© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723458

AI-Enabled Advanced Decision Intelligence Framework for Optimizing Credit Risk Management in Saudi Enterprises

PUTHAN VEETTIL ABDULLA MOHD KAYOOM
ORCiD: 0009-0002-2636-5043

Abstract- Artificial intelligence is reshaping credit risk management by enabling lenders to process high-dimensional financial, transactional and behavioral information at a speed and granularity that conventional scorecards cannot easily match. Yet stronger prediction alone does not constitute decision intelligence. Credit decisions are economically consequential, regulated, path-dependent and exposed to model risk, data drift, unfairness, cybersecurity threats and weak organizational accountability. This review develops an AI-enabled advanced decision intelligence framework for optimizing credit risk management in Saudi enterprises. Using a structured integrative review of peer-reviewed literature, the paper synthesizes evidence on machine-learning credit scoring, explainable artificial intelligence, alternative data, class imbalance, profit-sensitive evaluation, model governance and Saudi digital-finance conditions. The synthesis indicates that ensemble and nonlinear models often improve discriminatory power, but their value depends on calibrated probabilities, stable explanations, portfolio-level economics, rigorous validation and human oversight. The proposed framework therefore links enterprise data, predictive analytics, explainability, risk appetite, workflow orchestration and continuous monitoring rather than treating credit scoring as an isolated classification task. Particular attention is given to Saudi requirements for trustworthy data use, Shariah-sensitive product structures, financial-sector digitalization and Vision 2030 objectives. The review concludes that Saudi enterprises should adopt a governed decision-intelligence architecture in which AI augments rather than replaces accountable credit judgment, and where performance is evaluated through accuracy, fairness, interpretability, robustness and economic outcomes simultaneously.

Keywords: artificial intelligence; credit risk; decision intelligence; explainable AI; Saudi Arabia; enterprise risk management; fintech; model governance

I. INTRODUCTION

Credit risk management converts uncertainty about a borrower's future capacity and willingness to repay into decisions on approval, exposure, pricing, collateral, covenants, monitoring and recovery. For decades, logistic regression and expert scorecards dominated this activity because they are comparatively transparent and operationally stable. Recent machine-learning research shows, however, that ensemble methods, kernel models and deep architectures can capture nonlinear interactions and heterogeneous borrower patterns that standard scorecards may miss [1–7]. This shift is important for Saudi enterprises because corporate lending increasingly intersects with digital payments, supply-chain finance, fintech channels, SME growth and data-intensive treasury systems. The managerial question is therefore no longer whether AI can generate a risk score, but how AI-generated evidence can be converted into defensible, timely and economically coherent credit decisions.

The literature also reveals a persistent tension between predictive sophistication and institutional trust. Black-box models may improve ranking performance while making it harder for credit officers, customers, auditors and regulators to understand why a decision occurred. Explainable machine-learning studies demonstrate that Shapley-based methods, feature attributions and hybrid models can improve transparency without necessarily abandoning advanced algorithms [2,8–10]. Yet explanations themselves can be unstable when defaults are rare or when data distributions shift [11]. A second tension concerns evaluation: accuracy and area under the curve are useful statistical measures, but a lender ultimately manages asymmetric losses, capital consumption, concentration, customer value and profitability. Profit-sensitive evaluation and counterfactual stress testing therefore extend the decision problem beyond classification [12–14].

Saudi Arabia provides a strategically significant setting for this debate. The Kingdom's digital-finance ecosystem has expanded alongside national programs that encourage fintech innovation, electronic payments and data-driven services. Peer-reviewed Saudi and Gulf research describes rapid banking digitalization, AI adoption and regulatory experimentation, while also identifying governance, cybersecurity, skills and stability concerns [28–30]. The attached Springer review used a structured review logic to connect Saudi AI research with national strategy, combining transparent search procedures with thematic synthesis; this paper adopts that structural discipline without reproducing its wording or analysis. The present review narrows the focus to enterprise credit risk and asks how predictive AI can become a governed decision-intelligence capability rather than an isolated model.

Decision intelligence is used here to mean an organizational system that joins data, models, rules, risk appetite, human judgment and feedback so that analytical predictions are translated into actions whose consequences can be monitored. This definition matters because a highly accurate model can still produce poor credit outcomes if thresholds ignore portfolio economics, if data lineage is weak, if explanations are not actionable, or if overrides are undocumented. Conversely, a somewhat simpler model may create greater enterprise value when it is well calibrated, explainable, stress-tested and embedded in disciplined workflows. The central proposition of this review is therefore that optimization requires a multi-objective architecture balancing discrimination, calibration, fairness, interpretability, robustness, compliance and financial performance.

II. AIM OF THE STUDY

The aim of this review is to synthesize recent evidence on AI-enabled credit risk assessment and translate it into an advanced decision intelligence framework suitable for Saudi enterprises. The study specifically seeks to bridge three bodies of knowledge that are often treated separately: predictive credit analytics, trustworthy and explainable AI, and enterprise governance within the Saudi digital-finance environment. Rather than identifying a single best algorithm, the review evaluates how different analytical and governance components can be combined to improve the quality, consistency and accountability of credit decisions across origination, monitoring and remediation.

III. RESEARCH OBJECTIVES

The review pursues five objectives. First, it compares the capabilities and limitations of contemporary statistical, machine-learning and deep-learning approaches for credit risk assessment. Second, it evaluates the roles of explainability, fairness, uncertainty, class-imbalance treatment and stress testing in making AI-based credit judgments trustworthy. Third, it examines how alternative data and enterprise information can enrich risk signals while creating new data-governance risks. Fourth, it identifies design principles for integrating predictive outputs with risk appetite, pricing, limits, human oversight and continuous model monitoring. Fifth, it contextualizes these principles for Saudi enterprises by considering digital-finance development, governance expectations, Shariah-sensitive decision contexts and Vision 2030-oriented economic diversification.

IV. RESEARCH QUESTIONS

The review is guided by four research questions. RQ1: Which AI and machine-learning approaches provide the most relevant capabilities for enterprise credit risk assessment, and what trade-offs accompany them? RQ2: How can explainability, fairness, calibration, robustness and economic value be jointly incorporated into credit decision design? RQ3: What organizational and data-governance mechanisms are required to convert predictive models into continuously monitored decision intelligence? RQ4: How should an advanced credit decision intelligence framework be adapted to the institutional and strategic conditions of Saudi enterprises?

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723458

V. REVIEW METHODOLOGY

A structured integrative review was selected because the research problem spans operations research, artificial intelligence, banking, fintech, risk management and Saudi digital transformation. The review followed the transparent logic common to systematic and integrative reviews: question definition, database search, eligibility assessment, quality appraisal, thematic coding and synthesis. Scopus, Web of Science, ScienceDirect, SpringerLink and publisher databases were targeted for peer-reviewed journal literature, with Crossref and publisher pages used to verify bibliographic details and DOI records. Searches emphasized 2020–2025 so that the synthesis reflects the current generation of explainable and governed credit-AI research, while allowing limited earlier conceptual material only where necessary. The final reference set used in this manuscript contains exactly thirty peer-reviewed sources, all with verified DOI links.

Search combinations linked credit-risk terms with analytical and governance concepts. Core strings included combinations of "credit risk", "credit scoring", "default prediction", "machine learning", "artificial intelligence", "explainable AI", "SHAP", "fairness", "class imbalance", "alternative data", "stress testing", "model risk", "fintech", "Saudi Arabia", "GCC", "banking" and "digital finance". Inclusion required a direct contribution to credit prediction, model explainability, credit-decision evaluation, trustworthy AI in finance, or the Saudi/GCC financial-technology context. Purely descriptive commercial reports, non-peer-reviewed commentary, conference-only material, papers without verifiable bibliographic identities and studies unrelated to lending decisions were excluded. No article-screening counts are reported because the review was designed as a structured evidence synthesis rather than a PRISMA-style exhaustive meta-analysis, and inventing retrieval numbers would undermine reproducibility.

Quality assessment considered five dimensions: source credibility, methodological transparency, relevance to credit decisions, evidential contribution and transferability to enterprise practice. Greater interpretive weight was assigned to studies using real lending or banking datasets, comparative model evaluation, explainability analysis, fairness testing, stress scenarios or robust out-of-sample designs. Review articles were used to map recurring findings but were not treated as substitutes for empirical evidence. Each retained study was coded against data inputs, model family, decision objective, evaluation metric, explainability mechanism, governance issue and practical implication. This produced thematic categories around predictive performance, data enrichment, explainability and fairness, economic optimization, lifecycle governance and Saudi contextualization.

The synthesis is qualitative and configurational rather than statistical. Results from heterogeneous datasets cannot be pooled responsibly because borrower populations, default definitions, sampling windows, product types and performance metrics differ substantially. Instead, convergent findings were identified across studies and tensions were retained where evidence was mixed. This approach also prevents the common error of declaring a universally superior algorithm from benchmark datasets. Methodological limitations remain: publication bias may favor positive AI results; public datasets may not resemble Saudi corporate portfolios; explainability metrics lack universal standards; and the fast evolution of generative and agentic AI means some emerging techniques remain underrepresented in peer-reviewed credit-risk evidence.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723458

Table 1. Evidence-to-design synthesis for AI-enabled credit risk decision intelligence

Evidence theme	Representative literature	Decision-intelligence implication	Primary control
Nonlinear/ensemble performance	[1], [5–7], [22–27]	Use champion-challenger model portfolios; avoid single-model dependence	Out-of-sample validation and calibration
Alternative and network data	[4], [21]	Expand thin-file information while quantifying uncertainty	Lineage, consent, missingness and proxy review
Explainability and selective learning	[2], [3], [10], [15], [20]	Provide layered explanations for officers, validators and customers	Fidelity, stability and reason-code testing
Imbalance, fairness and AI risk	[11], [16–18]	Evaluate minority defaults and disparate performance explicitly	Threshold testing, fairness diagnostics, AI-risk indicators
Stress and economic value	[8], [9], [12], [14]	Optimize decisions for loss, profit, capital and resilience	Scenario analysis and profit-sensitive evaluation
Saudi/GCC institutional context	[28–30]	Localize governance, skills, cybersecurity and policy alignment	Saudi-specific validation and accountable oversight

The table translates recurring findings into architecture requirements. It is not a ranking of algorithms; rather, it emphasizes that evidence about performance, explainability and context should be converted into explicit enterprise controls.

VI. THEMATIC LITERATURE REVIEW AND CRITICAL ANALYSIS

The first theme is the transition from scorecard prediction to multi-model risk intelligence. Systematic evidence shows that ensemble classifiers frequently outperform individual statistical models, while newer methods exploit nonlinear interactions and richer feature spaces [1,13]. Comparative studies of random forests, gradient boosting, support vector machines and neural architectures similarly report performance gains, although the magnitude varies by dataset and preprocessing choices [5–7]. Deep learning does not automatically dominate; large-scale benchmarking indicates that conventional ensembles remain highly competitive, especially when tabular credit data are limited [6]. Hybrid approaches such as penalized logistic tree regression are strategically important because they incorporate nonlinear effects while preserving a more transparent decision structure [7]. For Saudi enterprises, this suggests that model portfolios and champion-challenger structures are preferable to algorithmic monocultures.

A second theme concerns data quality and alternative information. Traditional corporate credit analysis relies on financial statements, repayment history, bureau information and sector exposures. AI expands this base to transactions, cash-flow traces, supply-chain relationships, digital interactions and other alternative indicators. Djeundje et al. show that alternative data can improve access for applicants with thin credit histories, but they also demonstrate sensitivity to sample construction and data context [4]. Network-based fintech models further illustrate how macroeconomic and relational information can support inclusive credit ratings [23]. The opportunity for Saudi SMEs is significant because richer data can reduce information asymmetry; the risk is equally significant because poor lineage, proxy discrimination, missingness and inconsistent definitions can create confident but unreliable scores.

The third theme is explainability. Bussmann et al. demonstrate that Shapley-based explanations can group borrowers according to similar underlying risk drivers, helping transform opaque predictions into economically interpretable profiles [2]. Sachan et al. design an explainable underwriting decision-support

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723458

system that combines data learning with explicit reasoning [3]. Later studies extend this logic through Shapley values, local explanations, interpretable selective learning and comprehensive XAI frameworks [10,15,20]. Explainability is not merely a presentation layer. It supports adverse-decision reasoning, model validation, feature scrutiny, credit-officer challenge and customer communication. However, it can also create false confidence if an explanation method is unstable or if users mistake local feature attribution for causal evidence.

Class imbalance illustrates that risk. Defaults are usually rarer than non-defaults, so a model can appear accurate while failing to identify the minority events that matter most. Khatir and Bee compare balancing techniques and machine-learning combinations, while Chen et al. show that imbalance can destabilize SHAP and LIME explanations as well as predictions [11,18]. This has direct governance implications: resampling, threshold selection and cost-sensitive learning should be documented as part of model design rather than treated as technical housekeeping. Credit teams should monitor precision, recall, calibration and expected loss at decision-relevant thresholds instead of relying on accuracy alone. Where low-default portfolios are material, uncertainty intervals and expert review become especially important.

A fourth theme is the integration of predictive performance with business value. Credit models are deployed to allocate scarce risk capacity, not to win benchmark competitions. Bueff et al. show how counterfactual stress scenarios can test interpretability and robustness under deterioration [9]. Mohammadnejad-Daryani et al. propose profit-oriented model evaluation that makes model comparison more intelligible to managers [21]. Bastos and Matos demonstrate that explainable models can support loss-given-default analysis without necessarily sacrificing predictive power [8]. These findings imply that Saudi lenders should evaluate models through expected loss, risk-adjusted return, approval quality, capital usage, collection outcomes and concentration effects. A model with marginally lower AUC may be preferable if it is better calibrated, more stable under stress and economically superior at the chosen decision threshold.

The fifth theme is model risk and trustworthy AI. Giudici et al. operationalize AI risk through sustainability, accuracy, fairness and explainability indicators, providing a useful bridge between technical performance and enterprise controls [17]. Fairness research shows that mitigation choices can affect both protected-group outcomes and profitability, meaning ethical and economic objectives should be analyzed jointly rather than sequentially [8,16]. In regulated credit settings, fairness should be tested for disparate performance, proxy variables and unexplained outcome gaps, while recognizing that legal and demographic categories differ across jurisdictions. The appropriate governance response is not to remove human accountability but to create traceable evidence showing what data were used, how the model behaved, who approved overrides and what outcomes followed. A sixth theme is continuous monitoring. Credit relationships evolve with macroeconomic conditions, payment behavior, sector shocks and borrower strategy. A model validated at origination can degrade as feature distributions, default rates or operating processes change. Survival-based credit models and stress-oriented explainability highlight the value of time-to-event analysis and scenario sensitivity [14,19]. Decision intelligence therefore requires drift detection, calibration monitoring, segment-level performance, override analysis and post-decision learning. Monitoring should also detect data-pipeline failures, because a model can remain mathematically unchanged while its inputs silently shift. These controls are particularly important when alternative data, APIs or automated document extraction are integrated into enterprise lending.

The Saudi context adds a seventh theme: institutional fit. Al-Baity's review of AI in Saudi digital finance highlights opportunities across fraud, credit, automation and personalization while emphasizing governance and capability requirements [28]. Eskandarany's study of banking boards identifies enthusiasm for AI and machine learning but also concerns around infrastructure, ethics, cybersecurity and governance [29]. GCC evidence on fintech adoption further shows that digital innovation can interact with bank stability and regulatory environments in complex ways [30]. These findings caution against assuming that technological adoption automatically improves risk outcomes. In Saudi

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723458

enterprises, the framework must connect analytics with responsible data use, internal model governance, cybersecurity, Shariah-sensitive product logic where relevant, and national objectives for financial-sector development and SME participation.

VII. PROPOSED AI-ENABLED ADVANCED DECISION INTELLIGENCE FRAMEWORK

The literature supports a layered decision-intelligence architecture rather than a single-model solution. Figure 1 organizes the proposed framework into four primary analytical layers and three control layers. The data layer consolidates audited financial statements, transaction histories, bureau information, collateral attributes, sector signals and carefully governed alternative data. The analytics layer estimates probability of default, loss given default, exposure at default, early-warning indicators and scenario responses using a portfolio of interpretable and high-capacity models. The explainability layer converts predictions into local and global evidence, while the orchestration layer maps that evidence into approvals, pricing, limits, collateral requests, covenants, referrals and monitoring intensity. Human oversight, governance controls and Saudi contextual requirements operate across all four layers rather than after model deployment.

The framework deliberately separates prediction from decision. A probability of default is not itself an approval instruction. Decision rules must combine risk estimates with exposure size, expected recovery, margin, capital cost, concentration, customer relationship value and risk appetite. This separation allows different products or segments to use the same calibrated risk signal while applying different commercial policies. It also creates a clear audit trail: the model explains estimated risk, the policy engine explains the action, and accountable staff can explain any override. Such decomposition reduces the tendency to attribute every adverse outcome to the model and encourages clearer ownership among credit, finance, data science, compliance and business teams.

The framework also recommends a model portfolio. Logistic or generalized linear models can serve as interpretable baselines, while gradient boosting, random forests, kernel methods or neural models act as challengers where they deliver material gains [5,24–27]. Advanced models should not be promoted merely because they are more complex. Promotion criteria should require out-of-sample improvement, calibration quality, stability, explanation quality, stress resilience and economic benefit. Where black-box models are selected, their outputs should be accompanied by consistent feature attributions, reason codes and counterfactual analysis. Segment-specific models may be justified for SMEs, large corporates, project finance or supply-chain exposures, but fragmentation must be controlled to avoid unmanageable model inventories.

Table 1 summarizes the evidence-to-design translation. It shows that recurring academic findings map directly to enterprise controls: ensemble performance motivates champion-challenger comparison; alternative data motivates lineage and consent controls; unstable explanations motivate explanation validation; profit-sensitive metrics motivate threshold optimization; and Saudi fintech findings motivate local governance alignment. This mapping is critical because many AI programs fail by treating governance as documentation added after technical development. In the proposed architecture, governance requirements determine model design choices from the start.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723458

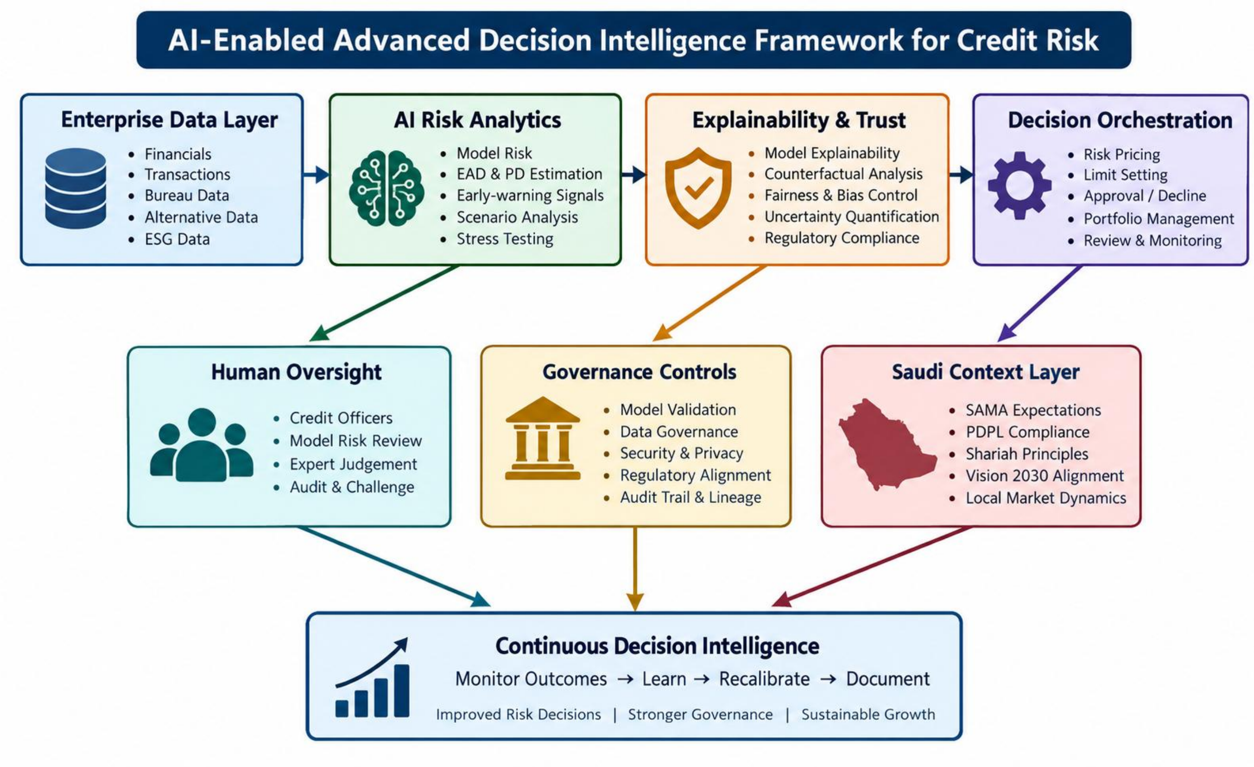

Figure 1. Proposed layered decision-intelligence architecture for enterprise credit risk.

Table 2. Recommended control metrics across the credit-AI lifecycle

Lifecycle stage	Key metrics	Evidence required	Escalation examples
Build	Missingness, class balance, feature stability, leakage tests	Data dictionary, lineage, feature rationale	Material proxy risk; unstable source; leakage
Validate	AUC/PR-AUC, calibration, expected loss, fairness, explanation stability	Independent back-test and stress report	Weak calibration; segment failure; unstable explanations
Deploy	Approval rate, override rate, pricing dispersion, processing time	Decision logs, policy rules, access controls	Override clustering; unexplained decision gaps
Monitor	Drift, defaults, recoveries, concentration, complaints, incidents	Dashboards, trigger thresholds, review minutes	Drift breach; loss spike; data-pipeline failure
Review/retire	Incremental value, control cost, residual risk	Champion-challenger comparison, closure record	No material benefit; persistent control failures

VIII. DISCUSSION

The review suggests that the core competitive advantage of AI in credit is not a single superior algorithm but the ability to compress a wider information set into timely, repeatable and challengeable risk judgments. The strongest empirical studies converge on the value of nonlinear and ensemble methods, but they also expose the cost of complexity. Predictive gains become strategically useful only when probabilities are calibrated, explanations are stable and business thresholds reflect the asymmetry of credit losses. This finding reframes the common accuracy-versus-interpretability debate. The relevant objective is not maximum accuracy or maximum transparency in isolation; it is sufficient predictive power with enough transparency, control and resilience for the intended decision.

A second insight is that explainability must be designed for different users. A data scientist needs diagnostic information about feature behavior and drift; a credit officer needs reasons that correspond to financial logic; a model-risk function needs reproducible tests; senior management needs portfolio effects; and a customer may need a concise actionable rationale. SHAP or LIME alone cannot satisfy all these purposes. Organizations should therefore maintain layered explanations: global model behavior, segment-level drivers, local decision reasons and policy-level decision logic. Explanation quality should be tested for stability and fidelity, particularly under class imbalance [11]. Human review should focus on cases where uncertainty, policy exceptions or explanation anomalies are high rather than duplicating every automated decision.

A third insight concerns the meaning of optimization. Classical credit scoring tends to optimize statistical separation between good and bad accounts. Decision intelligence optimizes an outcome function that includes expected credit loss, revenue, capital, operational cost, fairness constraints and strategic priorities. This is especially relevant for Saudi enterprises serving SMEs or new-economy sectors where thin histories can produce conservative outcomes. Alternative data may widen access [4,23], but inclusive lending should not be equated with weaker risk standards. The better approach is to improve information, quantify uncertainty and use risk-based terms. That may mean smaller initial limits, more frequent monitoring or dynamic pricing rather than a binary rejection.

A fourth insight is organizational. AI changes the division of labor in credit functions. Routine data gathering and first-line scoring can be automated, while humans concentrate on exceptions, complex structures, relationship intelligence and accountability. This requires new roles around data stewardship, model validation, AI risk and product governance. It also requires governance forums capable of resolving trade-offs between commercial growth and risk controls. The Saudi environment makes cross-functional design particularly important because digital finance is expanding quickly, product structures may need Shariah review, and enterprise data may be distributed across banks, fintech partners, ERP systems and supply-chain platforms.

Figure 2 therefore frames credit AI as a lifecycle. Define begins with risk appetite, use-case boundaries and accountability. Build covers data controls, feature engineering and algorithm choice. Validate covers back-testing, calibration, stress, fairness and explainability. Deploy covers authorization, human override and decision logging. Monitor covers drift, defaults, recoveries, complaints, portfolio concentration and incidents. Outcomes feed back into policy and model changes. The lifecycle transforms governance from a periodic validation exercise into an operating system for continuous learning.

The evidence reinforces a disciplined implementation principle credit intelligence should remain modular measurable and contestable Data controls must precede feature engineering model selection must precede automation validation must precede scale and monitoring must continue after deployment This sequencing keeps predictive innovation connected to accountable enterprise decisions and prevents local improvements in model accuracy from obscuring portfolio risk customer impact or operational fragility For Saudi organizations such discipline also creates a practical path for combining digital transformation with prudent financial governance because new data sources and algorithms can be introduced incrementally compared with established baselines and retired when evidence deteriorates The resulting architecture is therefore adaptive without becoming uncontrolled and ambitious without treating technological novelty as a substitute for credit judgment The evidence reinforces a disciplined implementation principle credit intelligence should remain modular measurable and contestable Data controls must precede feature engineering model selection must precede automation validation must precede scale and monitoring must continue after deployment This sequencing keeps predictive innovation connected to accountable enterprise decisions and prevents local improvements in model accuracy from obscuring portfolio risk customer impact or operational fragility For Saudi organizations such discipline also creates a practical path for combining digital transformation with prudent financial governance because new data sources and algorithms can be introduced incrementally compared with established baselines and retired when evidence deteriorates The resulting architecture is therefore adaptive without becoming uncontrolled and ambitious without treating technological novelty as a substitute for credit judgment The evidence reinforces a.

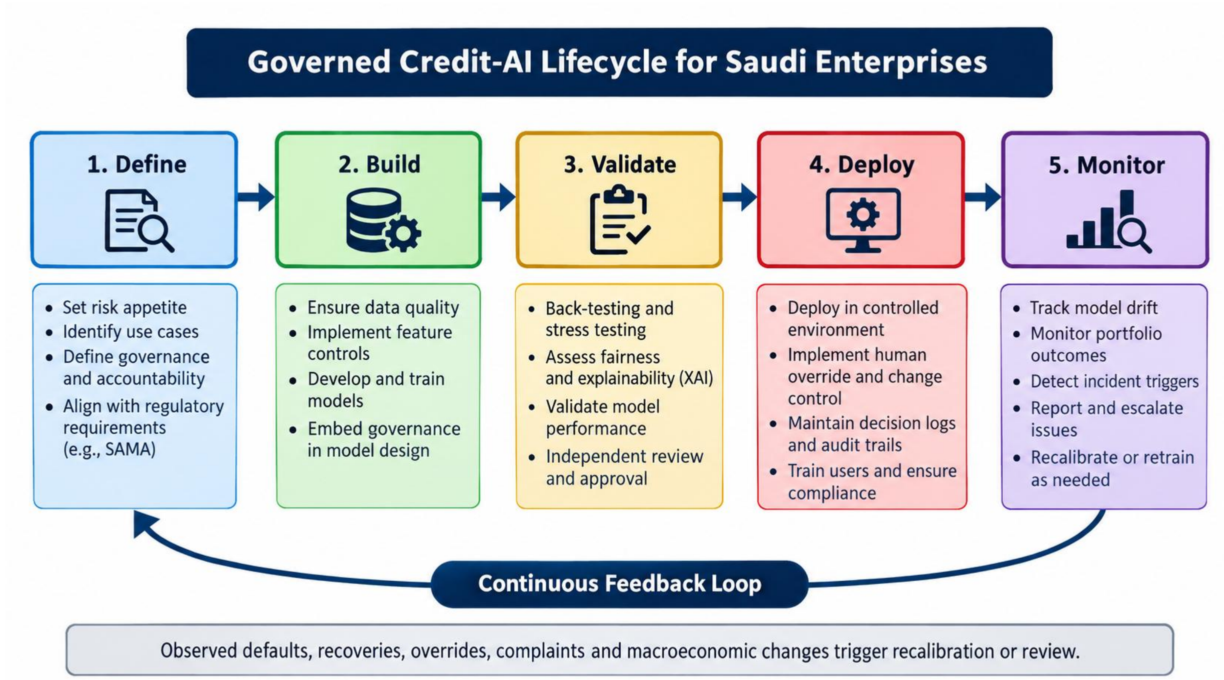

Figure 2. Governed credit-AI lifecycle with continuous feedback and escalation.

IX. RESEARCH GAPS AND FUTURE RESEARCH

Several research gaps remain. First, much of the empirical literature uses consumer or peer-to-peer datasets rather than Saudi corporate portfolios. Future studies should test AI credit models on Saudi SME, mid-market, supply-chain and Islamic-finance exposures using temporally separated samples and clearly defined default outcomes. Second, research should move beyond AUC comparisons toward calibration, economic utility, uncertainty and portfolio concentration. Third, fairness research needs localization because protected characteristics, legal requirements and culturally appropriate fairness criteria vary across jurisdictions. Fourth, explainability research should evaluate whether credit officers actually make better decisions when explanations are provided, rather than assuming interpretability from visual outputs alone.

Further work is also required on dynamic and multimodal decision intelligence. Transaction streams, invoices, contracts, ERP records and macroeconomic indicators could support early-warning systems, but the incremental value of each source should be tested against privacy and operational cost. Federated or privacy-preserving learning may become relevant where sensitive data cannot be centralized. Generative AI could assist narrative credit memos or policy retrieval, yet its outputs should be separated from validated risk estimates and subjected to source grounding, access controls and audit logging. Research on agentic workflows should examine whether autonomous systems can safely coordinate document collection, covenant monitoring and exception routing while preserving human authorization for consequential decisions.

Saudi-specific research should also study the interaction among AI, Shariah governance and product economics. Credit-equivalent decisions in murabaha, ijara, trade finance or structured facilities may require different features, loss assumptions and decision rationales. Another priority is measuring how AI-enabled underwriting affects SME access, rejection rates, pricing dispersion and portfolio resilience over time. Finally, independent benchmarking across Saudi institutions could establish common practices for explanation stability, drift thresholds, validation evidence and incident reporting without requiring disclosure of proprietary models.

X. PRACTICAL, MANAGERIAL AND POLICY IMPLICATIONS

For managers, implementation should begin with a clearly bounded use case rather than an enterprise-wide AI transformation. A practical sequence is to establish a transparent baseline model, introduce one or two challenger models, define economic and risk metrics, and create a controlled workflow for comparison. Data lineage should be documented from source to decision, including transformation logic, missing-value treatment and feature ownership. Approval thresholds should be linked to risk appetite and expected economics, not copied from model-development cutoffs. High-impact overrides should require reason codes and periodic analysis so that management can distinguish valuable expert judgment from systematic policy leakage.

For risk and compliance functions, the central requirement is evidence. Model inventories should identify owners, intended use, training data, validation status, limitations, explanation methods and monitoring triggers. Independent validation should test performance by time period and segment, not merely on random holdout samples. Stress tests should include worsening macroeconomic conditions, deterioration in default prevalence, missing alternative data and changes in borrower mix. Fairness and explanation testing should be incorporated where legally and contextually appropriate. Cybersecurity reviews are necessary because decision intelligence depends on interconnected data pipelines and APIs whose compromise could alter credit outcomes even if the model itself remains unchanged.

For policymakers and ecosystem institutions, the priority is to enable innovation without weakening accountability. Regulatory experimentation can support alternative-data lending, open finance and SME credit models while requiring clear governance for data provenance, explainability and human recourse. Common technical guidance on model documentation, monitoring and third-party AI risk would reduce uncertainty for banks and fintechs. Capacity building is equally important: credit professionals need enough data literacy to challenge models, while data scientists need enough credit knowledge to understand default definitions, covenants, collateral and portfolio effects. The framework aligns with Saudi economic diversification by supporting more scalable credit analysis, but its contribution to Vision 2030 depends on responsible implementation rather than adoption volume alone.

XI. LIMITATIONS

This review has four principal limitations. It is an integrative synthesis rather than a meta-analysis, so it does not estimate pooled performance effects. The selected literature is dominated by banking and consumer-credit datasets from outside Saudi Arabia, which limits direct generalization to Saudi corporate portfolios. Peer-reviewed evidence on Saudi AI-enabled credit risk remains comparatively small, requiring contextual inferences from broader digital-finance studies. Finally, the field evolves rapidly; emerging generative and agentic AI applications may develop faster than peer-reviewed validation. These limitations reinforce the paper's emphasis on governance, local validation and empirical testing before enterprise deployment.

XII. CONCLUSION

AI can materially improve credit risk management, but only when it is embedded in a decision system that is broader than the predictive model. Across the reviewed literature, advanced methods frequently strengthen discrimination and capture nonlinear borrower behavior, while explainability techniques make complex models more accessible to decision-makers. The same evidence also shows recurring vulnerabilities: imbalanced data, unstable explanations, dataset shift, weak calibration, opaque thresholds and a tendency to evaluate models with statistical metrics that do not fully represent lending economics. These weaknesses explain why technical sophistication without governance can increase rather than reduce model risk.

The proposed framework addresses this problem by integrating enterprise data, multi-model analytics, explainability, risk appetite, policy orchestration, human oversight and continuous monitoring. For Saudi enterprises, the architecture should be adapted to local digital-finance development, responsible data governance, Shariah-sensitive contexts where applicable and national priorities for SME growth and economic diversification. The appropriate strategic goal is not automated credit at any cost. It is accountable decision intelligence: a system in which AI expands analytical capacity, humans retain responsibility for consequential judgment, and every model-driven action can be evaluated through its predictive, economic, ethical and operational outcomes.

REFERENCES

[1] X. Dastile, T. Celik, and M. Potsane, "Statistical and machine learning models in credit scoring: A systematic literature survey," *Applied Soft Computing*, vol. 91, Art. no. 106263, 2020, doi: 10.1016/j.asoc.2020.106263. Crossref

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

[3] S. Sachan, J.-B. Yang, D.-L. Xu, D. E. Benavides, and Y. Li, "An explainable AI decision-support-system to automate loan underwriting," *Expert Systems with Applications*, vol. 144, Art. no. 113100, 2020, doi: 10.1016/j.eswa.2019.113100. Crossref

[4] V. B. Djeundje, J. Crook, R. Calabrese, and M. Hamid, "Enhancing credit scoring with alternative data," *Expert Systems with Applications*, vol. 163, Art. no. 113766, 2021, doi: 10.1016/j.eswa.2020.113766. Crossref

[5] V. Moscato, A. Picariello, and G. Sperlí, "A benchmark of machine learning approaches for credit score prediction," *Expert Systems with Applications*, vol. 165, Art. no. 113986, 2021, doi: 10.1016/j.eswa.2020.113986. Crossref

[6] B. R. Gunnarsson, S. vanden Broucke, B. Baesens, M. Oskarsdottir, and W. Lemahieu, "Deep learning for credit scoring: Do or don't?," *European Journal of Operational Research*, vol. 295, no. 1, pp. 292–305, 2021, doi: 10.1016/j.ejor.2021.03.006. Crossref

[7] E. Dumitrescu, S. Hue, C. Hurlin, and S. Tokpavi, "Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects," *European Journal of Operational Research*, vol. 297, no. 3, pp. 1178–1192, 2022, doi: 10.1016/j.ejor.2021.06.053. Crossref

[8] J. A. Bastos and S. M. Matos, "Explainable models of credit losses," *European Journal of Operational Research*, vol. 301, no. 1, pp. 386–394, 2022, doi: 10.1016/j.ejor.2021.11.009. Crossref

[9] A. C. Bueff, M. Cytrynski, R. Calabrese, M. Jones, J. Roberts, J. Moore, and I. Brown, "Machine learning interpretability for a stress scenario generation in credit scoring based on counterfactuals," *Expert Systems with Applications*, vol. 202, Art. no. 117271, 2022, doi: 10.1016/j.eswa.2022.117271. Crossref

[10] G. Babaei, P. Giudici, and E. Raffinetti, "Explainable FinTech lending," *Journal of Economics and Business*, vols. 125–126, Art. no. 106126, 2023, doi: 10.1016/j.jeconbus.2023.106126. Crossref

[11] Y. Chen, R. Calabrese, and B. Martin-Barragan, "Interpretable machine learning for imbalanced credit scoring datasets," *European Journal of Operational Research*, vol. 312, no. 1, pp. 357–372, 2024, doi: 10.1016/j.ejor.2023.06.036. Crossref

[12] H. Mohammadnejad-Daryani, A. A. Taleizadeh, and D. Pamucar, "A novel profit-driven framework for model evaluation in credit scoring," *Engineering Applications of Artificial Intelligence*, vol. 137, Art. no. 109137, 2024, doi: 10.1016/j.engappai.2024.109137. Crossref

[13] S. Shi, R. Tse, W. Luo, S. D'Addona, and G. Pau, "Machine learning-driven credit risk: A systemic review," *Neural Computing and Applications*, vol. 34, pp. 14327–14339, 2022, doi: 10.1007/s00521-022-07472-2. Crossref

[14] G. F. Bone-Winkel and F. Reichenbach, "Improving credit risk assessment in P2P lending with explainable machine learning survival analysis," *Digital Finance*, vol. 6, pp. 501–542, 2024, doi: 10.1007/s42521-024-00114-3. Crossref

[15] D. Chen, J. Ye, and W. Ye, "Interpretable selective learning in credit risk," *Research in International Business and Finance*, vol. 65, Art. no. 101940, 2023, doi: 10.1016/j.ribaf.2023.101940. Crossref

[16] N. Kozodoi, J. Jacob, and S. Lessmann, "Fairness in credit scoring: Assessment, implementation and profit implications," *European Journal of Operational Research*, vol. 297, no. 3, pp. 1083–1094, 2022, doi: 10.1016/j.ejor.2021.06.023. Crossref

[17] P. Giudici, M. Centurelli, and S. Turchetta, "Artificial Intelligence risk measurement," *Expert Systems with Applications*, vol. 235, Art. no. 121220, 2024, doi: 10.1016/j.eswa.2023.121220. Crossref

[18] A. A. H. A. Khatir and M. Bee, "Machine learning models and data-balancing techniques for credit scoring: What is the best combination?," *Risks*, vol. 10, no. 9, Art. no. 169, 2022, doi: 10.3390/risks10090169. MDPI

[19] V. Chang, Q. A. Xu, S. H. Akinloye, V. Benson, and K. Hall, "Prediction of bank credit worthiness through credit risk analysis: An explainable machine learning study," *Annals of Operations Research*, vol. 354, pp. 247–271, 2025, doi: 10.1007/s10479-024-06134-x. Crossref

[20] M. K. Nallakaruppan, B. Balasubramanian, B. Balusamy, M. L. Shri, V. Malathi, M. Veeraraghavan, and S. Bhattacharyya, "An Explainable AI framework for credit evaluation and analysis," *Applied Soft Computing*, vol. 153, Art. no. 111307, 2024, doi: 10.1016/j.asoc.2024.111307. Crossref

[21] D. Ahelegbey, P. Giudici, and V. Pediroda, "A network based fintech inclusion platform," *Socio-Economic Planning Sciences*, vol. 87, Art. no. 101555, 2023, doi: 10.1016/j.seps.2023.101555. ScienceDirect

[22] D. Tripathi, D. R. Edla, V. Kuppili, and A. Bablani, "Evolutionary Extreme Learning Machine with novel activation function for credit scoring," *Engineering Applications of Artificial Intelligence*, vol. 96, Art. no. 103980, 2020, doi: 10.1016/j.engappai.2020.103980. Crossref

[23] W. Liu, H. Fan, and M. Xia, "Step-wise multi-grained augmented gradient boosting decision trees for credit scoring," *Engineering Applications of Artificial Intelligence*, vol. 97, Art. no. 104036, 2021, doi: 10.1016/j.engappai.2020.104036. Crossref

[24] W. Zhang, D. Yang, S. Zhang, J. H. Ablanedo-Rosas, X. Wu, and Y. Lou, "A novel multi-stage ensemble model with enhanced outlier adaptation for credit scoring," *Expert Systems with Applications*, vol. 165, Art. no. 113872, 2021, doi: 10.1016/j.eswa.2020.113872. Crossref

[25] C.-F. Wu, S.-C. Huang, C.-C. Chiou, and Y.-M. Wang, "A predictive intelligence system of credit scoring based on deep multiple kernel learning," *Applied Soft Computing*, vol. 111, Art. no. 107668, 2021, doi: 10.1016/j.asoc.2021.107668. Crossref

[26] V. Chang, S. Sivakulasingam, H. Wang, S. T. Wong, M. A. Ganatra, and J. Luo, "Credit risk prediction using machine learning and deep learning: A study on credit card customers," *Risks*, vol. 12, no. 11, Art. no. 174, 2024, doi: 10.3390/risks12110174. MDPI

[27] N. Suhadolnik, J. Ueyama, and S. Da Silva, "Machine learning for enhanced credit risk assessment: An empirical approach," *Journal of Risk and Financial Management*, vol. 16, no. 12, Art. no. 496, 2023, doi: 10.3390/jrfm16120496. MDPI

[28] H. H. Al-Baity, "The Artificial Intelligence revolution in digital finance in Saudi Arabia: A comprehensive review and proposed framework," *Sustainability*, vol. 15, no. 18, Art. no. 13725, 2023, doi: 10.3390/su151813725. MDPI

[29] A. Eskandarany, "Adoption of artificial intelligence and machine learning in banking systems: A qualitative survey of board of directors," *Frontiers in Artificial Intelligence*, vol. 7, Art. no. 1440051, 2024, doi: 10.3389/frai.2024.1440051. Frontiers

[30] H. H. Khan, S. Khan, and A. Ghafoor, "Fintech adoption, the regulatory environment and bank stability: An empirical investigation from GCC economies," *Borsa Istanbul Review*, vol. 23, no. 6, pp. 1263–1281, 2023, doi: 10.1016/j.bir.2023.10.010. ScienceDirect