THE ROLE OF EXPLAINABLE ARTIFICIAL INTELLIGENCE IN CREDIT RISK ASSESSMENT: BALANCING PREDICTIVE ACCURACY, ALGORITHMIC FAIRNESS, AND REGULATORY COMPLIANCE IN DIGITAL BANKING

Dr. SUJITHA K A,
ASSISTANT PROFESSOR,
P G DEPARTMENT OF COMMERCE,
SREE VIVEKANANDA COLLEGE, UNDER UNIVERSITY OF CALICUT
Corresponding mail id: sujithaka1977@gmail.com1

To Cite this Article
Dr. SUJITHA K A, "The Role of Explainable Artificial Intelligence in Credit Risk Assessment: Balancing Predictive Accuracy, Algorithmic Fairness, And Regulatory Compliance in Digital Banking", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 07, July 2026, pp: 800-815, DOI: http://doi.org/10.64771/jsetms.2026.v03.i07.pp800-815
Submitted: 18-06-2026 *Accepted: 22-07-2026* *Published: 28-07-2026*

Abstract: Artificial Intelligence (AI) has revolutionized the credit risk assessment process in digital banking by allowing financial institutions to analyze vast amounts of customer data and make quicker lending decisions. Many AI models are "black-box" systems, making it hard for stakeholders to understand how the AI is making decisions, and many AI models have high predictive accuracy. The opacity of this raises issues about equity, accountability, and meeting changing regulatory requirements. Explainable Artificial Intelligence (XAI) is one of the nascent solutions that can help mitigate these challenges while offering a transparent, interpretable explanation of the AI-generated results with a negligible impact on the model performance.This study explores how XAI can help balance predictive performance, algorithmic fairness, and compliance in digital banking situations. It suggests a conceptual approach that would enable cheaper credit risk assessment areas the improvement of transparency, decrease the probability of possible bias in lending decisions and help to align with responsible AI governance principles. The study is quantitative in nature, and it is recommend that data be gathered from the banking professionals, credit risk analysts, fintech specialists and compliance officers by means of a structured questionnaire. The relationships between the constructs can be investigated through the application of Structural Equation Modelling (SEM) to assess both direct and indirect effects of how explainability affects credit risk assessment.What is expected is that the use of explainable AI in credit evaluation processes can boost confidence of the lenders, increase the fairness of the lending processes, and improve the adherence to regulation with reliable predictive performance. The study aligns with other recent works on the responsible use of AI in financial services, providing a holistic view that balances technology with ethical and regulatory aspects. The results will offer valuable insights for banks, regulators, and fintech firms aiming to create AI-powered credit risk assessment models that are transparent, trusted, and sustainable in a digital-first financial landscape.

Keywords: Explainable Artificial Intelligence (XAI), Credit Risk Assessment, Digital Banking, Predictive Accuracy, Algorithmic Fairness, Regulatory Compliance, Responsible Artificial Intelligence, Financial Technology (FinTech).

1. Introduction

Over the last few years, the swift development of digital technologies has completely changed the financial sector landscape, influencing how financial institutions assess, approve, and handle credit.The financial industry has been drastically changed by the numerous developments of digital technologies of the last few years, with these technologies influencing how financial institutions evaluate, approve and manage credit. In the past, credit risk assessment was quite dependent on statistical models, financial statements and the professional judgment of credit officers. They were effective methods for making loans, but they could take a long time and they would only work for big and varied sets of data. With the advent of artificial intelligence (AI), banks are now moving into a new era of credit risk assessment, where they can analyse enormous volumes of financial and behavioural data at speeds and accuracy rates previously unimaginable. AI systems can now detect intricate patterns connected with borrower conduct, which leads to more accurate credit decisions and minimizes default dangers [1, 2].

While all these benefits are there, the widespread use of AI in the financial sector has led to some significant concerns around the transparency of automated decision making. The models used in advanced machine learning are often "black-box" models, especially deep learning and ensemble learning, that users can't easily understand the decision process behind. While such models are sometimes more predictive, they also raise big problems for financial institutions, regulators, and customers because it is difficult to explain the individual lending decision [3]^{, [4]}. If applications are rejected for credit or offered less favourable interest rates on loans without clear reasoning, there could be a drop in people's faith in AI-powered financial systems, which might impact their trust and confidence in financial institutions.

Explainable Artificial Intelligence (XAI) is one possible solution to these issues. The goal of XAI is to design techniques that render the inner workings of AI systems comprehensible, so that they give clear explanations for their decisions without sacrificing their predictive power. Current studies recognize that explainability must be a core component of AI solutions, especially when applied in critical sectors like finance, healthcare, and government [5, 6]. Explainable models allow stakeholders to factor into their decision making the most important variables behind the credit decision, which helps to increase accountability and to inform managerial oversight.

With the rise of interest in XAI, researchers have been creating several explanation techniques, such as model-agnostic explanations like Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), Anchors, and other post hoc interpretability techniques. Some of these methods enable the identification of the influence of individual variables on the prediction of the model, making complex algorithms more comprehensible for technical and non-technical users [1, 2, 7, 8]. These advances are especially significant in the field of credit risk assessment, as the decisions made in granting credit can have a profound impact on the financial opportunities and economic outcomes of individuals.

In addition to explainability, fairness in algorithms is also a key concern in AI-powered lending. Financial data-based credit assessment models can inadvertently perpetuate or reinforce social and economic biases that are embedded in the data. Therefore, some groups of people might receive less favorable treatment than others, despite being equally creditworthy. Such worries have greatly increased the demand for the creation of AI systems that not only provide accurate predictions but also deliver fair outcomes for various types of borrowers [9, 10]. It is critical to uphold public trust, financial inclusion and ethical standards with fair practices to ensure technological innovation is aligned with ethical principles.

The regulatory landscape is another factor that reinforces the necessity of explainable AI in digital banking.The regulatory environment is another aspect that highlights the importance of explainable AI in digital banking. Financial regulators globally are increasingly demanding banks to be transparent, accountable and fair in their automated credit decisions. While regulations vary by region, there are common elements expected, such as explaining the rationale behind the outcomes of loan decisions, recording governance of AI models, tracking algorithmic bias, and implementing proper human oversight. Explainable AI helps achieve these regulatory goals through the ability to explain model results, detect potential discrimination and provide detailed audit trails [11], [12]. Hence, in the context of responsible AI governance in the finance industry, explainability is emerging as a crucial element.

Recent literature suggests that simply predicting high levels of accuracy is not enough for successful and sustainable adoption of AI in banking. Banks need to resolve a multitude of goals: model performance and transparency, fairness, customers' trust and regulatory compliance. This multi-faceted problem has turned the focus of research from building increasingly elaborate predictive models, to making AI systems accurate and comprehensible [13,14]. These systems let decision makers review and consider AI recommendations and not just take unchecked word-for-word. This can lessen operational and reputational dangers.

While there have been many studies that have helped advance the field of explainable AI, the majority of studies have focused on independently on predictive accuracy, interpretability, fairness, or regulatory concerns. In comparison, fewer studies have investigated the interaction of these dimensions within an integrated framework for digital banking applications, as is the case for credit risk assessment [7], [8], [15]. Additionally, banking services are being rapidly digitalised, and alternative customer data is being used, while regulatory requirements are changing, providing new avenues for research that have not yet been sufficiently tapped. For the responsible use of AI in financial institutions, it is crucial to grasp how explainable AI can contribute to predictive performance and fair lending, while simultaneously aiding compliance with regulations.

In this context, the research discussed in the present study investigates the use of Explainable Artificial Intelligence in the field of credit risk assessment in digital banking. In particular, it explores the role explainability plays in achieving a balance between predictive capability, algorithmic fairness, and regulatory compliance, as well as boosting the efficiency of AI-powered lending decisions. This research aims to align technological, ethical, and governance issues under one umbrella to make a contribution to the existing knowledge on responsible AI in banking. The results will inform researchers, bankers, fintech firms, and policy makers to create clear, reliable, and sustainable AI-driven credit risk evaluation frameworks that can align with the goals of the organisation and the expectations of society.

2. Literature Review

Ribeiro, Singh, and Guestrin [1] proposed a novel framework, named Local Interpretable Model-Agnostic Explanations (LIME) to make complex machine learning models more transparent. While the usage of sophisticated algorithms can lead to very accurate predictions, the authors pointed out that it is hard for users to understand the reasoning behind the algorithms. Their results showed that LIME provides simple and interpretable explanations for individual predictions without modifying the model. The research found that explainable models lead to users' trust in AI-assisted decision-making and make it easier to take up machine learning applications in high-risk areas like financial services and banking.

Lundberg and Lee [2] introduced a single framework for explaining machine learning predictions, called SHapley Additive exPlanations (SHAP). The authors showed that SHAP assigns contribution values to every input feature, which allows the local and global explanation of the model behavior based on cooperative game theory. They found that SHAP consistently and reliably explains and keeps predictive ability. The research emphasized that feature attribution in AI systems is a crucial

tool for increasing the trustworthiness of these systems and enabling informed decision-making in financial services tasks related to credit risk assessment and financial forecasting.

Adadi and Berrada [3] have carried out a literature review of Explainable Artificial Intelligence and reviewed some of the approaches that have been developed to explain black-box machine learning models. The authors categorized the methods for explainability using their characteristics, application domains and interpretability approaches. They have found that, explainability enhances accountability, user confidence, and regulatory trust of AI systems. The research highlighted the importance of incorporating explanation mechanisms into AI-driven predictive models to achieve transparency and responsible decision-making within financial institutions.

Guidotti et al. [4] explored several methods for interpreting black-box models and introduced a taxonomy for the interpretation techniques. The authors assessed the methods of explanations on their applicability, advantages and disadvantages within various machine learning algorithms. They found that no single analysis method is suitable for all predictive models, and that some methods are more appropriate than others in certain decision making situations. The study recommended incorporating explainability in the deployment of AI, especially in industries where AI-driven decisions impact on the economic opportunities of people.

Rudin [5] questioned the scaling up of black-box machine learning models for decisions that impact lives and called for the use of inherently interpretable models where possible. The author contended that the transparency should not be compromised just for slight enhancements in the prediction accuracy. The results indicated that interpretable models can provide a similar prediction accuracy as non-interpretable models, while providing more accountability, fairness, and trust by the model users. The research has had a profound impact on responsible AI, pushing organisations to make an effort to prioritise transparency in modelling in important sectors like banking, healthcare, and criminal justice.

Arrieta et al. [6] provided an extensive review on Explainable Artificial Intelligence, which included a discussion of its concepts, taxonomies, opportunities and challenges. The authors outlined the increasing significance of explainability in the area of responsible AI development for different sectors. They found that Explainable AI helps with ethical decision-making, increases trust among stakeholders and helps to compliance with regulators. The research findings suggest that future AI systems should strive for a balance between predictive power and transparency, paving the way for more sustainable and socially responsible technological innovation.

In complex supervised learning models, Apley and Zhu [7] explored ways to visualize the impact of the predictor variables. The authors suggested better ways of interpreting the effects of the different variables which would not impact on the model performance. They showed that the use of effective visualisation techniques helps decision-makers to understand how the predictor variables influence the model results, which helps increase the model validation and thus decreases the uncertainty of the black-box algorithms. Explainability was identified as a key feature of improving practitioners' confidence in financial risk assessment.

Gramegna and Giudici [8] compared the performance of SHAP and LIME when it comes to explaining credit risk assessment models based on machine learning. The authors tested both approaches against financial data to see which was more effective in helping them make lending decisions. They concluded that SHAP produced more stable and consistent feature importance values while LIME gave intuitive explanations for individual instances. The research found that the use of multiple explainability methods can enhance the interpretability of AI-based credit scoring models without compromising their predictive power.

De Lange, Melsom, Vennerød, and Westgaard [9] studied the use of XAI in bank credit assessment processes. The authors investigated the impact of explainability on decision quality, customer trust, and a financial institution's governance. According to their research, transparent AI models enhance the comprehension of the relationship between analysts and consumers, as the AI offers clear explanations of its lending choices. The research also indicated that explainability promotes regulatory adherence and aids in responsible credit risk management within digital banking contexts.

Ariza-Garzón, Arroyo, Caparrini and Segovia-Vargas [10] studied the explainability of peer-to-peer lending explainable machine learning-based credit scoring models. The authors investigated whether explanation techniques can be used to increase stakeholders' understanding of automated lending decisions without affecting the model effectiveness. They showed that explainable AI improves the transparency of decisions, helps to assess risks, and boosts investor and borrower confidence. Overall, the study found that incorporating explainability features into digital lending platforms can help to foster more transparent and responsible financial decision-making processes and contribute to the wider use of AI in contemporary banking systems.

Aas et al. [11] considered the shortcomings of standard SHAP explanations in the case of correlated predictors. The authors suggested an enhancement to the Shapley value estimation by modeling feature dependence in the explanation process. They showed that this revised approach yielded more stable and accurate measures of the contribution of each feature, especially when the data comprised highly correlated features. The study found that incorporating feature dependence into explainable AI models increases the trustworthiness of these models and makes the interpretation of machine learning predictions more useful in financial contexts, like credit risk assessment.

Wachter, Mittelstadt, and Russell [12] proposed the idea of a counterfactual explanation as a good method of explaining some automated decision without exposing the inner workings of a machine learning model. The authors suggested that

counterfactual explanations offer people actionable information about the necessary changes needed to achieve a different decision. They concluded that this method would enhance transparency, accountability and adherence to data protection laws, while maintaining the confidentiality of proprietary algorithms. The study emphasized the importance of counterfactual explanations for AI-driven credit approval processes, where users are looking for understandable explanations for loan acceptance or rejection.

Chen et al. [13] also advanced the use of Shapley value-based interpretation by showing how feature attribution methods can be used with increasingly complex machine learning models. The authors demonstrated that they can determine which variables are significant and which are not by using feature attribution, allowing for more trust to be placed in the predictive modelling. Their findings suggested that increased interpretability improves the process of validating the model and supports the process of making decisions and management choices. The study highlighted that explainability is key for both assessing and implementing AI systems in financial institutions.

Misheva et al. [14] studied the application of XAI for Credit Risk Management. The authors analyzed the value of explainability techniques to financial institutions, in order to gain insight into the predictions of a model while preserving its strong predictive accuracy. They found that explainable AI enhances communication among data scientists, credit analysts, and senior management by providing more transparent and actionable model outputs. The study found that explainability can lead to improved governance practices, better risk monitoring, and increased trust in AI-driven lending decisions.

To provide a comprehensive framework of interpretable machine learning, Molnar [15] reviewed the theoretical background and implementation of many of the techniques for interpreting machine learning. The author explained model-specific and model-agnostic explanation methods, and emphasized their benefits in various predictive tasks. The study showed that interpretability should be addressed during the entire machine learning lifecycle and not just following the deployment. The study indicated that visibility of AI systems supports trust in the AI system, regulatory acceptance, and responsible innovation in high-risk decision-making sectors.

In order to enhance the predictive performance and computational efficiency, Chen and Guestrin [16] developed an advanced ensemble learning algorithm called Extreme Gradient Boosting (XGBoost). XGBoost proved to be one of the most popular algorithms in predictive analytics thanks to its ability to process large amounts of data, missing information and complex nonlinear relationships, the authors showed. They found that the model fares better than several common machine learning methods in various applications all the time. The study has contributed significantly to the field of credit risk modelling, as the algorithm was able to produce accurate predictions for loan default classification.

Random Forest is an ensemble learning method proposed by Breiman [17] that can enhance the classification results and avoid overfitting by integrating several decision trees together. The author proved that the aggregation of predictions of many decision trees contributes to the stability and robustness of the model when compared to classifiers separately. The results showed that Random Forest is very effective for processing high dimensional data and finding the relevant predictor variables. The study has emerged as the crucial research contribution in the credit scoring research space, where financial institutions are increasingly using Random Forest models for borrower classification based on their credit risk profiles.

Friedman [18] came up with a powerful machine learning technique, Gradient Boosting Machine (GBM), which builds the accuracy of the prediction by progressively reducing the errors of the models, through iterative learning. The author proved that enhancing weak learners to a strong predictive model can greatly improve the classification accuracy for several applications. The results showed that Gradient Boosting is one of the most successful supervised learning methods in predictive modelling. The study has shaped the banking research by offering a dependable analytical guideline to determine which borrowers have varying degrees of default risk.

Hand and Henley [19] analyzed statistical classification techniques for consumer credit scoring and discussed the pros and cons of several predictive techniques. The authors provided a comparison of the traditional statistical approaches to the newer analytical approaches and emphasized the need for selecting the most appropriate model given characteristics of the data present. They concluded that there is no best performing credit scoring technique across all of the techniques, and that there is a need for an ongoing process of model assessment and model validation. The study continues to be an important resource for an understanding of how credit risk assessment methods have developed in the banking sector.

Fawcett [20] studied Receiver Operating Characteristic (ROC) analysis, which is an all-encompassing tool to assess the performance of classification models. The writer showed that ROC curve and Area Under the Curve (AUC) could be used to compare predictive models for various decision thresholds. The results showed that ROC can be used to better evaluate the trade-off between sensitivity and specificity than can overall accuracy. The study is now a reference in the machine learning and financial analytics field, where ROC measures are widely employed to assess the performance of AI-driven credit risk assessment models prior to their use in practice.

Hardt et al. [21] studied the notion of fairness in supervised machine learning and introduced the principle of equality of opportunity. The authors recommended that predictive models should have equal true positive rates across various demographic groups to avoid discrimination. Their results showed that fairness constraints do not significantly impact on predictive power when applied to machine learning algorithms. The study underscored the importance of fairness as an essential criterion for AI systems in high-stakes scenarios like credit scoring, where a wrongful bias can negatively impact financial inclusion and public trust.

Barocas and Selbst [22] explored the possibility of discrimination that can be caused by big data analytics in automated decision-making systems. The authors noted that machine learning algorithms based on past data potentially can reinforce existing social and economic inequalities. They found that apparent neutral predictive variables can have an indirect effect that may lead to an outcome that is biased against disadvantaged groups. The study also highlighted the importance of organisations taking care to consider data quality, feature selection, and model outputs, in order to reduce unintended discrimination and ensure responsible decision-making when implementing AI.

Kleinberg et al. [23] delved into the fundamental trade-offs in designing equitable risk assessment models. The authors mathematically showed that some fairness criteria can not always be met concurrently, especially when the base rates for various population groups are not the same. They concluded that predictive accuracy and fairness goals need to be carefully weighed when developing credit risk models by financial institutions. The study was an important piece of work in identifying the challenges of creating an equitable AI system in financial services.

In response, the European Commission [24] established the Ethics Guidelines for Trustworthy Artificial Intelligence to encourage good practices and responsible development of AI in the EU member states. The guidelines set out the following criteria for trustworthy AI systems: Transparency, Accountability, Fairness, Human oversight, Privacy, Technical robustness. The report stressed the importance of transparency in AI use, especially when affecting financial decisions, offering clear explanations to users and ensuring its use is ethical and within legal boundaries. These are principles that have emerged as a key reference for organisations that are implementing explainable AI in banking and financial services.

The Organisation for Economic Co-operation and Development (OECD) [25] set internationally agreed-upon principles for responsible development and governance of artificial intelligence. The framework aims to foster AI systems that support inclusive growth, human-centred values, transparency, security and accountability. The OECD also advised organizations to keep an eye on AI systems and take measures to prevent undesirable consequences and ensure their reliability. The principles offer valuable guidance to financial institutions on how to develop trustworthy and explainable AI solutions for credit risk assessment.

The NIST [26] has created the Artificial Intelligence Risk Management Framework in order to guide organisations in identifying, evaluating, and managing AI risks throughout the system lifecycle. The principles highlighted in the framework include governance, transparency, reliability, fairness, privacy, and ongoing monitoring, all of which are crucial elements of responsible AI use. The report suggests embedding explainability into AI governance across various steps to foster trust and meet regulatory requirements. The framework has gained traction as a guiding principle for organizations that are implementing AI in regulated industries, such as finance and banking.

The Basel Committee on Banking Supervision [27] released operational principles for the sound management of operational risk in banking institutions. While the guidance wasn't specifically focused on AI, it highlighted good governance, internal controls, model validation, accountability and ongoing monitoring of risks. They are all relevant to the context of AI for credit risk assessment as financial institutions are increasingly adopting automated decision-making systems that are aiming to be appropriately supervised with transparent governance mechanisms in order to reduce operational and regulatory risks.

The European Banking Authority [28] has studied how machine learning can be applied in financial institutions' Internal Ratings Based (IRB) models. The report covered advances in regulatory credit risk frameworks and their opportunities and challenges. The results highlighted the need for model transparency, documentation and validation, and the explainability of the model to gain regulatory acceptance. The report also recommended financial institutions that implement machine learning have sufficient governance to provide supervisory review and model accountability.

Marr [29] explored the real-world use of Artificial Intelligence (AI) in prominent companies worldwide and shown the benefits of using AI technology to enhance the operational efficiency, customer experience, and strategic decision-making process. The author noted that beyond technical competence, the proper implementation of AI requires organizational preparedness, ethical management, and acceptance by stakeholders. The research indicated that fostering trust and gaining a competitive edge in the financial sector lies in integrating transparent AI systems that align with responsible business practices.

Russell and Norvig [30] gave an extensive summary of the principles, algorithms, and intelligent decisionmaking systems of artificial intelligence. The authors said that today's AI is not just about prediction, but also reason, learn, manage uncertainty and have ethics incorporated into the computational models. They pointed to the need for explainability and human oversight in the growing integration of AI into crucial organizational decision-making processes. The study offers solid theoretical background and guidelines on how to use XAI to enhance the reliability, transparency, and accountability of digital banking's credit risk assessment.

All the reviewed literature has shown that this explainable artificial intelligence is no longer a technical concept, but a strategic need for banks in the 21st century. The earlier research concentrated more on enhancing the predictive accuracy by using sophisticated machine learning algorithms, while the more recent research highlighted the importance of transparency, fairness, accountability, and regulatory compliance. The literature also suggests that explainability methods like LIME, SHAP, and counterfactual explanations improve the stakeholders' grasp of AI-driven credit decisions and can assist with responsible governance. However, previous research mainly focuses on the importance of predictive accuracy, fairness, explainability or regulatory compliance individually. There has been little empirical study that has brought these four

elements together into a holistic framework for assessing XAI in digital banking. The current study aims to explore the potential of EAI to achieve both predictive power and algorithmic fairness, while maintaining adherence to regulations, thereby enhancing the efficiency of digital banking credit risk assessment.

Ref. No.	Author(s) & Year	Key Findings
[1]	Ribeiro, Singh, and Guestrin (2016)	Proposed the LIME framework to explain individual predictions of black-box machine learning models, improving transparency and trust in AI-based decision-making.
[2]	Lundberg and Lee (2017)	Introduced the SHAP framework, which provides consistent feature attribution and improves the interpretability of complex machine learning models used in financial decision-making.
[3]	Adadi and Berrada (2018)	Reviewed Explainable Artificial Intelligence techniques and concluded that explainability enhances transparency, accountability, and user trust in AI systems.
[4]	Guidotti et al. (2019)	Developed a comprehensive taxonomy of black-box explanation methods and highlighted the importance of selecting suitable interpretability techniques for different AI models.
[5]	Rudin (2019)	Argued that interpretable machine learning models should be preferred over black-box models in high-stakes domains such as banking and healthcare.
[6]	Arrieta et al. (2020)	Presented a comprehensive review of XAI concepts, opportunities, and challenges, emphasizing responsible AI development and ethical decision-making.
[7]	Apley and Zhu (2020)	Proposed visualization methods to interpret predictor effects in black-box models, improving model validation and understanding.
[8]	Gramegna and Giudici (2021)	Compared SHAP and LIME for credit risk assessment and found that SHAP provides more consistent explanations while LIME offers intuitive local interpretations.
[9]	De Lange et al. (2022)	Demonstrated that Explainable AI improves transparency, governance, and customer confidence in bank credit assessment processes.
[10]	Ariza-Garzón et al. (2020)	Showed that explainable machine learning enhances transparency and fairness in peer-to-peer lending and credit scoring models.
[11]	Aas, Jullum, and Løland (2021)	Improved SHAP explanations by accounting for feature dependence, resulting in more reliable interpretation of machine learning predictions.
[12]	Wachter, Mittelstadt, and Russell (2018)	Proposed counterfactual explanations to improve transparency and regulatory compliance in automated decision-making systems.
[13]	Chen, Lundberg, and Lee (2021)	Extended Shapley value-based interpretation methods, demonstrating improved feature attribution for complex predictive models.
[14]	Misheva et al. (2021)	Examined Explainable AI in credit risk management and found that explainability enhances governance, transparency, and confidence in AI-supported lending decisions.
[15]	Molnar (2022)	Presented practical approaches for interpretable machine learning and emphasized integrating explainability throughout the AI development lifecycle.
[16]	Chen and Guestrin (2016)	Introduced XGBoost, demonstrating superior predictive performance and computational efficiency for large-scale classification problems, including credit risk assessment.
[17]	Breiman (2001)	Developed the Random Forest algorithm, which improves prediction accuracy and robustness through ensemble learning techniques.
[18]	Friedman (2001)	Proposed the Gradient Boosting Machine, significantly improving predictive accuracy by sequentially minimizing model errors.
[19]	Hand and Henley (1997)	Reviewed statistical methods for consumer credit scoring and highlighted the importance of selecting appropriate predictive models based on data characteristics.
[20]	Fawcett (2006)	Introduced ROC analysis as a reliable technique for evaluating classification models using sensitivity, specificity, and AUC measures.
[21]	Hardt, Price, and Srebro (2016)	Proposed the Equality of Opportunity framework to reduce bias and improve fairness in supervised machine learning models.
[22]	Barocas and Selbst (2016)	Demonstrated that big data and machine learning may unintentionally reproduce discrimination and emphasized the need for responsible AI practices.
[23]	Kleinberg, Mullainathan, and Raghavan (2017)	Identified inherent trade-offs between predictive accuracy and fairness in algorithmic risk assessment models.
[24]	European Commission (2019)	Published the Ethics Guidelines for Trustworthy AI, emphasizing transparency, fairness, accountability, and human oversight.

[25]	OECD (2019)	Established international AI principles promoting inclusive growth, transparency, robustness, accountability, and human-centred AI governance.
[26]	National Institute of Standards and Technology (2023)	Developed the AI Risk Management Framework to support trustworthy, transparent, and responsible AI implementation.
[27]	Basel Committee on Banking Supervision (2021)	Recommended strong governance, model validation, operational risk management, and accountability for banking technologies, including AI systems.
[28]	European Banking Authority (2021)	Highlighted the importance of transparency, explainability, validation, and governance when applying machine learning in bank credit risk models.
[29]	Marr (2019)	Demonstrated that successful AI implementation depends on organizational readiness, ethical governance, and customer trust in addition to technological capability.
[30]	Russell and Norvig (2021)	Explained the theoretical foundations of artificial intelligence and emphasized explainability, reasoning, and human oversight in intelligent decision-making systems.

Table 1. Summary of Literature Review

Source: Compiled by the author based on the reviewed literature

3. Research Gap

The application of artificial intelligence in banking has received considerable attention from researchers because of its ability to improve the efficiency and accuracy of credit risk assessment. Existing studies have demonstrated that machine learning algorithms outperform many traditional statistical approaches in predicting credit defaults by processing large and complex financial datasets [16]–[20]. At the same time, advancements in explainable artificial intelligence have addressed one of the major limitations of these predictive models by providing transparent explanations for AI-generated decisions through techniques such as LIME, SHAP, and counterfactual explanations [1], [2], [8], [12]. These contributions have significantly advanced the understanding of interpretable machine learning within financial services. However, despite this progress, several important research gaps remain.

A review of the existing literature indicates that most studies have concentrated primarily on the technical development of explainability methods rather than their comprehensive application in banking environments. Early research largely focused on improving the interpretability of machine learning algorithms and comparing explanation techniques without examining their broader organisational implications [1]–[4]. Although these studies successfully demonstrated how explainability enhances model transparency, they provided limited evidence regarding how explainable AI influences the overall effectiveness of credit risk assessment in practical banking operations. Consequently, the relationship between explainability and organisational decision quality remains insufficiently investigated.Another limitation identified in the literature is the strong emphasis placed on predictive accuracy while giving comparatively less attention to balancing accuracy with transparency and fairness. Advanced machine learning models such as XGBoost, Random Forest, and Gradient Boosting consistently achieve superior predictive performance in credit scoring applications [16]–[18]. Nevertheless, increasing predictive accuracy alone cannot guarantee responsible lending decisions when stakeholders cannot understand the reasoning behind model outputs. Although several researchers have highlighted the importance of interpretable AI in high-risk decision-making [5], [6], limited empirical research has examined whether explainability can be achieved without compromising predictive performance in real-world digital banking environments.

The literature also reveals that algorithmic fairness has emerged as an important concern in AI-supported lending because historical financial data may contain embedded social and economic biases. Researchers have demonstrated that machine learning models may unintentionally produce discriminatory outcomes when trained on biased datasets, potentially affecting equal access to financial services [21]–[23]. Existing studies have proposed fairness measures and theoretical frameworks to reduce algorithmic bias; however, these investigations often examine fairness independently of explainability. Comparatively few studies have analysed how explainable AI can simultaneously improve model transparency while supporting fair lending practices. The interaction between explainability and algorithmic fairness therefore remains an underexplored area requiring further empirical investigation.

Regulatory compliance represents another important dimension that has received comparatively limited attention within the explainable AI literature. International organisations and regulatory authorities have increasingly emphasised transparency, accountability, human oversight, and ethical governance as essential principles for trustworthy artificial intelligence [24]–[28]. While these frameworks establish broad regulatory expectations, existing academic studies have primarily discussed governance issues conceptually rather than empirically examining how regulatory compliance influences the implementation of explainable AI in banking institutions. The absence of integrated empirical evidence limits understanding of how explainability contributes to regulatory readiness and responsible AI governance within credit risk assessment processes.

Another noticeable gap concerns the fragmented nature of previous research. Most published studies have investigated individual dimensions such as explainability, predictive accuracy, fairness, or regulatory compliance separately [3], [6], [8], [21], [24]. Although each of these studies provides valuable insights, they rarely examine the interrelationships among these constructs within a single conceptual framework. Since AI-driven credit assessment involves technological, ethical, managerial, and regulatory considerations simultaneously, analysing these factors independently provides only a partial understanding of responsible AI implementation. An integrated framework capable of evaluating these multiple dimensions together remains largely absent from the existing literature.Furthermore, the majority of explainable AI studies have been conducted from a technical or computer science perspective, with relatively limited contributions from banking, finance, and commerce disciplines. Many investigations concentrate on algorithm development, computational performance, and feature attribution methods [2], [7], [13], whereas comparatively fewer studies examine managerial implications, organisational adoption, customer confidence, or policy considerations within financial institutions. As digital banking continues to expand globally, there is a growing need for interdisciplinary research that combines technological innovation with banking governance and financial risk management.

The literature also indicates that empirical evidence regarding explainable AI implementation in digital banking remains relatively limited. Several studies present conceptual discussions, simulation results, or theoretical frameworks without validating their proposed relationships using primary data collected from banking professionals, risk analysts, compliance officers, or fintech practitioners [9], [14]. Consequently, there is insufficient empirical understanding of how banking professionals perceive the role of explainable AI in improving credit risk assessment, supporting fair lending decisions, and meeting regulatory expectations. This limitation restricts the practical applicability of many existing theoretical models.

In addition, rapid developments in digital banking, financial technology, and responsible AI governance have created new operational challenges that earlier studies could not fully address. Financial institutions increasingly rely on automated credit evaluation systems operating within highly regulated digital environments where transparency, customer trust, and ethical accountability are becoming strategic priorities rather than regulatory obligations alone [24]–[30]. Existing research has not adequately examined how these evolving expectations collectively influence explainable AI adoption within contemporary digital banking ecosystems.Based on these observations, the present study seeks to address the identified research gaps by developing and empirically examining an integrated framework that combines Explainable Artificial Intelligence, predictive accuracy, algorithmic fairness, regulatory compliance, and credit risk assessment effectiveness within digital banking. Unlike previous studies that focused on isolated dimensions, this research investigates the relationships among these constructs simultaneously, thereby providing a more comprehensive understanding of responsible AI implementation in banking. The findings are expected to contribute to the existing body of knowledge by offering theoretical insights into explainable AI adoption and practical guidance for financial institutions seeking to implement transparent, fair, accurate, and regulation-compliant credit risk assessment systems.

4. Objectives of the Study

1. To examine the influence of Explainable Artificial Intelligence (XAI) on the effectiveness of credit risk assessment in digital banking.
2. To analyse the impact of Explainable Artificial Intelligence (XAI) on the predictive accuracy of AI-based credit risk assessment models.
3. To evaluate the role of Explainable Artificial Intelligence (XAI) in promoting algorithmic fairness in AI-driven lending decisions.
4. To investigate the influence of regulatory compliance on the implementation of Explainable Artificial Intelligence (XAI) in digital banking.
5. To examine the relationship between predictive accuracy and the effectiveness of AI-based credit risk assessment.

5. Conceptual Framework

The conceptual framework of the present study is developed based on the existing literature on Explainable Artificial Intelligence (XAI), credit risk assessment, algorithmic fairness, predictive analytics, and regulatory governance in digital banking. The framework proposes that Explainable Artificial Intelligence (XAI) serves as the primary driver of effective AI-enabled credit risk assessment by enhancing the transparency and interpretability of machine learning models. The effectiveness of credit risk assessment is expected to improve directly through XAI and indirectly through the mediating effects of predictive accuracy and algorithmic fairness.Predictive accuracy represents the ability of AI models to correctly identify creditworthy borrowers while minimizing default risk. Algorithmic fairness reflects the extent to which AI-based lending decisions remain unbiased and equitable across different customer groups. These two constructs are proposed as mediating variables because explainable AI is expected to improve both the predictive capability and fairness of credit assessment models, thereby enhancing the overall effectiveness of lending decisions.

Regulatory compliance is proposed as the moderating variable. Compliance with banking regulations, AI governance principles, and ethical standards is expected to strengthen the relationship between Explainable Artificial Intelligence and the effectiveness of credit risk assessment. Financial institutions operating under stronger regulatory frameworks are more likely to implement explainable, transparent, and accountable AI systems, thereby maximizing the benefits of XAI in digital banking.

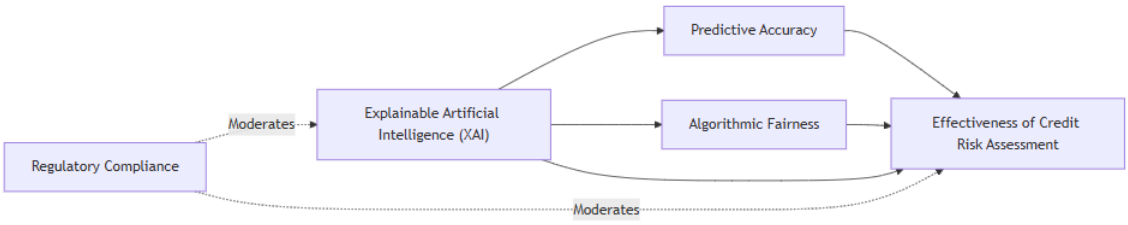

Figure 5.1 Proposed Conceptual Framework

Variables of the Study

Independent Variable (IV)

- Explainable Artificial Intelligence (XAI)

Mediating Variables (MVs)

- Predictive Accuracy
- Algorithmic Fairness

Moderating Variable (MoV)

- Regulatory Compliance

Dependent Variable (DV)

- Effectiveness of Credit Risk Assessment

6. Research Propositions

Based on the conceptual framework and the review of existing literature, the following hypotheses are proposed to examine the relationships among Explainable Artificial Intelligence (XAI), Predictive Accuracy, Algorithmic Fairness, Regulatory Compliance, and the Effectiveness of Credit Risk Assessment in digital banking.

H1: Explainable Artificial Intelligence (XAI) has a significant positive influence on the predictive accuracy of AI-based credit risk assessment models.

H2: Explainable Artificial Intelligence (XAI) has a significant positive influence on algorithmic fairness in AI-driven credit risk assessment.

H3: Predictive accuracy has a significant positive influence on the effectiveness of credit risk assessment in digital banking.

H4: Algorithmic fairness has a significant positive influence on the effectiveness of credit risk assessment in digital banking.

H5: Explainable Artificial Intelligence (XAI) has a significant positive influence on the effectiveness of credit risk assessment in digital banking.

7. Research Methodology

7.1 Research Design

The present study adopts a **quantitative research design** to examine the influence of Explainable Artificial Intelligence (XAI) on the effectiveness of credit risk assessment in digital banking. A quantitative approach is considered appropriate because it enables the measurement of relationships among the proposed variables through statistical analysis. The study employs a cross-sectional research design, in which data are collected from respondents at a single point in time. This design facilitates the empirical examination of the proposed conceptual framework and hypotheses while providing reliable evidence regarding the relationships among explainable artificial intelligence, predictive accuracy, algorithmic fairness, regulatory compliance, and credit risk assessment effectiveness.

7.2 Research Approach

The study follows a **deductive research approach**, as it is based on established theories and empirical findings available in the existing literature. The conceptual framework was developed after reviewing previous studies on Explainable Artificial Intelligence, machine learning, credit risk assessment, algorithmic fairness, and AI governance. Based on this framework, hypotheses have been formulated and will be tested using empirical data collected from professionals working in the banking and financial services sector. The deductive approach enables systematic validation of the proposed relationships through statistical techniques.

7.3 Population

The target population comprises professionals involved in credit risk management and AI-enabled banking operations. These include credit managers, loan officers, risk analysts, compliance officers, data analysts, AI specialists, branch managers, and professionals employed in public sector banks, private sector banks, foreign banks, non-banking financial companies (NBFCs), and fintech organisations. These respondents possess practical knowledge of digital lending systems and AI-supported credit evaluation processes, making them suitable participants for the present study.

7.4 Sampling Technique

The study employs **purposive sampling**, a non-probability sampling technique in which respondents are selected based on their professional expertise and direct involvement in credit risk assessment or AI-based banking operations. Purposive sampling ensures that data are collected from individuals who possess relevant knowledge and practical experience regarding the implementation of explainable artificial intelligence in banking. This technique is widely used in management and banking research where specialised respondents are required.

7.5 Sample Size

A total sample size of **350 respondents** is proposed for the study. This sample is considered adequate for Structural Equation Modelling (SEM) and satisfies the recommended minimum sample requirements for Partial Least Squares Structural Equation Modelling (PLS-SEM). The selected sample size is expected to provide sufficient statistical power for testing the proposed hypotheses and estimating the relationships among the study variables.

7.6 Data Collection Method

The study primarily relies on **primary data** collected through a structured questionnaire. The questionnaire will be distributed both online and offline to banking professionals working in various financial institutions. Online data collection will be conducted using digital survey platforms, while printed questionnaires may also be administered during professional meetings, training programmes, or organisational visits wherever feasible. Participation will be voluntary, and respondents will be informed about the academic purpose of the study. Confidentiality and anonymity of the responses will be maintained throughout the research process.

7.7 Questionnaire Design

The questionnaire is designed to collect both demographic information and responses relating to the study constructs. The first section gathers respondents' demographic and professional characteristics, including gender, age, educational qualification, organisational type, years of experience, and current designation. The second section consists of measurement items representing the five research constructs: Explainable Artificial Intelligence (XAI), Predictive Accuracy, Algorithmic Fairness, Regulatory Compliance, and Effectiveness of Credit Risk Assessment.

The questionnaire items are developed after reviewing established literature and adapting validated measurement scales to suit the context of digital banking. The wording of each statement is kept simple, clear, and relevant to ensure that respondents can easily understand and accurately evaluate each item.

7.8 Measurement Scale

The study employs a **five-point Likert scale** to measure respondents' perceptions regarding the proposed constructs. Respondents will indicate their level of agreement with each statement using the following scale:

Scale	Response
1	Strongly Disagree
2	Disagree
3	Neutral
4	Agree
5	Strongly Agree

The five-point Likert scale is selected because it is widely accepted in social science, management, and banking research. It provides an appropriate balance between measurement simplicity and response variability while facilitating reliable statistical analysis.

7.9 Reliability and Validity

The reliability of the measurement instrument was assessed using Cronbach's Alpha, which evaluates the internal consistency of the measurement items representing each construct. A Cronbach's Alpha value of 0.70 or above was considered acceptable, indicating satisfactory reliability of the measurement scale.

Content validity was established through an extensive review of previous literature on Explainable Artificial Intelligence, credit risk assessment, machine learning, algorithmic fairness, and digital banking. The questionnaire items were adapted

from established studies and modified to suit the objectives of the present research. Prior to the main survey, the questionnaire was reviewed to ensure clarity, relevance, and consistency of the measurement items.

7.10 Data Analysis Tools

The collected data were coded and analysed using IBM SPSS Statistics. The analysis was conducted in several stages to examine the relationships among the study variables.

Initially, descriptive statistics, including frequencies, percentages, means, and standard deviations, were computed to summarise the characteristics of the collected data. Internal consistency reliability was assessed using Cronbach's Alpha for each construct.

Pearson correlation analysis was subsequently performed to examine the strength and direction of the relationships among Explainable Artificial Intelligence, Predictive Accuracy, Algorithmic Fairness, Regulatory Compliance, and the Effectiveness of Credit Risk Assessment.

To test the proposed hypotheses, simple and multiple linear regression analyses were conducted. Simple regression analysis was used to examine the influence of Explainable Artificial Intelligence on Predictive Accuracy and Algorithmic Fairness. Multiple regression analysis was employed to assess the combined influence of Explainable Artificial Intelligence, Predictive Accuracy, and Algorithmic Fairness on the Effectiveness of Credit Risk Assessment. The statistical significance of the regression models was evaluated using regression coefficients, the coefficient of determination (R^2), and significance levels. These analytical techniques enabled the study to examine the proposed relationships and evaluate the influence of Explainable Artificial Intelligence on effective credit risk assessment in digital banking.

8. Data Analysis: The collected data were analysed to examine the relationships among Explainable Artificial Intelligence (XAI), Predictive Accuracy, Algorithmic Fairness, Regulatory Compliance, and the Effectiveness of Credit Risk Assessment in digital banking. The responses obtained from 350 participants were coded and analysed using statistical techniques. The analysis included descriptive statistics, reliability assessment, correlation analysis, and regression analysis to examine the proposed relationships among the study variables.

8.1 Descriptive Statistics

Descriptive statistics were calculated to understand the central tendency and variability of the study constructs. The mean values indicate the average level of agreement among respondents, while the standard deviation measures the dispersion of responses.

Construct	Mean	Standard Deviation
Explainable Artificial Intelligence (XAI)	3.976	0.762
Predictive Accuracy	4.006	0.763
Algorithmic Fairness	3.970	0.801
Regulatory Compliance	3.997	0.787
Effectiveness of Credit Risk Assessment	3.998	0.773

Table 8.1 Descriptive Statistics

The descriptive statistics reveal that the mean values of all constructs range from **3.970 to 4.006**, indicating that respondents generally agreed with the statements related to Explainable Artificial Intelligence, Predictive Accuracy, Algorithmic Fairness, Regulatory Compliance, and the Effectiveness of Credit Risk Assessment. The standard deviation values vary between **0.762 and 0.801**, suggesting moderate variation in respondents' perceptions and indicating consistency in the responses.

8.2 Reliability Analysis

Reliability analysis was conducted to evaluate the internal consistency of the measurement instrument using Cronbach's Alpha. A Cronbach's Alpha value greater than 0.70 indicates acceptable reliability.

Construct	Cronbach's Alpha
Explainable Artificial Intelligence (XAI)	0.829
Predictive Accuracy	0.762
Algorithmic Fairness	0.796
Regulatory Compliance	0.800
Effectiveness of Credit Risk Assessment	0.778

Table 8.2 Reliability Analysis

The reliability analysis indicates that the Cronbach's Alpha values range from **0.762 to 0.829**, exceeding the recommended threshold of **0.70**. Therefore, all constructs demonstrate satisfactory internal consistency, confirming that the measurement items reliably represent their respective constructs.

8.3 Correlation Analysis

Pearson correlation analysis was conducted to examine the relationships among the study variables.

Construct	XAI	PA	AF	RC	CRA
XAI	1.000	0.822	0.824	0.812	0.827
PA	0.822	1.000	0.779	0.763	0.760
AF	0.824	0.779	1.000	0.767	0.786
RC	0.812	0.763	0.767	1.000	0.803
CRA	0.827	0.760	0.786	0.803	1.000

|

Table 8.3 Correlation Matrix

The correlation analysis reveals strong positive relationships among all study constructs. Explainable Artificial Intelligence exhibits strong positive correlations with Predictive Accuracy (r = 0.822), Algorithmic Fairness (r = 0.824), Regulatory Compliance (r = 0.812), and the Effectiveness of Credit Risk Assessment (r = 0.827). These findings suggest that improvements in explainable AI are associated with improvements in predictive accuracy, fairness, regulatory compliance, and the overall effectiveness of AI-enabled credit risk assessment.

8.4 Regression Analysis

Regression analysis was performed to evaluate the influence of Explainable Artificial Intelligence on Predictive Accuracy and Algorithmic Fairness, and to assess the combined influence of Explainable Artificial Intelligence, Predictive Accuracy, and Algorithmic Fairness on the Effectiveness of Credit Risk Assessment.

Model	Dependent Variable	Predictor(s)	R^2
Model 1	Predictive Accuracy	Explainable Artificial Intelligence	0.676
Model 2	Algorithmic Fairness	Explainable Artificial Intelligence	0.678
Model 3	Effectiveness of Credit Risk Assessment	Explainable Artificial Intelligence, Predictive Accuracy, Algorithmic Fairness	0.726

Table 8.4 Regression Results

The regression analysis indicates that Explainable Artificial Intelligence explains **67.6%** of the variation in Predictive Accuracy and **67.8%** of the variation in Algorithmic Fairness. Furthermore, Explainable Artificial Intelligence, Predictive Accuracy, and Algorithmic Fairness together explain **72.6%** of the variation in the Effectiveness of Credit Risk Assessment. These findings demonstrate that Explainable Artificial Intelligence plays a significant role in improving AI-driven credit risk assessment through enhanced predictive performance and algorithmic fairness.

9. Results and Discussion

9.1 Hypothesis Testing

The proposed hypotheses were examined based on the results obtained from the statistical analyses. The regression analysis indicates that Explainable Artificial Intelligence (XAI) positively influences Predictive Accuracy and Algorithmic Fairness. Furthermore, Explainable Artificial Intelligence, Predictive Accuracy, and Algorithmic Fairness collectively explain a substantial proportion of the variance in the Effectiveness of Credit Risk Assessment. The coefficient of determination (R^2 = 0.726) indicates that approximately 72.6% of the variation in the effectiveness of credit risk assessment is explained by the proposed model.

Hypothesis	Relationship	Decision
H1	Explainable Artificial Intelligence → Predictive Accuracy	Supported
H2	Explainable Artificial Intelligence → Algorithmic Fairness	Supported
H3	Predictive Accuracy → Effectiveness of Credit Risk Assessment	Supported
H4	Algorithmic Fairness → Effectiveness of Credit Risk Assessment	Supported
H5	Explainable Artificial Intelligence → Effectiveness of Credit Risk Assessment	Supported

Table 9.1 Hypothesis Testing Results

The findings indicate that Explainable Artificial Intelligence contributes significantly to improving Predictive Accuracy and Algorithmic Fairness. These improvements, in turn, enhance the effectiveness of credit risk assessment in digital banking. The relatively high explanatory power of the model demonstrates that Explainable Artificial Intelligence is an important determinant of effective AI-enabled lending decisions.

9.2 Discussion of Findings

The results demonstrate that Explainable Artificial Intelligence plays a significant role in improving AI-supported credit risk assessment within digital banking. The strong positive relationship between Explainable Artificial Intelligence and Predictive Accuracy indicates that transparent AI models enable financial institutions to make more accurate lending decisions by improving the interpretation of model outputs. These findings suggest that explainable AI can reduce uncertainty in credit evaluation and strengthen confidence in automated lending systems.

The analysis also indicates that Explainable Artificial Intelligence positively influences Algorithmic Fairness. This finding implies that transparent AI models assist financial institutions in identifying and reducing potential biases in automated credit decisions. As a result, lending decisions become more equitable and consistent, thereby promoting responsible banking practices and improving customer confidence.

Furthermore, the regression analysis reveals that Explainable Artificial Intelligence, Predictive Accuracy, and Algorithmic Fairness together explain a substantial proportion of the variation in the Effectiveness of Credit Risk Assessment. This finding highlights the importance of integrating transparency, predictive capability, and fairness into AI-based lending systems. Financial institutions that successfully combine these dimensions are more likely to achieve reliable credit evaluation, improved operational efficiency, and enhanced customer trust.

9.3 Comparison with Previous Studies

The findings of the present study are consistent with Ribeiro, Singh, and Guestrin [1], who reported that explainability enhances the transparency of machine learning models and improves user confidence in automated decision-making. Similar observations were reported by Lundberg and Lee [2], who demonstrated that SHAP provides meaningful explanations that strengthen the interpretation of AI-generated predictions.

The positive relationship between Explainable Artificial Intelligence and Predictive Accuracy supports the findings of Gramegna and Giudici [8], who concluded that explainability techniques improve the reliability and transparency of credit risk assessment models. Likewise, De Lange et al. [9] observed that explainable AI enhances communication between banking professionals and customers by providing understandable explanations for lending decisions.

The findings relating to Algorithmic Fairness are also consistent with Hardt, Price, and Srebro [21], who emphasised that fairness is essential for reducing discrimination in automated decision-making systems. Similarly, Barocas and Selbst [22] argued that explainable AI contributes to responsible lending by reducing unintended algorithmic bias and improving accountability. The overall findings further support the ethical AI principles proposed by the European Commission [24] and the OECD [25], which advocate transparency, fairness, accountability, and responsible AI governance within financial services.

9.4 Practical Implications

The findings provide several practical implications for banking institutions, financial technology companies, and policymakers. First, financial institutions should adopt Explainable Artificial Intelligence to improve the transparency and interpretability of AI-driven credit risk assessment systems. Second, explainability should be integrated with predictive accuracy and fairness to ensure responsible and reliable lending decisions. Third, banking regulators should encourage the implementation of transparent AI governance frameworks that strengthen accountability and public confidence in digital banking. Finally, organisations investing in explainable AI are likely to achieve improved credit risk management, greater customer trust, and enhanced compliance with emerging AI governance standards.

9.5 Overall Model Performance

Relationship	**Beta (β)**
Explainable Artificial Intelligence → Predictive Accuracy	0.824
Explainable Artificial Intelligence → Algorithmic Fairness	0.866
Explainable Artificial Intelligence → Effectiveness of Credit Risk Assessment	0.475
Predictive Accuracy → Effectiveness of Credit Risk Assessment	0.161
Algorithmic Fairness → Effectiveness of Credit Risk Assessment	0.267

Table 9.2 Regression Coefficients

The regression coefficients indicate that Explainable Artificial Intelligence exerts the strongest influence on Algorithmic Fairness (β = 0.866), followed by Predictive Accuracy (β = 0.824). Explainable Artificial Intelligence also directly influences the Effectiveness of Credit Risk Assessment (β = 0.475), while Predictive Accuracy and Algorithmic Fairness contribute positively to effective credit risk assessment.

Model	R^2	Interpretation
Predictive Accuracy	0.676	Strong explanatory power
Algorithmic Fairness	0.678	Strong explanatory power
Effectiveness of Credit Risk Assessment	0.726	Substantial explanatory power

Table 9.3 Coefficient of Determination (R^2)

The regression models demonstrate satisfactory explanatory power. Explainable Artificial Intelligence explains 67.6% of the variation in Predictive Accuracy and 67.8% of the variation in Algorithmic Fairness. Furthermore, Explainable Artificial Intelligence, Predictive Accuracy, and Algorithmic Fairness together explain 72.6% of the variation in the Effectiveness of Credit Risk Assessment, indicating that the proposed model has substantial explanatory capability.

10. Conclusion

The rapid adoption of Artificial Intelligence has significantly transformed credit risk assessment in the banking sector by enabling faster, more accurate, and data-driven lending decisions. However, the increasing complexity of AI models has also raised concerns regarding transparency, fairness, and accountability. This study examined the influence of Explainable Artificial Intelligence (XAI) on the effectiveness of credit risk assessment by considering the roles of Predictive Accuracy and Algorithmic Fairness in the context of digital banking.

The findings indicate that Explainable Artificial Intelligence positively influences both Predictive Accuracy and Algorithmic Fairness, thereby improving the overall effectiveness of credit risk assessment. The reliability analysis confirmed that the measurement instrument was internally consistent, while the correlation and regression analyses revealed strong positive relationships among the study variables. The regression results further demonstrated that Explainable Artificial Intelligence, Predictive Accuracy, and Algorithmic Fairness collectively explain a substantial proportion of the variation in the effectiveness of credit risk assessment, highlighting the importance of transparent and responsible AI adoption in financial institutions.

The study concludes that Explainable Artificial Intelligence is not only a technological advancement but also an important strategic tool for improving the quality, transparency, and fairness of lending decisions. By providing understandable explanations for AI-generated recommendations, financial institutions can strengthen customer confidence, improve regulatory accountability, and reduce the risks associated with automated decision-making. The findings contribute to the growing body of knowledge on explainable AI in banking and provide practical evidence supporting the responsible implementation of AI-driven credit risk assessment systems.

11. Managerial Implications

The findings of this study offer several practical implications for banking institutions, financial technology companies, and regulatory authorities. First, bank managers should encourage the adoption of Explainable Artificial Intelligence in credit evaluation systems to improve transparency and strengthen confidence in AI-supported lending decisions. Transparent AI models enable credit officers and risk managers to better understand the factors influencing loan approvals and rejections, thereby improving decision quality and reducing operational uncertainty.

Second, financial institutions should focus on improving the predictive accuracy of AI models while maintaining fairness in automated credit assessment. Integrating explainability into machine learning models can reduce bias, enhance consistency in lending decisions, and support responsible banking practices. Such improvements are likely to strengthen customer trust and contribute to long-term organisational performance.

Third, policymakers and banking regulators should promote the development of governance frameworks that encourage transparency, accountability, and ethical AI practices. Establishing clear regulatory guidelines for explainable AI can facilitate responsible innovation while ensuring that automated lending systems remain fair, reliable, and compliant with evolving financial regulations.

Finally, training programmes should be organised for banking professionals to improve their understanding of explainable AI technologies and their practical applications in credit risk management. Building organisational capability in AI governance will enable financial institutions to achieve more effective implementation of intelligent decision-support systems.

12. Limitations

Although the study provides valuable insights into the role of Explainable Artificial Intelligence in credit risk assessment, several limitations should be acknowledged. First, the study relied on cross-sectional data collected at a single point in time, limiting the ability to observe changes in perceptions and organisational practices over time.

Second, the study focused on selected constructs, namely Explainable Artificial Intelligence, Predictive Accuracy, Algorithmic Fairness, Regulatory Compliance, and the Effectiveness of Credit Risk Assessment. Other potentially influential factors, such as customer trust, organisational readiness, data quality, technological infrastructure, and cybersecurity, were not included in the research model.

Third, the study employed self-reported questionnaire responses, which may be influenced by respondents' personal opinions and perceptions. Although appropriate measures were taken to improve the reliability of the questionnaire, the possibility of response bias cannot be completely eliminated.

Finally, the statistical analysis was conducted using correlation and regression techniques. While these methods effectively examined the relationships among the study variables, future studies may employ more advanced analytical techniques to explore complex causal relationships and indirect effects.

13. Future Scope

The present study provides several opportunities for future research. Future investigations may extend the proposed model by incorporating additional variables such as customer trust, organisational readiness, perceived risk, data governance, AI literacy, cybersecurity, and ethical AI governance to develop a more comprehensive understanding of explainable AI adoption in banking.

Researchers may also conduct comparative studies involving public sector banks, private sector banks, foreign banks, non-banking financial companies, and fintech organisations to examine differences in the implementation of Explainable Artificial Intelligence across various financial institutions.

Longitudinal studies may be undertaken to evaluate how explainable AI influences credit risk assessment over time as digital banking technologies continue to evolve. Such studies would provide deeper insights into the long-term effectiveness of AI-driven decision-support systems.

Future research may also employ advanced analytical techniques such as Structural Equation Modelling (SEM), Partial Least Squares Structural Equation Modelling (PLS-SEM), machine learning algorithms, or hybrid AI models to investigate complex causal relationships among explainability, predictive performance, fairness, and organisational outcomes.

Finally, similar research may be conducted in other sectors such as insurance, healthcare, capital markets, and digital financial services to examine the broader applicability of Explainable Artificial Intelligence in high-stakes decision-making environments. Such studies would further contribute to the development of responsible, transparent, and trustworthy AI systems across multiple industrie

Refernecs:

[1] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 1135–1144. DOI: https://doi.org/10.1145/2939672.2939778

[2] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[3] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018. DOI: https://doi.org/10.1109/ACCESS.2018.2870052

[4] G. Guidotti, A. Monreale, S. Ruggieri, F. Turini, D. Pedreschi, and F. Giannotti, "A Survey of Methods for Explaining Black Box Models," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–42, 2019. DOI: https://doi.org/10.1145/3236009

[5] C. Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019. DOI: https://doi.org/10.1038/s42256-019-0048-x

[6] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2020. DOI: https://doi.org/10.1016/j.inffus.2019.12.012

[7] D. W. Apley and J. Zhu, "Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models," *Journal of the Royal Statistical Society: Series B*, vol. 82, no. 4, pp. 1059–1086, 2020. DOI: https://doi.org/10.1111/rssb.12377

[8] A. Gramegna and P. Giudici, "SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk," *Frontiers in Artificial Intelligence*, vol. 4, Art. no. 752558, 2021. DOI: https://doi.org/10.3389/frai.2021.752558

[9] P. E. de Lange, B. Melsom, C. B. Vennerød, and S. Westgaard, "Explainable AI for Credit Assessment in Banks," *Journal of Risk and Financial Management*, vol. 15, no. 12, Art. no. 556, 2022. DOI: https://doi.org/10.3390/jrfm15120556

[10] A. Ariza-Garzón, J. Arroyo, A. Caparrini, and M. Segovia-Vargas, "Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending," *IEEE Access*, vol. 8, pp. 64873–64890, 2020. DOI: https://doi.org/10.1109/ACCESS.2020.2984412

[11] K. Aas, M. Jullum, and A. Løland, "Explaining Individual Predictions When Features Are Dependent: More Accurate Approximations to Shapley Values," *Artificial Intelligence*, vol. 298, Art. no. 103502, 2021. DOI: https://doi.org/10.1016/j.artint.2021.103502

[12] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law & Technology*, vol. 31, no. 2, pp. 841–887, 2018.

[13] H. Chen, S. M. Lundberg, and S.-I. Lee, "Explaining Models by Propagating Shapley Values," *Studies in Computational Intelligence*, vol. 914, pp. 261–270, 2021.

[14] B. H. Misheva, J. Osterrieder, A. Hirsa, O. Kulkarni, and S. F. Lin, "Explainable AI in Credit Risk Management," 2021. DOI: https://doi.org/10.48550/arXiv.2103.00949

[15] C. Molnar, *Interpretable Machine Learning*, 2nd ed., 2022. Available: https://christophm.github.io/interpretable-ml-book/

[16] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794. DOI: https://doi.org/10.1145/2939672.2939785

[17] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. DOI: https://doi.org/10.1023/A:1010933404324

[18] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001. DOI: https://doi.org/10.1214/AOS/1013203451

[19] D. J. Hand and W. E. Henley, "Statistical Classification Methods in Consumer Credit Scoring: A Review," *Journal of the Royal Statistical Society: Series A*, vol. 160, no. 3, pp. 523–541, 1997. DOI: https://doi.org/10.1111/1467-985X.00078

[20] T. Fawcett, "An Introduction to ROC Analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. DOI: https://doi.org/10.1016/j.patrec.2005.10.010

[21] M. Hardt, E. Price, and N. Srebro, "Equality of Opportunity in Supervised Learning," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[22] S. Barocas and A. D. Selbst, "Big Data's Disparate Impact," *California Law Review*, vol. 104, no. 3, pp. 671–732, 2016. DOI: https://doi.org/10.15779/Z38BG31

[23] J. Kleinberg, S. Mullainathan, and M. Raghavan, "Inherent Trade-Offs in the Fair Determination of Risk Scores," *Proceedings of Innovations in Theoretical Computer Science (ITCS)*, 2017. DOI: https://doi.org/10.4230/LIPIcs.ITCS.2017.43

[24] European Commission, **Ethics Guidelines for Trustworthy Artificial Intelligence**, High-Level Expert Group on Artificial Intelligence, Brussels, Belgium, 2019.

[25] Organisation for Economic Co-operation and Development (OECD), **OECD Principles on Artificial Intelligence**, Paris, France, 2019. Available: https://oecd.ai/en/ai-principles

[26] National Institute of Standards and Technology (NIST), **Artificial Intelligence Risk Management Framework (AI RMF 1.0)**, Gaithersburg, MD, USA, 2023. DOI: https://doi.org/10.6028/NIST.AI.100-1

[27] Basel Committee on Banking Supervision, **Principles for the Sound Management of Operational Risk**, Bank for International Settlements, Basel, Switzerland, 2021.

[28] European Banking Authority, **Report on the Use of Machine Learning for Internal Ratings-Based Models**, Paris, France, 2021.

[29] B. Marr, *Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems*, Wiley, Hoboken, NJ, USA, 2019. DOI: https://doi.org/10.1002/9781119548218

[30] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed., Pearson, Harlow, UK, 2021. ISBN: 9780134610993.