholding privacy. Moreover, incorporating multiparty privacy-preserving protection in credit modelling carries significant business benefits by facilitating accurate credit assessment while ensuring the privacy of all involved parties.
- Trustworthiness and unbiased evaluation are essential for reliable credit assessment processes. However, the complex algorithms utilised in FL and blockchain-based consensus mechanisms can obscure the rationale behind credit assessments, posing challenges in meeting evolving regulatory requirements around explainability. While the use of blockchain can enable trustworthiness and transparency, XAI contributes to fairness in credit scoring.
- Enabling the reliability and impartiality of the credit assessment models by incorporating adherence to regulatory requirements. Therefore, its significance enables financial institutions to employ trustworthy, unbiased credit assessment models.

We propose an automated credit decision framework focusing on the robust integration of blockchain and XAI to achieve these goals. The primary contributions of this paper can be summarised as follows:

- Our research explores the fundamental features of XAI and blockchain for credit scoring. We have conducted an in-depth credit scoring analysis and presented a taxonomy of the blockchain and XAI, which has not been done before. Our comprehensive taxonomy of blockchain and XAI features highlights their importance and insights for use within credit assessment. This can assist researchers and practitioners in navigating and applying these evolving technologies effectively within the domain of credit scoring.
- Performed a comparative analysis of proposed architectures that combine blockchain, FL, and XAI technologies to construct credit scoring systems. We examine the difficulties of integrating these technologies into credit assessment, addressing fundamental challenges and examining the integration mechanisms implemented across diverse industrial applications. Our findings highlight that current solutions primarily focus on data storage security and privacy, with limited impact on model verification, behavioural reliability and explainability in intelligent credit assessment.
- We present a conceptual framework that combines blockchain, FL, and XAI technologies to establish an automated decision-making credit assessment process. By utilising the optimal features of these technologies, this framework aims to fulfil the requirements of the banking industry and regulatory standards. The result is a credit scoring system that is both effective and explainable, thereby enhancing reliability and transparency in the decision-making process.

After thoroughly examining the qualitative features required for designing an efficient credit scoring framework, we performed an initial complexity analysis of the proposed framework.

The rest of this paper is organised as follows: Section 2 presents the key blockchain concepts. Section 3 reviews XAI techniques, including the surveyed work on the XAI and CA. Furthermore, we analysed the integration of FL, blockchain and XAI. In Section 4, we present the credit assessment's principal functionalities, including the existing process's limitations. Section 5, we propose the conceptual framework that will address identified limitations and provide an outlook on future research. Section 6 presents the analysis of the key characteristics required to build a robust credit assessment. Finally, Section 7 concludes the paper. Fig. 1 presents an overview of related work integrating blockchain and XAI for credit assessment.

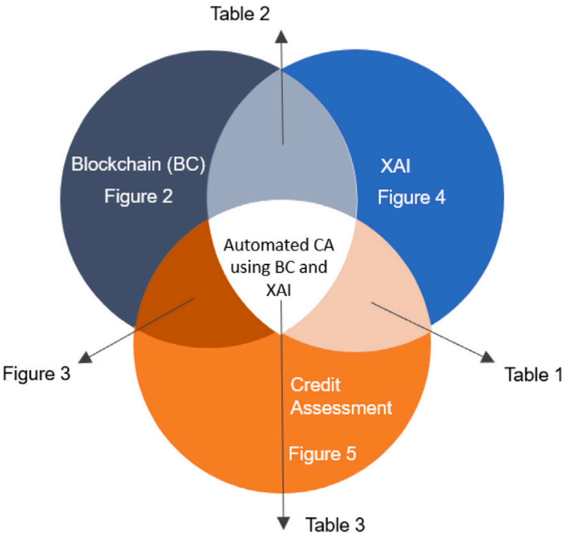


Fig. 1. Overview of the integration of technologies and credit assessment.

2. Blockchain features for credit assessment

This section describes the background of blockchain features and the possible contributions to the credit assessment.

2.1. Blockchain technology

Blockchain as a distributed ledger technology was introduced with Bitcoin [11] to solve the double-spending problem using a peer-to-peer network. The proposed peer-to-peer distributed timestamp server uses the *Proof-of-work (PoW)* system to record a chronological order of transactions into timestamp blocks. The author defined an electronic coin as a chain of digital signatures. The hash of the previous transaction and the owner's private key are required to sign their transactions digitally. The public key is used to verify the sender's identity. A proposed timestamp server consists of the hash of a block of timestamped items and the previous timestamp in its hash, constructing a chain. The *Proof-of-work* system was implemented by incrementing a nonce in the block until a value is found that gives the block's hash the required zero bits, consistent with the SHA-256 algorithm. The *Proof-of-work* requires CPU computation for mining the network nodes and finding a *Proof-of-work* for its block. Once the node considers the *Proof-of-work*, it broadcasts the block to all nodes. The nodes accept the block only if all transactions are validated and not spent. A hash of the accepted block is created to be used as a previous hash for the new block in the network. The block header contains the hash of the previous block validated and a hash of all transactions contained in the block (Merkle tree) as presented in Fig. 2. Privacy is preserved by keeping the public key anonymous.

The history of the blockchain, starting from Blockchain 1.0 to Blockchain 4.0, has been discussed by Tanwar [12]. Blockchain 5.0 is the latest generation of blockchain that has been applied together with AI, hyper-converged infrastructure, and industry 4.0 technologies for high security, efficiency, reliability, and scalability [12]. Verma et al. [13] evaluated the integration of blockchain with Industry 5.0 focusing on how the technology can enhance the security challenges of cyber–physical systems, such as security, trust and transparency.

A systematic literature review of the blockchain-based application across multiple domains such as supply chain, business process enactment [14], financial, healthcare, IoT, privacy and data management has been analysed [15–19]. Investigating trends in blockchain technology by applying text mining and clustering for register patents provides insights for researchers and inventors [20].

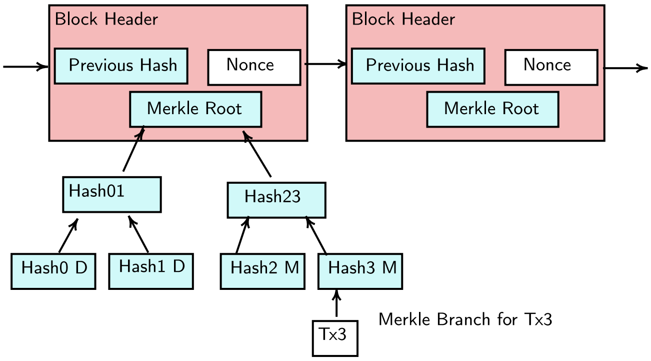


Fig. 2. Merkle tree of hash data in blockchain.

The taxonomy of the blockchain for business process enactment presented in [14] is based on the two characteristics of *capabilities* and *enforced guarantees*. The analysis categorised capabilities based on the different factors of model support, resource allocation and process flexibility. Model support is a notation for the business process. The authors used resource allocation to differentiate various source allocations and examined how different approaches impacted process flexibility. Regarding enforcement, the authors determine that control flow, resource allocation, and data-integrity aspects are enforced on-chain.

The benefits of using blockchain technology to improve security, transparency, and trust in different applications such as Multi-Agent Systems (MAS) [21], energy market [22], identity management [23, 24], multi-organisation collaboration system [25] and data store [26].

A detailed survey on blockchain applications for AI shows that adopting blockchain for AI applications is still in its infancy, [27]. There are many research challenges to be addressed in areas related to privacy, smart contract security, trusted oracles, scalability, consensus protocols, standardisation, interoperability, quantum computing resiliency, and governance [27–29].

The integration of FL with blockchain to address machine learning models' privacy, security and scalability challenges in distributed environments has been analysed. Aledhari et al. [30] and Qu et al. [31] provided an overview of the enabling technologies, protocols and applications of FL and blockchain-enabled FL, respectively. The comprehensive overview of research in blockchain-based FL with different consensus mechanisms and privacy-preserving techniques is presented in [32]. The blockchain approach to enhance security and privacy FL for IoT is proposed in [33,34]. Issa et al. [34] discussed the challenges and risks of using centralised storage and deep learning for IoT applications. While FL is a promising solution for preserving data privacy, it still has a challenge of the model vulnerability. Issa et al. [34] proposed utilising the blockchain smart contract to safeguard FL and reviewed the blockchain-based FL techniques securing IoT systems.

2.2. Blockchain technology and credit assessment

The prospect of integrating blockchain within the banking and financial sector has been presented in [35,36]. Nowadays, blockchain's breakthrough is in data storage and information transmission. Regulation, efficiency, and security are the challenges to be resolved for the anticipated integration of blockchain technology in the banking industry.

Blockchain integrates computer technologies, distributed data storage, information transmission, consensus mechanisms, and encryption algorithms. Fig. 3 presents the blockchain's taxonomy for the credit assessment application. The taxonomy is organised into four dimensions: type, storage, blockchain features and applications in credit assessment. A public blockchain is open and permissionless, and decentralised. A private blockchain is permissioned, and access to a network

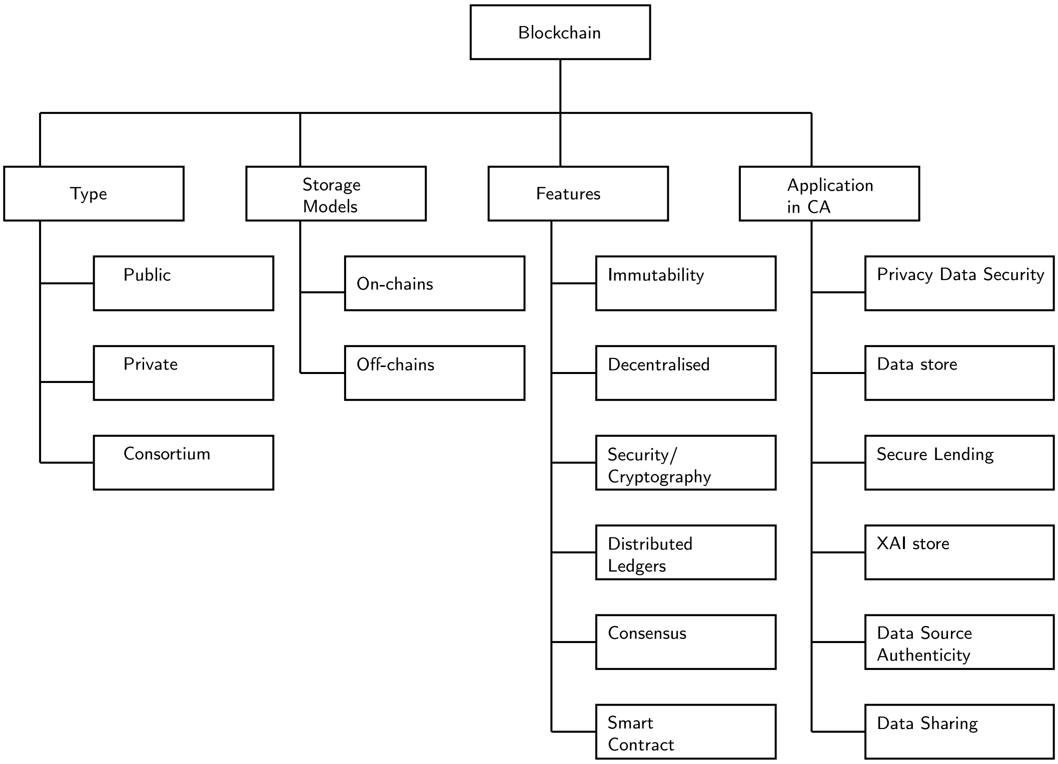


Fig. 3. Taxonomy of blockchain technology for credit assessment.

is restricted to authorised participants. A blockchain consortium is semi-decentralised and permissioned, meaning nodes from multiple organisations collectively own and manage the network. Storage on the blockchain can be “on-chain” or “off-chain”. On-chain refers to the data stored on the blockchain, which means the network nodes verify it. Off-chain refers to storing data outside of the blockchain in a separate system. The features of blockchains are Immutability, Decentralised, Security and Cryptography, Distributed Ledgers, Consensus and Smart Contracts. The immutability features of the blockchain ensure data integrity and transparency. Blockchains are decentralised networks without a central authority and are thus more resilient against attack. Blockchains use various cryptographic algorithms, such as asymmetric-key algorithms (digital signatures), hashing, public-key cryptography, elliptic curve cryptography, and the Merkel tree. Asymmetric-key algorithms (digital signatures) are used to authenticate transitions and ensure the parties approve them. Hashing uses mathematical algorithms to generate one-way functions while ensuring the immutability and integrity of the data stored on the network. Public-key cryptography is used to authenticate transactions and verify identity in the network. Elliptic curve cryptography ensures private keys’ security and authenticates transactions. The Merkle Tree verifies the integrity of the transaction data. Consensus in blockchain refers to the process whereby nodes in distributed networks work together to validate and process transactions, which is essential for the integrity and immutability of the blockchain. There are several blockchain consensus algorithms, such as Proof of work (PoW), proof of stake (PoS), delegated Proof of stake (DPoS), Proof-of-Authority (PoA), Proof of elapsed time (PoET), and Practical Byzantine Fault Tolerance (PBFT).

The blockchain-based credit assessment modelling [7,37,38] considers data privacy protection issues. The blockchain is introduced for storing credit data, which ensures full data traceability of the credit scoring process [39], while the consensus mechanism is used to assess whether the credit data is stored according to a predefined set of rules.

Walambe et al. [40] proposed a system that leverages blockchain’s secure and immutable nature to store machine learning model explanations for credit scoring. The proposed system aims to enable local interpretations of the global model to be publicly available to customers to access securely. The authors demonstrated the trustworthiness of an explained model prediction, with the security, reproducibility, traceability and transparency of blockchain, providing the end-user with a way to securely request an explanation for the credit-scoring decision. Blockchain tamper-proof characteristics ensure the authenticity of the data and minimise the impact of false data for credit evaluation modelling [37]. The blockchain-based framework that assists the gathering of information about the customers from the various financial institutions and calculates their score based on the consensus of multiple institutions improves the credit decision process [41]. For example, blockchain for the credit evaluation system of traders in the food supply chain has been analysed [42].

A blockchain-based credit score evaluation is proposed to ensure transparency in the lending process [7,43]. Blockchain and Decentralized Credit Scoring Model presents a theory to model the optimal investment strategy for different risk vs. return scenarios [7]. In [43] KiRTi, a deep learning-based credit recommender, is proposed to automate loan disbursements and repayments. This work is a step forward in eliminating the requirement of third-party credit rating agencies for credit score generation.

Cho et al. [44] designed a Verifiable Credential (VC) model for VC generation and revocation verification for credit scoring data. Blockchain-authorised data [1,45] and model sharing [46] enhances the security of credit reporting.

Nassar et al. [47] proposed a framework based on the principle that critical decisions in complex AI systems must be subject to consensus among distributed AI and XAI agents hosted in trusted oracles. Blockchain can fulfil trustworthy AI requirements for resilience to biases and adversarial attacks. Blockchain provides key features for XAI agents: Transparency and Visibility, Immutability, Tractability and Nonrepudiation and Smart Contracts.

3. Use of explainable AI in credit assessment

In this section, we first describe the overall XAI techniques. Following this, we present an overview of XAI methods and describe their characteristics for the explainable credit assessment. Finally, we describe the integration of the blockchain and XAI.

3.1. Background of XAI

Machine learning (ML) models are predominantly black boxes. The model-agnostic techniques have been developed to explain the predictions of any classifier in an interpretable form. Among this area's best known contributions is the Locally Interpretable Model-Agnostic Explanations (LIME) [48]. LIME constructs locally linear models around the predictions of a model to explain it by approximating it locally with an interpretable model. These contributions fall under model agnostic (MA) and local (L) explanations. Notably, the authors propose algorithms for individual predictions to solve the "trusting a prediction" problem known as the LIME algorithm by approximating it locally with an interpretable mode. Furthermore, the authors proposed a Submodular Pick SP-LIME algorithm to select a set of predictions (and explanations) to solve the "trusting the model" problem via submodule optimisation.

The LIME explanation is obtained by minimising the following objective function:

$$\xi(x) = \arg\min_{g \in G} \quad L(f, g, \pi_z) + \Omega(g) \tag{1}$$

where $L(f, g, \pi_z)$ measures faithfulness of *explanation model* g, in approximating *original model* f in the locality defined by π_z. G represents the class of the of potentially *interpretable* models, while $\pi_z(x)$ is proximity measure between an instance z to x, knows as locality around x. $\Omega(g)$ penalizes the complexity of the explanation g.

SHAP (Shapley Additive exPlanations) value is proposed by Lundberg and Lee [49] for interpreting and understanding the predictions made by the machine learning models. The proposed SHAP values measure the contribution of each feature to a model prediction. The feature importance explanation technique is a form of ranking the importance of each feature in the prediction output by the model to be explained. The SHAP method calculates an *additive feature attribution measures* that satisfies the set of required properties (*local accuracy, missingness and consistency*). The first property *local accuracy* requires explanation model g to at least match the original model f output for a simplified input. The second property *missingness* requires features missing in the original input to have no attributed impact. The third property *consistency* requires that if a model changes such that some simplified input's contribution increases or stays the same regardless of the other inputs, the input's attribution should be consistent. The Shapley value for each feature can be calculated using the following formula:

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} [f_x(z') - f_x(z' \backslash i)] \tag{2}$$

where M denotes a number of all features, f is a model. $|z'|$ is the number of non-zero entries in z', and $z' \subseteq x'$ represents all z' vectors where the non-zero entries are a subset of non-zero entries of x', while $z' \backslash i$ denotes $z' = 0$.

A comprehensive taxonomy of the XAI method is presented in [50–55]. The XAI is a subsection of AI that focuses on the transparency of the AI systems' decision-making. Integrating XAI into cybersecurity intends to improve the AI security system's trustworthiness, interpretability and resilience. The XAI methods to tackle cybersecurity issues have been presented in different areas, such as industrial IoT [56], advanced persistent threats [57], intrusion detection [58] and autonomous driving [59].

The survey of resampling techniques on feature importance in imbalanced blockchain data is presented in [60]. Rajbahadur et al. [61] explored the impact of feature importance measures on the interpretability and stability of the classifiers. The Neural-Backed Decision Trees (NBDT) model [62] trains a decision tree to represent a (deep) neural network and maintains a high level of model interpretability. The authors Hara and Hayashi [63] proposed a Bayesian model selection to improve the model interpretability of tree ensembles.

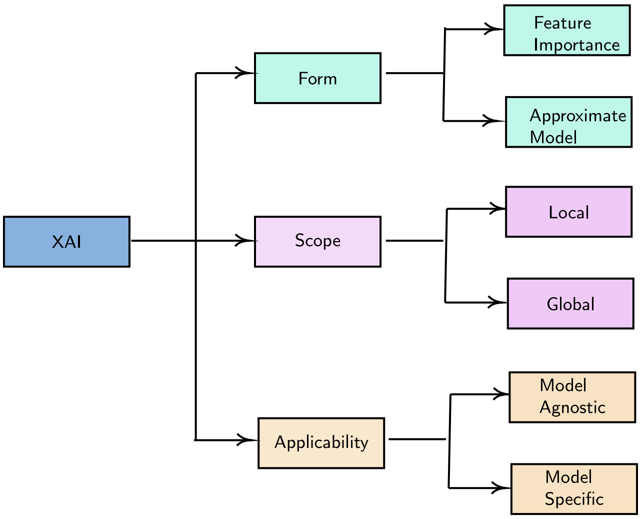


Fig. 4. Overall taxonomy of XAI.

3.2. XAI and credit assessment

The surveyed work on XAI and CA is presented in Fig. 4 and Table 1. Fig. 4 presents the different aspects of classifying XAI methods based on their characteristics. As a result, according to the proposed taxonomy, three main categories for the XAI are identified: *form, scope and applicability*. Another important aspect is the form of the XAI method: numeric or rule-based. Numerical expansion in the form of the importance of a specific feature to the overall performance of a model is called *feature importance* [64]. The explainability produced by rule-based explanations by exploiting several rule-extraction techniques, such as automated reasoning-based models [65–67], is known as *approximate model*. Based on the scope of interpretation, if the method explains a specific instance, it is known as *local*, and if the method explains the whole model, then it is *global*. An important aspect of separating XAI methods is the type of algorithms that could be applied. If the technique has restricted application to a specific family of algorithms, it is called *model-specific*. The method used for any possible algorithms is *model agnostic*. The recent work by Wan et al. [62] presents a Neural backed Decision Tree (NBDT), which explores the combination of neural nets and decision trees. Such an intersection would preserve high-level interpretability while neural networks provide high accuracy. Recent work by de Lange et al. [68] presents the combination of the LightGBM model with SHAP, which enables the interpretation of explanatory variables affecting credit predictions.

Table 1 presents the XAI characteristics and broader applications. XAI in credit risk applications is presented in [6,28,40,68–70].

3.3. Blockchain, federated learning and XAI

Federated Learning trains machine learning models on multiple datasets distributed across different clients without data sharing [46]. FL enables multiple clients to solve machine learning problems under the coordination of the central aggregator, which ensures data privacy [73].

Regarding communication delays, the global model in FL involves multiple iterative rounds of model updates from users, engendering significant communication overhead and incurring additional storage costs during network transmission, [74]. The FL is contingent upon the seamless communication between clients and servers. This communication involves the transmission of local learning models and multiple training iterations for model updates, making communication and training efficiency critical for FL performance [75]. To manage

Table 1
Related work on XAI and credit assessment and their characteristics. Yes (✓), No (×), Partial details on explainability (*).

Reference	Method	Feature importance	Approximate model	Local	Global	Model agnostic	Model specific	Credit application
Moscato et al. [69], Walambe et al. [40] and Bücker et al. [6]	LIME	✓	✓	✓	*	✓	×	XAI for credit score
Moscato et al. [69]	Anchors	✓	✓	✓	*	✓	×	XAI for credit score
Moscato et al. [69] and de Lange et al. [68]	SHAP	✓	×	✓	✓	✓	×	XAI for credit score
Moscato et al. [69]	BEEF	✓	×	×	✓	✓	×	XAI for credit score
Moscato et al. [69]	LORE	✓	*	✓	*	✓	×	XAI for credit score
Ma et al. [71]	MUC	✓	✓	✓	✓	×	✓	Loan application improvements
Srinivasan et al. [70]	ARAEGAN+GM	×	×	✓	×	×	✓	Credit loan denials
Sachan et al. [28]	MAKER	×	×	×	×	✓	×	Loan underwriting
Fahner [72]	TGAMT	×	×	×	✓	×	✓	Explainability by design credit score
Bride et al. [66]	Silas	✓	✓	×	✓	×	✓	XAI via logical reasoning on credit data
Zhang et al. [67]	OptExplain	×	✓	×	✓	×	✓	XAI via logical reasoning on credit data

upstream communication delay, a Sparse Ternary Compression (STC) framework is proposed by Sattler et al. [76] extends gradient sparsification with downstream compression, surpassing federated averaging in various scenarios and advocating for a transformative shift towards high-frequency, low-bandwidth communication in bandwidth-constrained learning environments. Hieu et al. [77] introduced the application of deep reinforcement learning in optimising system parameters for minimising delay, energy consumption and maximising total rewards.

A comprehensive and systematic Privacy-Preserving FL (PPFL) review is presented in [78]. The overview of the main characteristics of the Blockchain-Based Federated Learning (BCFL) framework, architectural design, deployed platforms and feasible applications for BCFL is presented in [79]. Li et al. [80] proposed a systematic study on privacy and security in blockchain-based FL methodologies and discussed the integration of blockchain with FL in various human-centric applications in IoT and intelligent environments.

Blockchain design that enables recording and secure incentives for distributed FL model training via Smart Contracts with Class-Sampled Validation ErrorScheme (CSVES) to validate the quality of gradients to determine reward is proposed in [81]. The advantages of this approach encompass increased trust in the FL process and enhanced incentives for participants during gradient validation. However, potential limitations may arise from centralised model aggregation, lack of explainability of the trained models, and the impact of new data on the training process.

The overview of FL and blockchain integration, called FLchain, can potentially transform intelligent mobile edge computing (MEC) networks into a decentralised, secure, and privacy-enhancing system [82]. The article presents four use cases that demonstrate the potential applications of FLchain in edge networks, including edge data sharing, edge content caching, and edge crowdsensing. However, research lacks a comprehensive evaluation of experiments to assess their effectiveness and limitations in a practical setting fully.

The proposed serverless function for training FL FedLess is detailed in [83], which utilised serverless technologies, AWS Lambda, Azure functions and Openwhisk to enable FL while providing authentification, authorisation and differential privacy. FedLess supports Local Differential Privacy, a technique that adds noise to the data before sharing it. The paper introduces a novel approach that leverages serverless computing to address the challenges of scalability, infrastructure management, and inactive client computing resources in FL. However, it is important to consider the limitations of FedLess in the specific context of the target domain and requirements. Further research and evaluation are needed to fully understand the effectiveness and limitations of FedLess in other areas, such as credit score modelling.

The behaviour attestation method is used to verify the consistency of the behaviour of each participating client during the training process for detecting poisoning attacks in FL [84]. The authors presented the AttestedFL algorithm for defence against untargeted model poisoning attacks in FL with contributions to reducing attack effectiveness, increasing accuracy, pattern-based detection, and flexibility in deployment. However, further research and optimisations are required to explore its efficacy under different scenarios.

Al Mallah and López [85] proposed techniques to address the latency challenges by decoupling the monitoring phase from the detection phase in decentralised FL approaches defences that protect against poisoning attacks in FL. The blockchain replaced the centralised aggregation of the traditional FL. It divided the blockchain network into two types of miners: *minersFL* responsible for FL, and *minersMON*, responsible for monitoring. *Workers* perform the FL and send their local model updates to *minersMON*, responsible for monitoring. The blockchain *minersFL* nodes randomly select a set of reliable *workers* to continue the FL process and calculate the average model using the updated model from the *workers and minersMON*. The proposed design does not store the model updates on the blockchain. Instead, the hash value is written on the blockchain and points towards the model updates. The blockchain stores the commitments of all *workers* on the model updates they worked on. A Merkle tree is used to authenticate the model updates submitted by the *workers*. The proposed approach is designed for resource-contained nodes like mobiles and the Internet of Things (IoT).

Walambe et al. [40] proposed a system that leverages blockchain's secure and immutable nature to store machine learning model explanations for credit scoring. The proposed system aims to enable local interpretations of the global model to be publicly available for customers to access securely. The authors demonstrated the trustworthiness of an explained model prediction, with the security, reproducibility, traceability and transparency of blockchain, providing the end-user with a way to securely request an explanation for the credit-scoring decision. However, the proposed solution considers only a single AI and XAI method: Random Forest (RF) and Locally Interpretable Model-Agnostic (LIME). FL may be regarded as preserving privacy in credit assessments. FL trains machine learning models on multiple datasets distributed across different institutions without data sharing. However, the system's efficiency proposed in [40] depends on the quality of the machine learning model used for credit scoring. Hence, if the model is unreliable, the explanation stored on the blockchain may be inaccurate, leading to incorrect credit scoring decisions. The limitation of the proposed system in [40] is that it does not ensure that the model is reliable and the performance of the model is validated.

Table 2
A comparison of properties of blockchain, AI and integration mechanisms, Yes (✓), No (×), Insufficient details (*).

Reference	Blockchain					AI				Integration			
	Immutability	Traceability	Transparency	Smart contract	Oracles	FL	FL method	XAI	XAI method	Storage	Decentralised ML	Privacy	Implementation
Salah et al. [27]	✓	×	×	×	×	×	*	✓	*	✓	✓	✓	×
Hossain et al. [87]	×	✓	✓	×	×	✓	DL	✓	LIMA	✓	✓	×	✓
Walambe et al. [40]	×	✓	×	✓	×	✓	RF	✓	LIME	✓	×	✓	✓
Patel et al. [43]	✓	×	✓	✓	×	✓	LSTM	×	*	✓	×	✓	✓
Nassar et al. [47]	✓	✓	✓	✓	×	✓	*	✓	*	✓	✓	×	×
Zhang et al. [37]	✓	×	×	×	×	✓	Logit	×	*	✓	✓	✓	✓
Polyviou et al. [36]	✓	×	×	✓	×	×	*	×	*	×	✓	✓	×
Calvaresi et al. [21]	×	×	✓	×	×	×	*	✓	*	×	✓	×	×
Hassija et al. [7]	✓	×	×	×	×	×	*	×	*	PoV	✓	✓	✓
Malhotra et al. [86]	✓	✓	✓	✓	×	✓	SVM	✓	LIME	✓	✓	✓	✓
Verma et al. [13]	✓	*	✓	✓	*	*	*	*	*	*	✓	✓	×
Bellagarda and Abu-Mahfouz [88]	*	*	*	✓	✓	✓	*	✓	*	*	*	✓	×
Yin et al. [78]	×	×	×	×	×	✓	*	×	*	×	✓	✓	×
Chen et al. [89]	×	×	×	×	×	✓	ESB-FL	×	*	PoS	✓		✓
Zhang et al. [73]	×	×	×	×	×	✓	*	×	*	×	✓	✓	×
Cheng et al. [10]	×	×	×	×	×	✓	RL-SecureBoost	×	*	×	×	✓	✓
Srinivasan et al. [70]	×	×	×	×	×	✓	SVM, Naive Bayes	✓	ARAEGAN +GM	×	×	×	✓
Bride et al. [66]	×	×	×	×	×	✓	Silas	✓	Logical Reasoning	×	×	×	✓
Sachan et al. [28]	×	×	×	×	×	✓	BRB	✓	MAKER	×	×	×	✓
Davis et al. [90]	×	×	×	×	×	✓	Optimal Tree, NN, RF	✓	LIME, SHAP, DiCE	×	×	×	✓

Recent research has presented uses of blockchain as a distributed data structure with major features summarised as immutability, transparency and encryption. The integration of blockchain and XAI is presented in Table 2. Integration of the blockchain and XAI can be achieved through decentralised data storage, smart contracts and decentralised model learning [30,31,33]. The primary use of blockchain for AI is for secure data storage [27,37,40,43] and audit trailing of XAI decisions [86].

4. Technologies for integrated credit assessment

In this section, we first describe the overall concepts of credit assessment. Following this, we present an overview of the related work on using XAI and blockchain for credit assessment.

4.1. Credit assessment fundamentals

Credit evaluation assesses a borrower's capacity to become eligible for a loan and the ability to repay. Credit evaluation is the process that assures the development of credit scorecards to assess the creditworthiness of the customers and loan applications following the policy of the lending institution. The banks are looking at making the process efficient, transparent and sustainable to reduce the model risk and provide adequate governance. The increasing competition and growing pressure for revenue generation are requiring banks to explore further effective integration that will result in quicker turnaround time while managing the authenticity of the data and privacy protection.

The book by Thomas et al. [91] has been recognised as a bible of credit scoring and reviews statistical and operational research methods used in building the scorecard. One of the first credit scoring approaches was developed to predict companies' bankruptcy risk [92].

Credit scoring is one of the earliest financial risk management tools [91] and is a method that is used to predict the probability that a borrower will default or become delinquent and to measure the profitability of granting loans. Traditional credit evaluation methods consist of judgmental models, statistical methods, regression analysis [93], discriminant analysis [92,93], logistic models and probit models. Recently, alternative machine learning methods such as artificial neural networks (ANNs) [94], neural networks (NN) [93,95], bayesian networks [93], support vector machines (SVMs) [96], decision trees [93, 95,97,98], XGBoost [99] and other methods have been introduced to build credit scoring models. The fairness of AI techniques in the context of the credit scoring model has been analysed by Hurlin et al. [5].

Credit scoring systems are based on the past performance of customers, similar to those who will be assessed under the scheme. When the customer applies for a loan, the financial institutions collect the customer details, known as application data. Fig. 5 presents an illustrative view of the data flow for the credit assessment process. The application data consists of variables such as the applicant's age, time at current/previous residence, time at current/previous job, housing status, occupation group, income, number of dependents, banking relationship, debt ratio, and credit references. Credit references or bureau information consist of the previous defaults, arrears and the customer's current status on other loans, including the number of enquiries, hardship information and repayment history information. The major credit bureau providers are Equifax, Illion and Experian. The comprehensive credit score is the number that models the data held in the credit bureau and indicates the likelihood of repaying the money to the credit applicant's credit bureau. The bureau has its method for modelling the comprehensive credit score. The customer application data is used to perform the calculations related to the serviceability of the customer, and that information is used in the *scorecard model.*

The credit scoring model uses any characteristic of the customers that aids prediction in the scoring system. The variables are mainly associated with default risks, such as previous defaults or arrears or the customer's current status on other loans and comprehensive credit score. Other variables present the stability of the consumer, such as time at address and time at present employment. A different group of variables gives a view of the consumer's residential status, spouse's employment, number of children, and number of dependents. A separate set of variables shows the consumer's serviceability, such as the Debt to Income ratio.

The good/bad flag is created based on the loan repayment history of the accepted population for the scorecard development. Borrowers who have missed payments or gone past a certain number of days, usually 90 days, are categorised as "bad" borrowers, while those who have not are classified as "good" borrowers. The good and bad flags are then used to develop a scorecard model. The Kolmogorov–Smirnov statistic determines the cut-off score and measures the distance between the cumulative distribution of goods and bads. The cut-off score is the

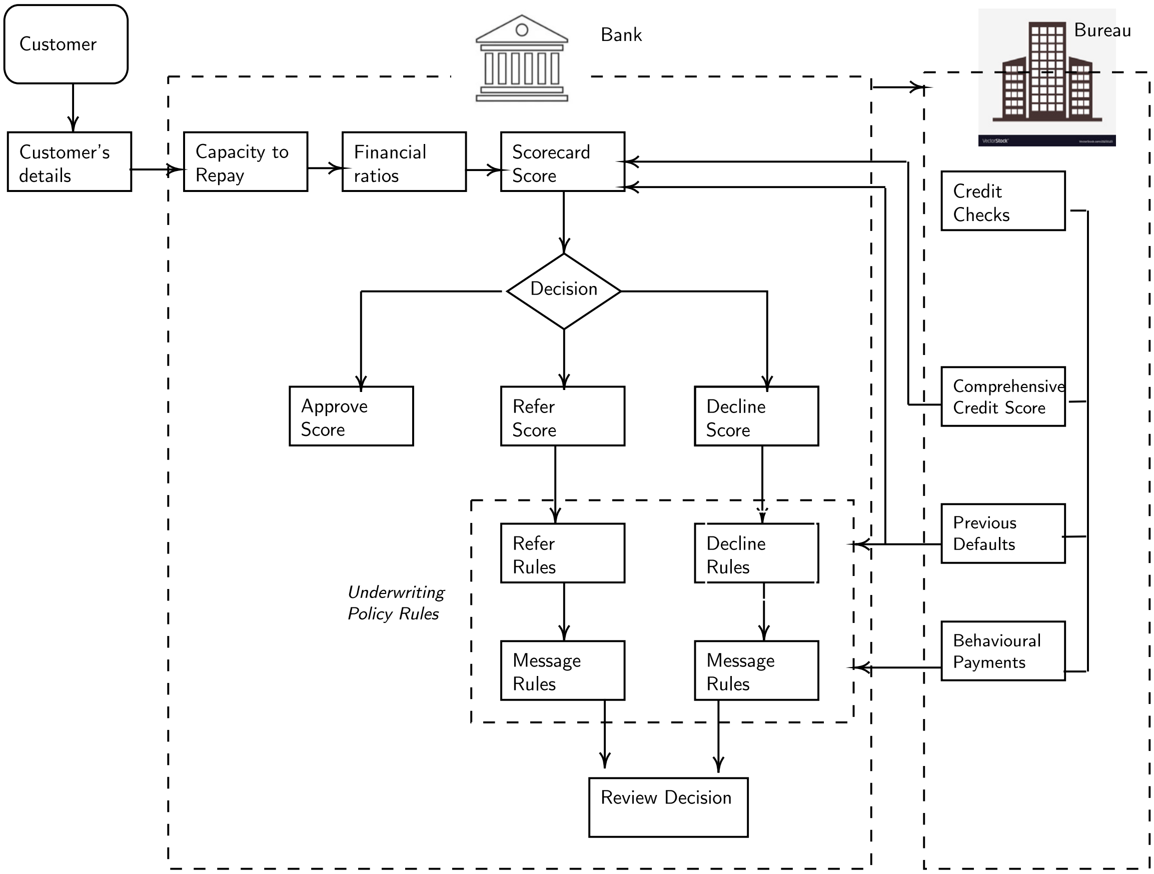


Fig. 5. Overview of the credit assessment process.

maximum distance between the distribution of good/bad and is used to predict the good/bad [100]. If the score of a customer is above or equal to the cut-off score, then the customer is predicted as a good borrower otherwise, a bad borrower. Subsequently, the scorecard is applied to the rejected population to predict good/bad, known as reject inference [101].

A loan underwriting process or *Underwriting Policy Rules* evaluates the information in a loan application following the scorecard cut-off score outcome and the policy of the lending institution as shown in Fig. 5.

A loan underwriting system containing coded underwriter guidelines decides acceptance or rejection when specific default rules in the rule base are triggered. The loan underwriting could be manual or automated. Manual underwriting refers to processing non-standard (higher risk) loans. The underwriting system consists of a codified set of rules based on the policy of the lending institution to assist in a final lending decision. The key limitation in the existing literature is that the credit scoring and underwriting process have been considered in isolation, while the automated intelligent credit evaluation should consider both.

Sachan et al. [28] proposed an XAI decision-support system to automate loan underwriting by a belief-rule-base (BRB) system. The solution proposed by the authors aims to enhance the efficiency and accuracy of the underwriting process while preserving transparency and fairness.

An intelligent credit risk scorecard approach based on statistical principles is needed for specific business objectives like predicting losses better [102]. A deeper view of creating, evaluating, and monitoring scorecards is presented in [101].

Credit scoring is a supervised learning problem. Specifically, it is a binary classification problem aiming to classify good and bad borrowers [100]. A systematic literature survey approach to statistical and machine learning models in credit scoring, identifying literature limitations, proposing a guiding machine learning framework and pointing to emerging directions have been proposed by Dastile et al. [100]. However, the LIME method covers the explainability of credit scoring methods to a limited extent.

The most popular technique in credit scoring modelling is Logistic Regression Eq. (3). The Logistic Regression assumes a linear relationship between the log of probability odds and inputs [103]. The logistic regression is sensitive to the correlation between the predicted variables. Thus, it should be ensured that no correlated variables are in the regression set. Logistic regression is the log of the probability odds by a linear combination of the input variables.

$$log(\frac{p}{1-p}) = w_0 + \sum_{i=1}^{m} w_i X_i \tag{3}$$

where p represents the proportional response, w_0 is the intercept, when $X = 0$ intercept is the log of the odds of having the outcome. $\boldsymbol{X}_i$ are application characteristics and weights $\boldsymbol{w}_i$ are the score of the characteristics.

Eq. (3) is considered a linear regression of the non-linear function of the probability of being a good customer. The score $s(x)$ of the scorecard presented in Eq. (3) is the following Equation:

$$s(x) = w_0 + \sum_{i=1}^{m} w_i X_i \tag{4}$$

Another important technique in credit scoring modelling is non-linear regression, known as a *probit* analysis. The *probit* model $\boldsymbol{N}(x)$ is given as the cumulative normal (standard Gaussian) distribution function defined below:

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \tag{5}$$

Table 3
A comparative analysis of the use of XAI and Blockchain (BC) for credit assessment: Yes (✓), No (×), Minor advancement (*).

Reference	XAI for CA	BC for CA	XAI and BC for CA	Key technologies	Considerations
Qiao et al. [38,107]	×	✓	*	PHE, SMPC algorithm	BC for privacy and data security
Hassija et al. [7]	×	✓	*	Prospect theory for risk vs. return	BC for secure lending
Walambe et al. [40]	✓	✓	✓	RF method, LIME	XAI with BC to store explanation in block
Patel et al. [43]	×	✓	*	LSTM	BC to update the credit score
Zhang et al. [1]	×	✓	*	PBFT consensus	Consortium BC for CA
Yang et al. [39]	×	✓	*	BACS	BC and AutoML for CA
Yang et al. [46]	×	✓	*	IPFS to store encrypted data	BC and FL for data sharing in CA

The goal is to estimate $\boldsymbol{N}^{-1}(p_i)$ as a linear function of the characteristics of the applicant, as follows:

$$N^{-1}(p_i) = w_0 + \sum_{i=1}^{m} w_i X_i \tag{6}$$

The value of $\boldsymbol{N}^{-1}$ indicates that the customer is good if the score is above a certain level. Linear programming is used as a classification approach for scorecard modelling.

The popular machine-learning techniques in credit scoring are Random Forest, Artificial Neural Networks, and Convolution Neural Networks. Random Forest (RF) is an ensemble of decision trees [104], such that K decision trees are built on different bootstrap samples of the data.

The traditional credit risk assessment process utilises the application data, while dynamic transactional data has recently been used to evaluate the credit application [105]. The authors proposed cost-sensitive multiple-instance learning (CSMIL) to build a credit scoring model incorporating customers' dynamic transactional data and static/personal information. This study is the first to apply a CSMIL model to credit risk assessment and considers the impact of dynamic transactional data and time-series information. The work presented in [105] is limited as it does not include explainability techniques in the credit scoring model while utilising dynamic transactional data. Furthermore, the model performance may deviate due to the data update; identifying those deviations may require model recalibration.

P2P lending is a business model involving borrowers, lenders, and a P2P platform. A P2P platform generally has a large number of users and frequent transactions. A benchmarking study of some of the most used credit risk scoring models to predict if a loan will be repaid in a P2P platform has been analysed by Moscato et al. [69]. The authors compared the obtained outcomes concerning the state-of-the-art approaches and also evaluated them in terms of their explainability through different XAI tools. Zhang et al. [106] proposed a new online integrated credit scoring model (OICSM) for P2P lending that integrates gradient-boosting decision trees and the neural network to make the credit scoring model handle two types of features (numerical and categorical) more effectively and update the model online. This is one of the first experiments considering the problem of the credit scoring model online update to avoid prediction deviation. The limitation of the OICSM scoring model is that it does not include XAI techniques to ensure transparency in the credit scoring model. Furthermore, the traceability of model updates is not considered.

4.2. *Blockchain and XAI for credit assessment*

Credit assessment requires an efficient, transparent, traceable, secure, and sustainable process to reduce the model risk and provide adequate governance. The surveyed work on integrating CA, blockchain and XAI is presented in Table 3. Table 3 presents limited work that has been done to examine the integration of blockchain and XAI for the credit assessment process. As discussed previously, the system proposed in [40] relies on the quality of the machine learning model used for credit scoring. If the model is unreliable, the explanation stored on the blockchain may be inaccurate, resulting in incorrect credit scoring decisions. The proposed mechanism lacks a technique to ensure that the model is reliable and its performance is validated, thus limiting its effectiveness.

The efficiency of the Credit evaluation has been addressed in the BACS scheme by Yang et al. [39]. The BACS scheme consists of credit data storage to the blockchain to ensure traceability. The random forest model effectively integrated the critical steps of credit data feature extraction, feature selection, credit model construction, and model evaluation. Blockchain technology as discussed in this article requires a consensus mechanism to determine whether credit data is stored and used within predefined rules. The consensus process is divided into the sorting service and the synchronised ledger. This work by Yang et al. [39] has a few limitations. Firstly, the paper does not explore alternative blockchain platforms beyond Fabric Hyper-ledger for ensuring consensus on updates to the model. Additionally, it does not consider XAI methods other than consensus for improving transparency in credit decision outcomes. Finally, the study relies solely on historical credit data and does not explore the potential benefits of using transaction data to identify early delinquent behaviour. Thus, identifying changes in data, model input assumptions, or scorecard model performance may also have limitations that have not yet been explored.

4.2.1. *Case studies*

Authors in [8] introduced an explainable federated learning and blockchain-based credit scoring system to address credit model sharing challenges and ensure data privacy. Their approach explains the FL mechanism, proposing a Decentralized Byzantine fault-tolerant Stochastic Gradient Descent algorithm (D-SGD). Mathematically, the study combines the Shapley value with Delegated Proof of Stake (DPOS) for a consensus protocol. The algorithm calculates the contribution values of the parties in the execution of the federated algorithm. Evaluation of the proposed Explainable Federated learning and blockchain-based Credit scoring System (EFCS) includes simulations and experiments using the "Give Me Some Credit" dataset from Kaggle. The dataset contains 150,000 credit card payments and income-related data, with 10,026 default customers. The performance assessment encompasses accuracy, precision, recall, F1 score, and AUC. The modelling process involves the coordinating party calculating contributions and recording them in the current block of transactions. Training results reflect aggregated data source outcomes, with each participant iterating locally 20 times before sending the gradient. Increased participants lead to longer training times due to heightened communication overhead and computational intensity in FL. The EFCS is evaluated using six credit datasets from traditional financial institutions and peer-to-peer lending platforms. The datasets from Germany, Taiwan, and Australia are available through the UCI machine learning repository. Additionally, P2P lending datasets and credit card datasets are employed for further validation. Specifically, two P2P datasets are collected from China's pioneering P2P lending platforms. This diverse set of datasets from various sources enhances the applicability of the evaluation process for EFCS.

Table 3 presents case studies of use cases of the XAI and blockchain for credit assessment.

5. Proposed conceptual framework

This section presents a conceptual framework based on decentralised blockchain as a solution to induce model verification, behavioural reliability and explainability for intelligent credit assessment.

This framework uses a blockchain-based FL solution to enable AI machine model learning and verification of the methods, and it is a machine-learning model built on distributed datasets. FL benefits blockchain with aspects of privacy-preserving data exchange. Our proposed conceptual framework considers blockchain-based FL through the consortium or private blockchain platform.

Robust integration of technologies can be defined as facilitating decentralised model learning, verification, and model aggregation on distributed multisource datasets. This involves fostering collaboration among different sources while preserving data privacy and enhancing the overall reliability and transparency of the credit scoring system. This definition effectively captures the essence of robust integration in the proposed conceptual framework, highlighting key aspects such as decentralised learning, verification, collaboration, data privacy, reliability, and transparency, leading to a trustworthy credit scoring decision-making process.

Blockchain enables FL to enhance the process of the global model aggregation such that model aggregation is to be computed by the nodes and miners, leaving the central aggregation unneeded.

Fig. 6 illustrates the architecture of the proposed intelligent automated credit assessment that enables AI model learning and blockchain miner verification of the model while ensuring privacy is protected. The architecture consists of the following roles:

- Scorecard Clients
- Federated Learning Local Model Miners
- Blockchain Miners verify the models and generate XAI models
- Serverless Aggregation Node(s) for the Global Model
- Distributed ledger for the Global Model, Local Model Updates, Re-train model

Our conceptual framework considers the architecture of *Flexible Couple Blockchain-based Federated Learning (FIC-BCFL), presented in* [32]. The architecture of the FIC-BCFL indicates the clients are responsible for collecting and training local models. The miners of the blockchain perform the verification of the local model updates. The FL can ensure the parameters of the models are stored on the blockchain, and the blockchain miners perform the aggregation of the global model.

Our conceptual proposal considers the *Predefine Nodes* performing the model training for FL. Those nodes are authorised to perform the model learning and are equipped with computational powers and storage to receive the data and train local models. Furthermore, our framework is an extension to FIC-BCFL as it incorporates XAI as well.

Specifically, future implementation will consider (1) IPFS for distributed storage of data, the global AI and XAI models' parameters, (2) Consortium blockchain for the system's logic and state, and (3) FL for the AI model learning.

5.1. Scorecard clients

Banks use various channels to gather data for credit scoring, such as credit reports from bureaus, loan applications, and income verification. They may also analyse bank account transactions, review public records for legal information, and take behavioural data into consideration when assessing creditworthiness. Some banks even explore social media and online presence. Credit scoring models weigh factors differently to calculate credit scores, incorporating information on payment history, outstanding debts, and financial behaviours. It is important for banks to comply with privacy regulations throughout the data collection process to ensure the protection of individuals' sensitive information.

A good credit assessment consists of multisource data, such as banks', bureaus' and enterprises' data. Multi-party data enables a broader platform to provide a comprehensive model learning foundation for a good credit evaluation system. The technology we consider is the consortium blockchain, which ensures a strict access mechanism. The participating nodes from banks, bureaus and enterprises are required to obtain a user's certificate for access to data. Specifically, scorecard clients are the distinct client functions that will collaborate in FL. Specifically, Hyperledger Fabric incorporates the ciphertext-policy attribute-based encryption (CP-ABE) access control scheme avoiding unauthorised access, [108]. All nodes in the Hyperledger Fabric network are generally assumed to be credible. As a result, the consensus mechanism employed by Fabric mainly focuses on ordering transaction proposals rather than validating them, [108].

In [109] presented ZeroTrustBlock, a comprehensive blockchain framework for secure and private health information exchange using the Hyperledger Fabric. The architecture and consensus protocols are designed to comply with security and confidentiality regulations.

Our proposed system uses FL and blockchain to enable the aggregation of information about customers without compromising customer privacy. To ensure privacy-preserving features in credit score modelling, our proposed framework utilises federated model learning, incorporating credit application data, customer transaction data, and credit bureau information. The original data associated with providers, banks, credit bureaus and enterprises are hashed, and the associated hash will be stored on the blockchain. Privacy protection will be achieved using the SHA256 hash algorithm. Our framework process considers the data is stored off-chain, and the hash of the data is stored on the blockchain to form a unique index to identify the corresponding off-chain data. Raw data is not shared in our proposed framework. Only the model, model parameters, accuracy, and updates will be shared. Data sharing is a common use case for IPFS due to its high availability and good performance, [110], hence supporting the idea of using IPFS in our proposed framework.

5.2. Role of Local Model Miners

Our proposed conceptual framework considers the *Predefine Nodes* performing the model training for FL. Those nodes are authorised to perform the model learning and are equipped with computational powers and storage to train to receive the data and to train local models. The nodes associated with the clients are randomly selected to perform the model training. The clients define the initial models. The nodes train the models on the local data and upload hash local model parameters in the on-chain blockchain. The client node updates the model parameters in off-chain IPFS for the same on-chain hashed local model parameters. The hash data and hash model generated locally will be stored on the blockchain and maintained on-chain.

To ensure the privacy of the training model is achieved, we will use Paillie's Cryptosystem, which is homomorphic encryption used in distributed machine learning.

5.3. Generation and verification of XAI models

The miners *validate* the local models by invoking the Smart Contract through an oracle to access the model parameters value in the table off-chain. The smart contact queries IPFS for the model parameters with the same on-chain hashed model parameters. The local model's *authenticity* is confirmed by training the selected model on its local data. Miners use the smart contract to invoke a pair of the local hash data and local hash model to obtain data and model parameters from off-chain IPFS. Off-chain data is used to train and compare the model with the authentic model. A range of different AI and XAI models will be considered, such as Logistic Regression (LR), Random Forest (RF), RidgeClassifier, GaussianNB, and SGDClassifier.

Most of the learning techniques in machine learning belong to *Non-convex Training*. Training neural networks can pose challenges, particularly due to factors like sensitivity to initialisation, step sizes, mini-batching, and optimiser selection. As a result, close monitoring and interpretation of the model's learning process are crucial due to its intricate black-box nature. To have model parameters representing the

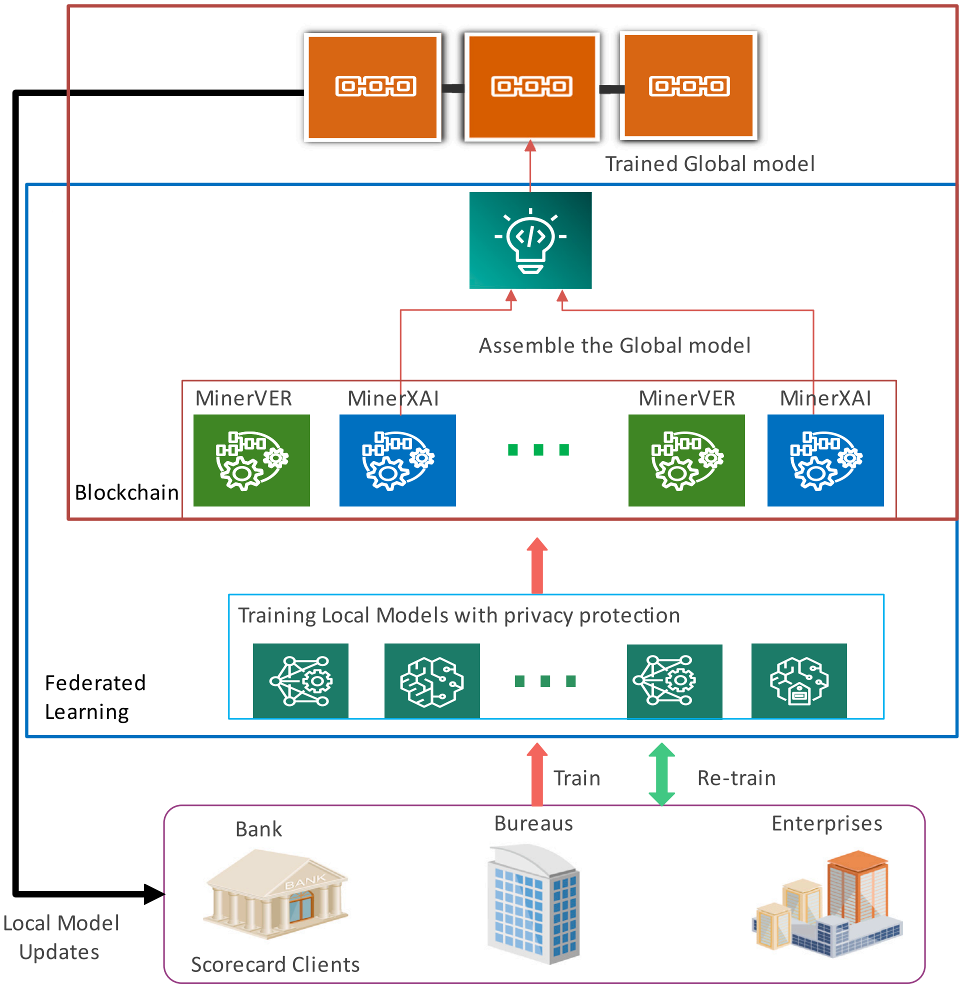


Fig. 6. Proposed conceptual framework.

global solution, it is necessary to use *Convex optimisation*, similar to the study presented in [111].

$$\min_{u} C(\boldsymbol{u}) \quad \text{s.t.} \quad \varphi_i(\boldsymbol{u}) \leq 0, \qquad \psi_j(\boldsymbol{u}) = 0, i = 1, \dots, l, j = 1, \dots, m \tag{7}$$

where C is the cost function and $\boldsymbol{u} \in \mathbb{R}^H$ is the optimisation (control) variable. The functions $C, \varphi_1, \dots, \varphi_l$ are convex while the functions $\psi_1, \dots, \psi_m$ are affine [112].

Selected miners will generate the XAI for the respected models. Our work will consider the proposed framework presented in [85]. This decoupled the monitoring phase from the detection phase in defence against poisoning attacks and replaced the *centralised Federated Learning - chief* with the *workers* that collaborate to train the global model. Al Mallah and López [85] proposed techniques to address the latency challenges by decoupling the monitoring phase from the detection phase in decentralised FL approaches defences that protect against poisoning attacks in FL. The blockchain replaced the centralised aggregation of the traditional FL. It divided the blockchain network into two types of miners: *minersFL* responsible for FL, and *minersMON*, responsible for monitoring. *Workers* perform the FL and send their local model updates to *minersMON*, responsible for monitoring. The blockchain *minersFL* nodes randomly select a set of reliable *workers* to continue the FL process and calculate the average model using the updated model from the *workers, and minersMON*. The proposed design does not store the model updates on the blockchain. Instead, the hash value is written on the blockchain and points towards the model updates. The blockchain stores the commitments of all *workers* on the model updates they worked on. A Merkle tree is used to authenticate the model updates submitted by the *workers*.

Hence, in our conceptual framework, we will consider decoupling the miners performing the model verification *minerVER* from those miners responsible for generating model explanation named *minerXAI*. This improvement will enable reliability, transparency and explainability of the credit assessment model. The minerVER validates models using a k-fold cross-validation technique. Regarding model reliability, each local model obtains its own set of metric functions after being trained. These metric functions are used to evaluate a model's performance based on specific objectives, and they play a critical role in assessing prediction errors. The SHAP model for explainability is used.

5.4. Functions of serverless aggregation nodes

We will use serverless FL and blockchain to enhance the process of the global model aggregation without a centralised aggregator. The serverless aggregation utilises the cloud provider to ensure the scalability of the proposed solution. Our approach will employ serverless computing, FL, and blockchain to enable privacy-preserving, decentralised machine learning. In this approach, each client runs a local machine learning model using serverless computing, and the updates from each client are securely aggregated using blockchain. This approach will enable a decentralised and secure model training process without needing a central server or data aggregator.

Grafberger et al. [83] presented the workflow for training multiple clients using FedLess in a single FL round. The FL admin selects the model, registered client functions, and hyperparameters. The FedLess controller requests a new invocation token from the Auth Server and uses it along with the credentials to access the parameter server to invoke the clients randomly selected for this round. The clients validate the signature and authorisation of the token and load the latest global model from the parameter server before performing local training, optionally using Local Differential Privacy (LDP). Once training is finished, the clients upload their parameters to the parameter server. The FedLess controller waits until all clients have completed training and starts the model aggregation by invoking the *aggregator function*. The aggregator loads the client results, aggregates the parameters, and stores the new global model. Finally, the controller starts the evaluation, either using the global test set or invoking a new selection of clients to evaluate their test set. It aggregates the returned metrics to resume the training process. Specifically, we will consider the FedLess [83] framework and extend its average model aggregation Federated Averaging (FedAvg) with optimisation. The serverless computing platform to be used is AWS Lambda [113]. The hash of the global model is to be stored on the blockchain. The miners who performed the verification were randomly selected for the global model aggregation and assembly. Federated Averaging (FedAvg) is the most common model aggregation technique in FL proposed by McMahan et al. [114], based on averaging the model weights across all clients.

Credit assessment requires an efficient, transparent, traceable, secure, and sustainable process to reduce the model risk and provide adequate governan

Convex optimisation is a branch of mathematical optimisation focused on problems where both the objective function and the constraints are convex. It deals with finding a convex function's minimum (or maximum) over a convex set. A set $\Omega \in \mathbb{R}^n$ is convex if, for all $\mathbb{x}$ and $\mathbb{y}$ in Ω and for all λ in $[0,1]$ it holds $\lambda\mathbb{x} + (1-\lambda)\mathbb{y}$.

A convex function has the property that the line segment between any two points on the function lie above the function itself. Formally, it is defined by the following theorem.

Lemma 1 (*[112]*)**.** *A function* $\mathbb{f} : \mathbb{R}^n \rightarrow \mathbb{R}$ *is convex if, for all* $\mathbb{x}$ *and* $\mathbb{y}$ *in the domain of* $\mathbb{f}$ *and for all* λ *in the interval* $[0,1]$*, the following holds*

$$\mathbb{f}(\lambda\mathbb{x} + (1-\lambda)\mathbb{y}) \leq \lambda\mathbb{f}(\mathbb{x}) + (1-\lambda)\mathbb{f}(\mathbb{y}) \tag{8}$$

The main property of convex optimisation is its ability to guarantee a global minimum, meaning it is possible to find the best solution to a the problem rather than just a local minimum.

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{p}^T\boldsymbol{w} - \tfrac{1}{2}\boldsymbol{w}^T Q\boldsymbol{w} \\ \text{s.t.} \quad & G\boldsymbol{w} \leq \boldsymbol{h}, \quad A\boldsymbol{w} = \boldsymbol{b} \end{aligned} \tag{9}$$

where, $\boldsymbol{p}$ is the mean accuracy of all accuracy types for each local model n-dimensional vector. Q is $n\times n$ covariance matrix of local model accuracies that consists of the accuracy classification score metrics used to measure the classification performance of considered classification models. A is $m \times n$ real matrix, G is $m \times n$ real matrix, $\boldsymbol{b}$ is a real-valued m-dimensional vector. Quadratic programming aims to find an n-dimensional vector $\boldsymbol{w}$ to meet the imposed constraints. The variable $\boldsymbol{w}$ in our framework symbolises the weights allocated to each model, reflecting their significance derived from the accuracy of local data. The weights assigned to each model emphasise their performance, contributing to a compelling ensemble that enhances the overall predictive power of the system. The Algorithm 1 presents local model training. The assembling of the global model is formalised in Algorithm 2.

Regarding the practical implementation of the Algorithm 2, time complexity is critical. The time complexity $T(n)$ of our proposed Algorithm 2 is linear to the training time of ML models. For example, if the chosen model is Random Forest, then the time complexity would be $O(N \cdot m \cdot log\ m \cdot d \cdot k)$, where m is the number of training samples, d is the dimension (number of features), k is the number of trees, and N is the number of local models. We assume that the input to the convex quadratic programming problem is much smaller than the size of the dataset, so its complexity is subsumed by the training time. The algorithm for assembling the global model based on the optimisation will follow specifications defined in the smart contract.

Algorithm 1 Local Model Training Algorithm

```
procedure TRAIN LOCAL MODEL $\Lambda_i(D_i)$
    Load data set $D_i$ that includes local features and labels
    Validate data
    Select type of classification model $\Lambda_i$
    Tuning the hyper-parameters of a model $\Lambda_i$
    Train a local model $\Lambda_i$
    Evaluate a vector of model accuracy metric $\mu_i$
    Evaluate a mean value $\bar{\mu}_i$ of a vector $\mu_i$
    return $\Lambda_i$, $\mu_i$, $\bar{\mu}_i$
end procedure
```

Algorithm 2 Global Model Aggregation Algorithm

```
procedure AGGREGATE GLOBAL MODEL $\Gamma$
    Request $N$ local models
    for $i \leftarrow 1, N$ do
        $\Lambda_i$, $\mu_i$, $\bar{\mu}_i \leftarrow$ TRAIN LOCAL MODEL $\Lambda_i$
    end for
    Concatenate accuracy vectors into matrix $X$
    for $i \leftarrow 1, N$ do
        $X \leftarrow concat(X, \mu_i)$
    end for
    Create matrix $Q \leftarrow X^T X$
    Create vector $p \leftarrow [\bar{\mu_1}, \dots, \bar{\mu_N}]^T$
    Calculate consensus weights $\boldsymbol{w}$

  $\max_{\boldsymbol{w}}$   $\boldsymbol{p}^T\boldsymbol{w} - \frac{1}{2}\boldsymbol{w}^T Q\boldsymbol{w}$
  s.t.   $G\boldsymbol{w} \leq \boldsymbol{h}$,   $A\boldsymbol{w} = \boldsymbol{b}$

    Compose a global model $\Gamma \leftarrow w_1\Lambda_1 + \ldots + w_N\Lambda_N$
    return $\Gamma$
end procedure
```

The uniqueness of our proposed approach lies in its non-iterative and parallel nature, suggesting potential efficiency gains over traditional iterative methods. An extension to our framework incorporates an integrated evaluation process, wherein local model prediction accuracy directly contributes to the assembly of the global model. To augment overall model accuracy, we propose an additional enhancement involving utilising XAI model input impact measures and an accuracy matrix during the global model assembly. This extension aims to provide a more comprehensive and accurate credit assessment mechanism.

5.5. Model distribution and retraining

The assembly of the global model will occur as an off-chain process. The hash global model parameters and weights will be stored on the blockchain. Similar to Li et al. [33], we consider the Committee Consensus Mechanism blocks to store the global model and local updates. Communication-based generated mechanisms reach an agreement before appending blocks. Selected nodes will validate the updates.

All clients can download the global model parameters and weights from the blockchain and continue to use them in the next round of learning models.

As the client's data changes, the models may need to be updated. Therefore, retraining will be performed as a fit method for new data.

At the same time, the original model parameters are to be used as a starting point in the retraining process. Re-training of the model will occur once a change of the statistical properties is detected, such as a change in the Population Stability Index (PSI).

6. Discussion and analysis

Table 4 presents a comprehensive list of key characteristics required to build reliable credit scoring modelling. Specifically, credit score modelling requires diverse data to enable collaborative modelling while ensuring privacy, transparency, and fairness. Our research proposes a novel conceptual framework that integrates these elements previously studied in isolation. Some previous research has explored the use of the blockchain in the context of credit data sharing [1,45], credit evaluation [37,115] and storing explanation on the blockchain [40]. However, our motivation aligns with a similar study [85], which discussed the use of the blockchain for model authentication. While our unique research focus is on the importance of model explainability for the specific credit score modelling application, which incorporates privacy-preserving decentralised model learning combined with reliability, transparency, and explainability features of the blockchain miners.

The authors in [85] used the blockchain to develop an immutable framework for decentralised, federated model learning. The Merkle tree was utilised to store the local model updates to verify the validity of the model updates. However, the study does not consider the explainability of the models and specific application of credit scoring.

The system proposed by Walambe et al. [40] relies on the quality of the machine learning model used for credit scoring. If the model is not verified, the explanation stored on the blockchain may be inaccurate, resulting in incorrect credit scoring decisions. The proposed mechanism lacks a technique to ensure diverse data is used in a decentralised FL model, and its performance is validated, thus limiting its effectiveness.

The serverless function for training Federated Learning FedLess is detailed in [83], which utilised serverless technologies, AWS Lambda, Azure functions and Openwhisk to enable multisource FL while providing model aggregation. However, it is important to consider the limitations of FedLess in the specific context of model verification and explainability in the domain of credit scoring. Further research and evaluation are needed to fully understand the effectiveness and limitations of FedLess in other areas, such as credit score modelling.

The work presented in [1] explored the use of blockchain in credit data sharing. However, this research is limited as it does not include model learning, verification, aggregation and explainability techniques in the credit scoring model while utilising dynamic multisource data.

A method to validate the quality of FL model gradients and to determine reward is proposed in [81]. The advantages of the proposed approach encompass increased trust in the FL process and enhanced incentives for participants during gradient validation. However, potential limitations may arise from centralised model aggregation, lack of explainability of the trained models, and the impact of new data on the training process in the domain of credit scoring.

The proposed conceptual framework identifies the need for combining different technologies to ensure model verification, behavioural reliability, and model explainability for intelligent credit scoring. Specifically, our framework uses a Blockchain-based Federated Learning solution to enable decentralised model learning, verification of the models and model aggregation on distributed multisource datasets.

Risks:. Implementing blockchain technology and FL for credit assessment has vulnerabilities that require attention. Regulatory compliance is a primary challenge that requires constant monitoring of evolving regulations and ensuring data privacy, model explainability and reliability across all parties involved. Technologies also face data privacy issues that demand careful management of sensitive information and privacy-preserving techniques. Interoperability and scalability concerns may arise when integrating these technologies with existing financial systems. The real-world challenge of achieving explainability and model interpretability persists due to the decentralised and collaborative nature of FL and the limited clarity of blockchain transactions. To mitigate these risks, it is essential to focus on robust security measures, careful technological design, and ongoing collaboration with industry stakeholders and regulators. By doing so, we may ensure that these technologies can be safely and securely integrated into existing systems while maintaining data privacy, model explainability and regulatory compliance.

7. Conclusion

Our research investigates the core features of XAI, blockchain, and credit scoring. Specifically, we examine recent efforts to integrate XAI, blockchain, and FL for credit scoring and identify limitations in these approaches. While these solutions primarily focus on enhancing data storage security and privacy, we identify the need for combining these technologies to ensure model verification, behavioural reliability, and model explainability for intelligent credit assessment. To address those challenges and create a reliable and explainable credit scoring process, we propose a novel framework that leverages the benefits of blockchain and FL. Our framework's distinctiveness lies in its holistic design, which incorporates privacy-preserving decentralised model learning coupled with the reliability, transparency, and explainability features of the blockchain.

We have thoroughly examined the qualitative features necessary for designing an efficient credit scoring framework. In our future work, we will employ the framework to quantify and evaluate the effectiveness of the proposed architecture, including communication delay in a real environment.

Our proposed framework has certain limitations that could be addressed and improved upon in future research. The study is based on Hyperledger Fabric, which is a suitable platform for credit scoring applications due to its scalability and compliance features. Its modular architecture and permissioned blockchain model facilitate efficient workload distribution and make it well-suited for scaling up to meet the demands of growing networks. By optimising smart contracts, computational overhead can be minimised to ensure effective transaction processing and contract execution. Its features, such as private channels and access controls, align with regulatory requirements for data privacy and confidentiality in credit scoring. Its interoperability and auditability features also support seamless integration with external systems and compliance with financial regulations, making it a reliable choice for building secure and scalable credit scoring applications. Implement privacy-preserving techniques like zero-knowledge proofs and homomorphic encryption to protect sensitive data. However, improvements related to performance, advanced cryptography, and real-world pilot testing will be addressed in future work.

CRediT authorship contribution statement

Zorka Jovanovic: Writing – original draft, Methodology, Formal analysis. **Zhe Hou:** Writing – review & editing, Supervision, Conceptualization. **Kamanashis Biswas:** Writing – review & editing, Supervision, Conceptualization. **Vallipuram Muthukkumarasamy:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Table 4
Comparison of credit assessment requirements among different models: Yes (✓), No (×).

Reference	Multisource	Model verification	Model aggregation	XAI model	Credit application
Al Mallah and López [85]	✓	✓	✓	×	×
Walambe et al. [40]	×	×	×	✓	✓
Grafberger et al. [83]	✓	×	✓	×	×
Zhang et al. [1]	✓	×	×	×	✓
Martinez et al. [81]	✓	✓	✓	×	×
Proposed approach	✓	✓	✓	✓	✓

References

[1] J. Zhang, R. Tan, C. Su, W. Si, Design and application of a personal credit information sharing platform based on consortium blockchain, J. Inf. Secur. Appl. 55 (2020) 102659, http://dx.doi.org/10.1016/j.jisa.2020.102659.

[2] M. Óskarsdóttir, C. Bravo, C. Sarraute, J. Vanthienen, B. Baesens, The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics, Appl. Soft Comput. 74 (2019) 26–39.

[3] M. Hurley, J. Adebayo, Credit scoring in the era of big data, Yale JL Tech. 18 (2016) 148.

[4] X. Dastile, T. Celik, H. Vandierendonck, Model-agnostic counterfactual explanations in credit scoring, IEEE Access 10 (2022) 69543–69554.

[5] C. Hurlin, C. Pérignon, S. Saurin, The fairness of credit scoring models, 2022, arXiv preprint arXiv:2205.10200.

[6] M. Bücker, G. Szepannek, A. Gosiewska, P. Biecek, Transparency, auditability, and explainability of machine learning models in credit scoring, J. Oper. Res. Soc. 73 (1) (2022) 70–90.

[7] V. Hassija, G. Bansal, V. Chamola, N. Kumar, M. Guizani, Secure lending: Blockchain and prospect theory-based decentralized credit scoring model, IEEE Trans. Netw. Sci. Eng. 7 (4) (2020) 2566–2575.

[8] F. Yang, M.Z. Abedin, P. Hajek, An explainable federated learning and blockchain-based secure credit modeling method, European J. Oper. Res. (2023).

[9] A. Imteaj, M.H. Amini, Leveraging asynchronous federated learning to predict customers financial distress, Intell. Syst. Appl. 14 (2022) 200064.

[10] K. Cheng, T. Fan, Y. Jin, Y. Liu, T. Chen, D. Papadopoulos, Q. Yang, SecureBoost: A lossless federated learning framework, IEEE Intell. Syst. 36 (6) (2021) 87–98, http://dx.doi.org/10.1109/MIS.2021.3082561.

[11] S. Nakamoto, A. Bitcoin, A peer-to-peer electronic cash system, Bitcoin 4 (2) (2008) 15, URL: https://bitcoin.org/bitcoin.pdf.

[12] S. Tanwar, Blockchain Technology: From Theory to Practice, Springer Nature, 2022.

[13] A. Verma, P. Bhattacharya, N. Madhani, C. Trivedi, B. Bhushan, S. Tanwar, G. Sharma, P.N. Bokoro, R. Sharma, Blockchain for industry 5.0: Vision, opportunities, key enablers, and future directions, IEEE Access (2022) http://dx.doi.org/10.1109/ACCESS.2022.3186892.

[14] F. Stiehle, I. Weber, Blockchain for business process enactment: A taxonomy and systematic literature review, in: A. Marrella, R. Matulevičius, R. Gabryelczyk, B. Axmann, V. Bosilj Vukšić, W. Gaaloul, M. Indihar Štemberger, A. Kő, Q. Lu (Eds.), International Conference on Business Process Management, Springer, 2022, pp. 5–20, http://dx.doi.org/10.1007/978-3-031-16168-1_1.

[15] F. Casino, T.K. Dasaklis, C. Patsakis, A systematic literature review of blockchain-based applications: Current status, classification and open issues, Telemat. Inform. 36 (2019) 55–81.

[16] M.A. Uddin, A. Stranieri, I. Gondal, V. Balasubramanian, A survey on the adoption of blockchain in IoT: challenges and solutions, Blockchain: Res. Appl. 2 (2) (2021) 100006, http://dx.doi.org/10.1016/j.bcra.2021.100006.

[17] S. Alam, M. Shuaib, W.Z. Khan, S. Garg, G. Kaddoum, M.S. Hossain, Y.B. Zikria, Blockchain-based initiatives: Current state and challenges, Comput. Netw. (Amsterdam, Netherlands : 1999) 198 (2021) 108395,

[18] M.J.M. Chowdhury, M.S. Ferdous, K. Biswas, N. Chowdhury, V. Muthukkumarasamy, A survey on blockchain-based platforms for IoT use-cases, Knowl. Eng. Rev. 35 (2020).

[19] S.F. Wamba, M.M. Queiroz, Blockchain in the operations and supply chain management: Benefits, challenges and future research opportunities, Int. J. Inf. Manage. 52 (2020) 102064, http://dx.doi.org/10.1016/j.ijinfomgt.2019.102064.

[20] S.M.H. Bamakan, A.B. Bondarti, P.B. Bondarti, Q. Qu, Blockchain technology forecasting by patent analytics and text mining, Blockchain: Res. Appl. 2 (2) (2021) 100019.

[21] D. Calvaresi, Y. Mualla, A. Najjar, S. Galland, M. Schumacher, Explainable multi-agent systems through blockchain technology, in: International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems, Springer, 2019, pp. 41–58.

[22] N. Karandikar, R. Abhishek, N. Saurabh, Z. Zhao, A. Lercher, N. Marina, R. Prodan, C. Rong, A. Chakravorty, Blockchain-based prosumer incentivization for peak mitigation through temporal aggregation and contextual clustering, Blockchain: Res. Appl. 2 (2) (2021)

[23] L. Stockburger, G. Kokosioulis, A. Mukkamala, R.R. Mukkamala, M. Avital, Blockchain-enabled decentralized identity management: The case of self-sovereign identity in public transportation, Blockchain: Res. Appl. 2 (2) (2021) 100014, http://dx.doi.org/10.1016/j.bcra.2021.100014.

[24] A. Moinet, B. Darties, J.-L. Baril, Blockchain based trust & authentication for decentralized sensor networks, 2017, arXiv preprint arXiv:1706.01730.

[25] S.K. Lo, M. Staples, X. Xu, Modelling schemes for multi-party blockchain-based systems to support integrity analysis, Blockchain: Res. Appl. 2 (2) (2021) 100024.

[26] S.K. Radha, I. Taylor, J. Nabrzyski, I. Barclay, Verifiable badging system for scientific data reproducibility, Blockchain: Res. Appl. 2 (2) (2021) 100015.

[27] K. Salah, M.H.U. Rehman, N. Nizamuddin, A. Al-Fuqaha, Blockchain for AI: Review and open research challenges, IEEE Access 7 (2019) 10127–10149.

[28] S. Sachan, J.-B. Yang, D.-L. Xu, D.E. Benavides, Y. Li, An explainable AI decision-support-system to automate loan underwriting, Expert Syst. Appl. 144 (2020) 113100.

[29] A. El Azzaoui, S.K. Singh, Y. Pan, J.H. Park, Block5GIntell: Blockchain for AI-enabled 5G networks, IEEE Access 8 (2020) 145918–145935.

[30] M. Aledhari, R. Razzak, R.M. Parizi, F. Saeed, Federated learning: A survey on enabling technologies, protocols, and applications, IEEE Access 8 (2020) 140699–140725.

[31] Y. Qu, M.P. Uddin, C. Gan, Y. Xiang, L. Gao, J. Yearwood, Blockchain-enabled federated learning: A survey, ACM Comput. Surv. 55 (4) (2022) 1–35.

[32] Z. Wang, Q. Hu, Blockchain-based federated learning: A comprehensive survey, 2021, arXiv preprint arXiv:2110.02182 URL https://arxiv.org/abs/2110.02182.

[33] Y. Li, C. Chen, N. Liu, H. Huang, Z. Zheng, Q. Yan, A blockchain-based decentralized federated learning framework with committee consensus, IEEE Netw. 35 (1) (2020) 234–241.

[34] W. Issa, N. Moustafa, B. Turnbull, N. Sohrabi, Z. Tari, Blockchain-based federated learning for securing internet of things: A comprehensive survey, ACM Comput. Surv. 55 (9) (2023) 1–43.

[35] Y. Guo, C. Liang, Blockchain application and outlook in the banking industry, Financ. Innov. 2 (1) (2016) 1–12.

[36] A. Polyviou, P. Velanas, J. Soldatos, Blockchain technology: financial sector applications beyond cryptocurrencies, Multidiscip. Digit. Publ. Inst. Proc. 28 (1) (2019) 7.

[37] J. Zhang, T. Lyu, R. Li, A study on SMIE credit evaluation model based on blockchain technology, Proc. CIRP 83 (2019) 616–623, http://dx.doi.org/10.1016/j.procir.2019.05.003.

[38] Y. Qiao, Q. Lan, Z. Zhou, C. Ma, Privacy-preserving credit evaluation system based on blockchain, Expert Syst. Appl. 188 (2022) 115989.

[39] F. Yang, Y. Qiao, Y. Qi, J. Bo, X. Wang, BACS: blockchain and AutoML-based technology for efficient credit scoring classification, Ann. Oper. Res. (2022) 1–21, http://dx.doi.org/10.1007/s10479-022-04531-8.

[40] R. Walambe, A. Kolhatkar, M. Ojha, A. Kademani, M. Pandya, S. Kathote, K. Kotecha, Integration of explainable AI and blockchain for secure storage of human readable justifications for credit risk assessment, in: D. Garg, K. Wong, J. Sarangapani, S.K. Gupta (Eds.), International Advanced Computing Conference, Springer, Singapore, 2020, pp. 55–72, http://dx.doi.org/10.1007/978-981-16-0404-1_5.

[41] S. Chakraborty, S. Aich, S.J. Seong, H.-C. Kim, A blockchain based credit analysis framework for efficient financial systems, in: 2019 21st International Conference on Advanced Communication Technology, ICACT, IEEE, 2019, pp. 56–60.

[42] D. Mao, F. Wang, Z. Hao, H. Li, Credit evaluation system based on blockchain for multiple stakeholders in the food supply chain, Int. J. Environ. Res. Public Health 15 (8) (2018) 1627.

[43] S.B. Patel, P. Bhattacharya, S. Tanwar, N. Kumar, Kirti: A blockchain-based credit recommender system for financial institutions, IEEE Trans. Netw. Sci. Eng. 8 (2) (2020) 1044–1054.

[44] K.W. Cho, B.-G. Jeong, S.U. Shin, Verifiable credential proof generation and verification model for decentralized SSI-based credit scoring data, IEICE Trans. Inf. Syst. 104 (11) (2021) 1857–1868.

[45] X. Zhu, Blockchain-based identity authentication and intelligent credit reporting, J. Phys. Conf. Ser. 1437 (1) (2020) 012086, http://dx.doi.org/10.1088/1742-6596/1437/1/012086.

[46] F. Yang, Y. Qiao, M.Z. Abedin, C. Huang, Privacy-preserved credit data sharing integrating blockchain and federated learning for industrial 4.0, IEEE Trans. Ind. Inform. 18 (12) (2022) http://dx.doi.org/10.1109/TII.2022.3151917.

[47] M. Nassar, K. Salah, M.H. ur Rehman, D. Svetinovic, Blockchain for explainable and trustworthy artificial intelligence, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 10 (1) (2020) e1340.

[48] M.T. Ribeiro, S. Singh, C. Guestrin, "Why should i trust you?" Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.

[49] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).

[50] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115.

[51] G. Schwalbe, B. Finzel, A comprehensive taxonomy for explainable artificial intelligence: A systematic survey of surveys on methods and concepts, 2021, http://dx.doi.org/10.1007/s10618-022-00867-8, arXiv e-prints, arXiv–2105.

[52] M.R. Islam, M.U. Ahmed, S. Barua, S. Begum, A systematic review of explainable artificial intelligence in terms of different application domains and tasks, Appl. Sci. 12 (3) (2022) 1353.

[53] M. Sahakyan, Z. Aung, T. Rahwan, Explainable artificial intelligence for tabular data: A survey, IEEE Access 9 (2021) 135392–135422.

[54] G. Vilone, L. Longo, Explainable artificial intelligence: a systematic review, 2020, http://dx.doi.org/10.48550/arXiv:2006.00093, arXiv preprint arXiv:2006.00093.

[55] F. Lampathaki, C. Agostinho, Y. Glikman, M. Sesana, Moving from 'black box'to 'glass box'artificial intelligence in manufacturing with XMANAI, in: 2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), IEEE, 2021, pp. 1–6.

[56] I.A. Khan, N. Moustafa, D. Pi, K.M. Sallam, A.Y. Zomaya, B. Li, A new explainable deep learning framework for cyber threat discovery in industrial IoT networks, IEEE Internet Things J. (2021).

[57] H. Li, J. Wu, H. Xu, G. Li, M. Guizani, Explainable intelligence-driven defense mechanism against advanced persistent threats: A joint edge game and AI approach, IEEE Trans. Dependable Secure Comput. 19 (2) (2021) 757–775.

[58] A. Oseni, N. Moustafa, G. Creech, N. Sohrabi, A. Strelzoff, Z. Tari, I. Linkov, An explainable deep learning framework for resilient intrusion detection in IoT-enabled transportation networks, IEEE Trans. Intell. Transp. Syst. (2022).

[59] G. Rjoub, J. Bentahar, O.A. Wahab, Explainable AI-based federated deep reinforcement learning for trusted autonomous driving, in: 2022 International Wireless Communications and Mobile Computing, IWCMC, IEEE, 2022, pp. 318–323.

[60] I. Alarab, S. Prakoonwit, Effect of data resampling on feature importance in imbalanced blockchain data: Comparison studies of resampling techniques, Data Sci. Manage. (2022).

[61] G.K. Rajbahadur, S. Wang, G.A. Oliva, Y. Kamei, A.E. Hassan, The impact of feature importance methods on the interpretation of defect classifiers, IEEE Trans. Softw. Eng. 48 (7) (2022) 2245–2261.

[62] A. Wan, L. Dunlap, D. Ho, J. Yin, S. Lee, H. Jin, S. Petryk, S.A. Bargal, J.E. Gonzalez, NBDT: neural-backed decision trees, 2020, arXiv preprint arXiv:2004.00221 URL https://arxiv.org/abs/2004.00221.

[63] S. Hara, K. Hayashi, Making tree ensembles interpretable: A bayesian model selection approach, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2018, pp. 77–85.

[64] A. Altmann, L. Toloşi, O. Sander, T. Lengauer, Permutation importance: a corrected feature importance measure, Bioinformatics 26 (10) (2010) 1340–1347.

[65] H. Bride, J. Dong, J.S. Dong, Z. Hóu, Towards dependable and explainable machine learning using automated reasoning, in: International Conference on Formal Engineering Methods, Springer, 2018, pp. 412–416.

[66] H. Bride, C.-H. Cai, J. Dong, J.S. Dong, Z. Hóu, S. Mirjalili, J. Sun, Silas: A high-performance machine learning foundation for logical reasoning and verification, Expert Syst. Appl. 176 (2021) 114806.

[67] G. Zhang, Z. Hou, Y. Huang, J. Shi, H. Bride, J.S. Dong, Y. Gao, Extracting optimal explanations for ensemble trees via logical reasoning, 2021, arXiv preprint arXiv:2103.02191 URL https://arxiv.org/abs/2103.02191.

[68] P.E. de Lange, B. Melsom, C.B. Vennerø d, S. Westgaard, Explainable AI for credit assessment in banks, J. Risk Financ. Manage. 15 (12) (2022) 556.

[69] V. Moscato, A. Picariello, G. Sperlí, A benchmark of machine learning approaches for credit score prediction, Expert Syst. Appl. 165 (2021) 113986.

[70] R. Srinivasan, A. Chander, P. Pezeshkpour, Generating user-friendly explanations for loan denials using GANs, 2019, http://dx.doi.org/10.48550/arXiv.1906.10244, arXiv preprint arXiv:1906.10244.

[71] S. Ma, J. Shi, Y. Huang, S. Qin, Z. Hou, MUC-driven feature importance measurement and adversarial analysis for random forest, 2022, arXiv preprint arXiv:2202.12512.

[72] G. Fahner, Developing transparent credit risk scorecards more effectively: An explainable artificial intelligence approach, Data Anal. 2018 (2018) 17.

[73] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, Y. Gao, A survey on federated learning, Knowl.-Based Syst. 216 (2021) 106775, http://dx.doi.org/10.1016/j.knosys.2021.106775.

[74] J. Konečnỳ, H.B. McMahan, F.X. Yu, P. Richtárik, A.T. Suresh, D. Bacon, Federated learning: Strategies for improving communication efficiency, 2016, arXiv preprint arXiv:1610.05492.

[75] L. Wu, W. Ruan, J. Hu, Y. He, A survey on blockchain-based federated learning, Future Internet 15 (12) (2023) 400.

[76] F. Sattler, S. Wiedemann, K.-R. Müller, W. Samek, Robust and communication-efficient federated learning from non-iid data, IEEE Trans. Neural Netw. Learn. Syst. 31 (9) (2019) 3400–3413.

[77] N.Q. Hieu, T.T. Anh, N.C. Luong, D. Niyato, D.I. Kim, E. Elmroth, Resource management for blockchain-enabled federated learning: A deep reinforcement learning approach, 2020, arXiv preprint arXiv:2004.04104.

[78] X. Yin, Y. Zhu, J. Hu, A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions, ACM Comput. Surv. 54 (6) (2021) 1–36, http://dx.doi.org/10.1145/3460427.

[79] D. Li, D. Han, T.-H. Weng, Z. Zheng, H. Li, H. Liu, A. Castiglione, K.-C. Li, Blockchain for federated learning toward secure distributed machine learning systems: a systemic survey, Soft Comput. 26 (9) (2022) 4423–4440.

[80] D. Li, Z. Luo, B. Cao, Blockchain-based federated learning methodologies in smart environments, Cluster Comput. 25 (4) (2022) 2585–2599.

[81] I. Martinez, S. Francis, A.S. Hafid, Record and reward federated learning contributions with blockchain, in: 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), IEEE, 2019, pp. 50–57.

[82] D.C. Nguyen, M. Ding, Q.-V. Pham, P.N. Pathirana, L.B. Le, A. Seneviratne, J. Li, D. Niyato, H.V. Poor, Federated learning meets blockchain in edge computing: Opportunities and challenges, IEEE Internet Things J. 8 (16) (2021) 12806–12825.

[83] A. Grafberger, M. Chadha, A. Jindal, J. Gu, M. Gerndt, FedLess: Secure and scalable federated learning using serverless computing, in: 2021 IEEE International Conference on Big Data (Big Data), 2021, pp. 164–173, http://dx.doi.org/10.1109/BigData52589.2021.9672067.

[84] R.A. Mallah, D. Lopez, G.B. Marfo, B. Farooq, Untargeted poisoning attack detection in federated learning via behavior attestation, 2021, arXiv preprint arXiv:2101.10904.

[85] R. Al Mallah, D. López, Blockchain-based monitoring for poison attack detection in decentralized federated learning, in: 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME, IEEE, 2022, pp. 1–6.

[86] D. Malhotra, S. Srivastava, P. Saini, A.K. Singh, Blockchain based audit trailing of XAI decisions: Storing on IPFS and ethereum blockchain, in: 2021 International Conference on COMmunication Systems and NETworkS, COMSNETS, 2021, pp. 1–5, http://dx.doi.org/10.1109/COMSNETS51098.2021.9352908.

[87] M.S. Hossain, G. Muhammad, N. Guizani, Explainable AI and mass surveillance system-based healthcare framework to combat COVID-I9 like pandemics, IEEE Netw. 34 (4) (2020) 126–132.

[88] J.S. Bellagarda, A.M. Abu-Mahfouz, An updated survey on the convergence of distributed ledger technology and artificial intelligence: Current state, major challenges and future direction, IEEE Access 10 (2022) 50774–50793, http://dx.doi.org/10.1109/ACCESS.2022.3173297.

[89] B. Chen, H. Zeng, T. Xiang, S. Guo, T. Zhang, Y. Liu, ESB-FL: Efficient and secure blockchain-based federated learning with fair payment, IEEE Trans. Big Data (2022) 1, http://dx.doi.org/10.1109/TBDATA.2022.3177170.

[90] R. Davis, A.W. Lo, S. Mishra, A. Nourian, M. Singh, N. Wu, R. Zhang, Explainable machine learning models of consumer credit risk, 2022, Available at SSRN.

[91] L. Thomas, J. Crook, D. Edelman, Credit Scoring and its Applications, SIAM, 2017.

[92] E.I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, J. Financ. 23 (4) (1968) 589–609.

[93] R.E. Turkson, E.Y. Baagyere, G.E. Wenya, A machine learning approach for predicting bank credit worthiness, in: 2016 Third International Conference on Artificial Intelligence and Pattern Recognition, AIPR, IEEE, 2016, pp. 1–7, http://dx.doi.org/10.1109/ICAIPR.2016.7585216.

[94] D. West, Neural network credit scoring models, Comput. Oper. Res. 27 (11–12) (2000) 1131–1152, http://dx.doi.org/10.1016/S0305-0548(99)00149-5.

[95] M.B. Yobas, J.N. Crook, P. Ross, Credit scoring using neural and evolutionary techniques, IMA J. Manag. Math. 11 (2) (2000) 111–125, http://dx.doi.org/10.1093/imaman/11.2.111.

[96] T. Harris, Credit scoring using the clustered support vector machine, Expert Syst. Appl. 42 (2) (2015) 741–750.

[97] G. Nie, W. Rowe, L. Zhang, Y. Tian, Y. Shi, Credit card churn forecasting by logistic regression and decision tree, Expert Syst. Appl. 38 (12) (2011) 15273–15285.

[98] E. Dumitrescu, S. Hué, C. Hurlin, S. Tokpavi, Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects, European J. Oper. Res. 297 (3) (2022) 1178–1192.

[99] Y. Xia, C. Liu, Y. Li, N. Liu, A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring, Expert Syst. Appl. 78 (2017) 225–241, http://dx.doi.org/10.1016/j.eswa.2017.02.017.

[100] X. Dastile, T. Celik, M. Potsane, Statistical and machine learning models in credit scoring: A systematic literature survey, Appl. Soft Comput. 91 (2020) 106263.

[101] N. Siddiqi, Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards, John Wiley & Sons, 2017.

[102] N. Siddiqi, Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring, vol. 3, John Wiley & Sons, 2012.

[103] L.C. Thomas, A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers, Int. J. Forecast. 16 (2) (2000) 149–172, http://dx.doi.org/10.1016/S0169-2070(00)00034-0.

[104] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[105] W. Zhang, W. Xu, H. Hao, D. Zhu, Cost-sensitive multiple-instance learning method with dynamic transactional data for personal credit scoring, Expert Syst. Appl. 157 (2020) 113489.

[106] Z. Zhang, K. Niu, Y. Liu, A deep learning based online credit scoring model for P2P lending, IEEE Access 8 (2020) 177307–177317, http://dx.doi.org/10.1109/ACCESS.2020.3027337.

[107] Y. Qiao, Q. Lan, Y. Wang, S. Jia, X. Kuang, Z. Yang, C. Ma, PEvaChain: Privacy-preserving ridge regression-based credit evaluation system using hyperledger fabric blockchain, Expert Syst. Appl. (2023) 119844.

[108] Y. Qiao, Q. Lan, Y. Wang, S. Jia, X. Kuang, Z. Yang, C. Ma, PEvaChain: Privacy-preserving ridge regression-based credit evaluation system using hyperledger fabric blockchain, Expert Syst. Appl. 223 (2023) 119844.

[109] P. Thantharate, A. Thantharate, ZeroTrustBlock: Enhancing security, privacy, and interoperability of sensitive data through ZeroTrust permissioned blockchain, Big Data Cogn. Comput. 7 (4) (2023) 165.

[110] E. Nyaletey, R.M. Parizi, Q. Zhang, K.-K.R. Choo, BlockIPFS-blockchain-enabled interplanetary file system for forensic and trusted data traceability, in: 2019 IEEE International Conference on Blockchain (Blockchain), IEEE, 2019, pp. 18–25.

[111] S. Chen, Y. Wang, Convolutional Neural Network and Convex Optimization, Tech. Rep., Dept. of Elect. and Comput. Eng., Univ. of California at San Diego, San Diego, CA, USA, 2014.

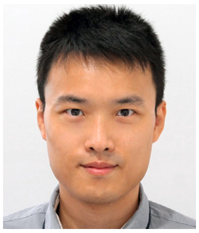
[112] S. Boyd, S.P. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[113] AWS, AWS lambda, 2022, URL https://aws.amazon.com/lambda/.

[114] B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in: Artificial Intelligence and Statistics, PMLR, 2017, pp. 1273–1282.

[115] X. Zeng, N. Hao, J. Zheng, X. Xu, A consortium blockchain paradigm on hyperledger-based peer-to-peer lending system, China Commun. 16 (8) (2019) 38–50, http://dx.doi.org/10.23919/JCC.2019.08.004.

Zorka Jovanovic received her B.Sc. in Mathematics and M.Sc. in Financial Mathematics. She is experienced quantitative analytics professional with deep credit risk modelling and software engineering knowledge. Her background includes extensive experience in developing advanced machine-learning models across credit risk. Zorka's current research interest is the integration of Blockchain, Artificial Intelligence, and Explainable AI, aiming to facilitate automated credit decision-making processes.



Dr Zhe Hou (common: "Zee Ho", correct-ish: "Hojer") obtained his Ph.D. from the Australian National University on the topic of automated reasoning for separation logic -- a logic for reasoning about pointers and other mutable data structures. In 2015, he joined Nanyang Technological University, Singapore, as a postdoc to work on formal verification of information flow security for instruction set architecture and weak memory models. He joined Griffith University, Australia, in 2017 on a project to develop trusted autonomous systems and advanced model checking techniques in collaboration with Australia Defense Science and Technology. He joined the faculty of Griffith University in late 2019. His research interests include logic, automated reasoning, formal methods, AI, sports analytics, blockchain, and quantum computing.

Dr Kamanashis Biswas received his M.Sc. in Computer Science with specialisation in Security Engineering from Blekinge Institute of Technology (BTH), Sweden and Ph.D. in Information and Communication Technology from Griffith University, Australia. Currently, he is working as a Senior Lecturer in Information Technology (Cybersecurity) at Australian Catholic University. He is also an adjunct lecturer at the School of Information and Communication Technology, Griffith University. Before starting his role at ACU, he worked as an Associate Lecturer at Griffith University for one and a half years. He also worked as a faculty at Daffodil International University, Bangladesh for about four and a half years. His research interests include blockchain technology, privacy issues in the metaverse, explainable AI, design and development of lightweight security protocols, energy efficient and secure routing, intrusion detection systems, and security issues in SDNs, MANETs, and WSNs. He has published numerous papers in the areas of blockchain, WSN, security, and privacy in reputed venues and high-impact journals including IEEE Internet of Things, Computer Networks, Journal of Network and Computer Applications, Future Generation Computer Systems, and Nature Scientific Report. He has served in different leading capacities, such as a keynote speaker, general chair, session chair, organising committee member and/or TPC member of several international conferences including TrustCom, GridCom, SDLT, CISIS, ICCS, ACM BlockSys, and PRDC-23. Dr Biswas has been a recipient of several internal and external research grants. He has also been editor of special issues in international journals and organised several workshops in the past.

Professor Vallipuram Muthukkumarasamy received a Ph.D. degree from Cambridge University, England, and a B.Sc.Eng. (Hons.) from the University of Peradeniya, Sri Lanka. He is currently attached to the School of ICT, Griffith University, Australia, as Professor. His research areas include cyber security, blockchain technology, and wireless sensor networks. He has pioneered the Network Security and Blockchain Research Group at the Institute for Integrated and Intelligent Systems.