



Artificial Intelligence And Machine Learning Predictive Models In Financial Risk Management: A Comparative Study Of Iraqi Banks

Saad Dyih Muhsin

*Assistant Lecturer, Ibn Khaldun Private University College, Baghdad – Iraq
Email: Saadisamohsin1961@gmail.com*

Muhsin, S.D., (2026). Artificial Intelligence And Machine Learning Predictive Models In Financial Risk Management: A Comparative Study Of Iraqi Banks, *Journal of Language and Linguistic Studies*, 22(1), 59-74; 2026.

ABSTRACT

This study deals with assessing the feasibility and effectiveness of using AI and ML models in Iraqi banks for financial risk management, which are currently facing a critical situation regarding the non-performing loan ratio, reaching 14.2%, compared to 4.5% on average in the MENA region.

This research follows a mixed-methods research design, and this paper analyzes 10 banks from Iraq, which constitute 75% of industry assets, using various machine learning models on 125,487 loans between 2015 and 2024, and 38 interviews were conducted on stakeholders. Various machine learning models have been applied, such as Random Forest, XGBoost, Neural Networks, and LSTM.

Moreover, the results show that ML techniques were significantly better than traditional techniques in terms of accuracy, i.e., 84.6% vs. 68.4% (+16.2% pp) for the classification, 71.8% default detection vs. 42.8% (+29.0% pp), and AUC-ROC 0.885 vs. 0.642, at a $p < 0.001$ level. The economic gains show that the system can win 227.2 million dollars annually or cut 50.6% error costs. The cost savings from the system implementation will be 3 million dollars over a period of 3 years, rewarding a NPV and IRR of 1.9 million dollars and 54.3%, respectively.

The research validates three research hypotheses: technological readiness and data quality affect the feasibility of the implementation ($R^2=0.847$), ML models have a significant superiority over conventional models ($d=2.83$), and a phased implementation approach can achieve a success rate of 85%.

The study identified that the key factors to the success of the implementation are: executive support ($\beta=0.68$), data quality

Keywords: Artificial Intelligence, Machine Learning, Financial Risk Management, Credit Risk, Iraqi Banks, XGBoost, Predictive Models, Emerging Markets.

1. INTRODUCTION

The world financial environment has undergone an unprecedented revolution with the incorporation of artificial intelligence (AI) and machine learning (ML). The financial environment of Iraq is very challenging and unique, with economic and geopolitical risks and structural weaknesses. Most of the banks in Iraq still use traditional risk measurement systems like Value at Risk (VaR) and other traditional risk measurement systems, which are found to be inadequate to handle the complex and nonlinear relationships of the financial environment of Iraq.

A very powerful resource which can help support the banks in this task is the function of machine learning, which facilitates the processing of tremendous volumes of information, establishment of complex relationship recognition, and executing decisions in real time. Banks in Iraq have their particular problems, which include the lack of both technological and human assets, in addition to regulatory risks.

Such research would help overcome the said challenge with an in-depth examination of the viability, efficacy, and implementation of AI/ML-based risk management systems in the banking sector within the financial landscape of Iraq.

2. RESEARCH PROBLEM

The predictive powers of Iraqi banks are inadequate to identify and measure financial risks properly and address them timely and precisely. This results in a considerably higher default loan ratio than other MENA region countries (14.2% versus the average of 4.5%), unexpected losses during economic shocks, and capital inefficiency. Traditional risk models undershoot tail risks considerably and do not detect weakening credit qualities in a timely way.

Research Questions:

RQ1: How feasible is the implementation of machine learning models within Iraqi banks?

RQ2: To what extent do machine learning models show performance that is considerably better than the traditional methods?

RQ3: What recommendations are there, which can aid in the successful implementation of AI/ML-based risk management systems in Iraqi banks?

3. RESEARCH SIGNIFICANCE

Theoretical Significance

- Expands the current AI/ML risk management models to an extreme emerging market scenario
- Combines technical model performance and organizational implementation factors
- Expands existing technology adoption research carried out in banks of developing countries

Practical Significance:

- Provides bankers in Iraq concrete proposals to modernize their risk management systems
- Informs Central Bank of Iraq regulatory policies in advanced risk models
- Helps in the economic development of Iraq through credit allocation and financial stabilization

Timeliness: The enhanced digitalization stemming from the post-pandemic, in concert with the imperative of finance regulation compliance in Iraq and its financial diversification, signals the timeliness of such research.

4. RESEARCH OBJECTIVES

Primary Objective: The primary objective of this research is to evaluate the feasibility and effectiveness of the application of Artificial Intelligence and Machine Learning models in the financial risk management of Iraqi Banks.

Secondary Objectives:

Objective 1 (→RQ1): To assess the current levels of technology, data, human capital, and legal infrastructure within the Iraqi banking system, which can be utilized for deploying an ML model.

Objective 2 (→RQ2): To experiment and compare the efficiency of various ML methods and the traditional approach using practical bank data from the Iraqi banking sector.

Objective 3 (→RQ3): Give detailed suggestions on the application of ML models in the Iraqi banking industry based on evidence.

5. RESEARCH METHODOLOGY

Research Approach: The researcher is going to use a mixed research approach in quantitative empirical research and qualitative case study research.

Data Collection:

- Quantitative: Data collection for this study involves loan-related data collected from 10 banks in Iraq, covering as many as 125,487 loans and extending from
- Qualitative: Semi-structured interviews with 38 Risk Managers, IT Directors, and Executives.
- Secondary: The publications from the Central Bank, market information, and macroeconomic indicators would be utilized.

Data Analysis:

- Implementation of Machine Learning: Random forests, XGBoost, neural networks, and LSTM will be used among other algorithms in the implementation of machine learning.

Performance Metrics: The different metrics chosen for measuring the performance of the created models are accuracy, precision, recall, F1 score, AUC ROC, MAE, and RMSE.

- Statistical Testing: Regression analysis will be used to perform the statistical testing for the success factors.

Thematic Analysis: Thematic analysis shall be conducted on the data collected from interviews, which shall be carried out during the research process.

Validity and reliability:

- Cross-validation will be used for its internal validity, and a wide range of banks will.

6. RESEARCH POPULATION AND SAMPLE

Population: All 58 licensed commercial banks in Iraq, (7 State-owned, 43 Private, 8 Islamic)

Sampling Method: Purposive sampling on the basis of available data, asset size, and willingness to participate

Sample Composition:

Bank Category	Population	Sample	Sampling Rate
Large State Banks	3	2	67%
Medium State Banks	4	1	25%
Large Private Banks	8	3	38%
Medium Private Banks	15	2	13%
Islamic Banks	8	2	25%
TOTAL	38	10	26%

Sample represents: 75% of the total banking sector's asset base due to the diversity of the sample, though not so extensive an amount that makes it hard to scrutinize.

7. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Today, the modern theory of banking places a big belief in the management of risk, that is to say, measurement, pricing, and managing the risks, not eliminating them per se. Traditional theories are based upon normal distribution curves and a stable state. This condition is not feasible for an emerging economy like Iraq with the volatile state of its markets.

International evidence: Khandani et al. (2010) show that the ML model can raise the accuracy of the US markets' default predictions by 6-15%. Lessmann et al. (2015) prove that tree-based ensemble models are bound to rank within the top three on credit data sets.

Evidence of Emerging Markets: Moscatelli et al. (2020) noted that when contingency situations occur, ML models based on previous crisis information showed poor performance and reinforced the need for placing more emphasis on stress tests in uncertain markets.

Bharti et al. (2022) found that the best approach was XGBoost, although the simpler approaches were more explainable to the regulators.

Challenges in the Implementation of AI Solutions: According to Deloitte, in a report released in 2019, financial institutions were able to successfully implement AI solutions on a large scale only to the extent of 32%, where the challenges faced were related to data quality (67%), skills (54%), and resistance from users (52%).

Research Gap: There is no specific literature on the applications of AI/ML in Iraqi banks, which results in a research gap that this paper aims to fulfill.

HYPOTHESIS DEVELOPMENT:

H1: Feasibility Hypothesis (→RQ1 & Objective 1)

H1: There is a strong positive relationship between the level of maturity of the technology infrastructure, the availability of human capital resources and the quality of the data, and the feasibility of implementing the machine learning model of risk management in banks operating in Iraq.

Rationale

Given that when an organization adopts technology, the success of that technology is subject to the level of preparation of the organization, this is according to the theory of technology adoption, and in the case of Iraqi banks, different dimensions exist.

H2: Performance Hypothesis (→RQ2 & Objective 2)

H2: Machine learning models (Random Forest, XGBoost, Neural Networks) will yield significantly higher accuracy and lower risk costs compared to traditional models (logistic regression, simple VaR) in Iraqi bank credit and market risk management.

Rationale: Prior studies worldwide have proven the superiority of ML models in a stable environment. This hypothesis aims at validating these findings for a highly volatile environment like Iraq.

H3: Implementation Success Hypothesis (→RQ3 & Objective 3)

H3: Phased implementation methods beginning with credit risk applications, using dedicated AI/ML teams, and securing executive management commitment yield significantly better implementation success rates compared with alternative methods for banks in Iraq.

Rationale: The choice of implementation strategies is an important factor in implementation success. There are resource and capability constraints for banks in Iraq, which might make some implementation methods more appropriate than others.

RESEARCH TERMINOLOGY

Artificial Intelligence (AI): Computer systems with the ability to carry out tasks that are normally associated with human intelligence, such as perception, recognition of speech, making decisions, and recognizing patterns, among others (Goodfellow, Bengio, & Courville, 2016). AI in the banking sector involves different technologies, such as machine learning, natural language processing, and expert systems, for the purpose of automation in finance (Ernst & Young, 2020).

Machine Learning (ML): A subset of artificial intelligence that allows computer systems to learn and improve with experience, as opposed to traditional programming (James, Witten, Hastie, & Tibshirani, 2013). Machine learning algorithms use sample data, referred to as training data, to build mathematical models that predict outcomes or make decisions without explicit programming (Hastie, Tibshirani, & Friedman, 2009).

Supervised Learning: "A machine learning approach in which the algorithm learns from the training data with the goal of mapping inputs to the correct outputs" (James et al., 2013). In the context of credit risk prediction, supervised learning involves training the model to learn from the historical loans that were either "defaulted" or "repaid" to make predictions about future loans (Baesens et al., 2003; Lessmann et al., 2015).

Ensemble Methods: "Machine learning methods that combine multiple models to improve the performance of the prediction" (Hastie et al., 2009). Some popular ensemble methods include Random Forest, which is the combination of decision trees, and XGBoost, which is the combination of weak learners with the help of the gradient boosting algorithm (Breiman, 2001; Chen & Guestrin, 2016).

Random Forest: "An ensemble learning approach that involves creating multiple decision trees during the training phase, and the predicted output is the mode of the predicted classes for classification problems, and the average prediction of the decision trees in the ensemble for regression problems" (Breiman, 2001). It is a very accurate approach with very minimal chances of overfitting; hence, it is suitable for credit scoring problems according to Brown & Mues, 2012.

XGBoost or Extreme Gradient Boosting: Improved gradient boosting, where models are constructed greedily one by one, correcting mistakes of its predecessor. Proposed by Chen & Guestrin in 2016. Quite popular in finance, since it is the most predictive and computationally efficient according to Lessmann et al. (2015).

Neural Networks (ANN): Computing systems that are inspired by biological networks, with nodes called neurons that are connected in layers. Introduced by LeCun, Bengio, & Hinton in 2015. Neural networks are able to learn complex non-linear relationships between inputs and outputs using a process called backpropagation, as explained in Goodfellow et al. (2016).

LSTM (Long Short-Term Memory): A type of recurrent neural network that is able to learn from sequential data while avoiding the problem of vanishing gradients. Introduced by Hochreiter &

Schmidhuber in 1997. Used for time series forecasting in financial markets, as reported in Fischer & Krauss (2018) and Kim & Won (2018).

Deep Learning: A type of machine learning based on artificial neural networks with many layers (deep networks) (LeCun et al., 2015). Able to automatically learn hierarchical representations of the world and perform complex pattern recognition tasks (Goodfellow et al., 2016).

Credit Risk: The possibility of loss due to the non-payment of a loan or the failure to fulfill a contractual obligation (Hull, 2018). It encompasses default risk, concentration risk, and country risk and is considered one of the most important risks for financial institutions (Basel Committee on Banking Supervision, 2006).

Market Risk: The possibility of loss in financial positions due to unfavorable changes in market prices such as interest rates, foreign exchange rates, stock prices, and commodity prices (Hull, 2018; Basel Committee on Banking Supervision, 2019).

Value at Risk (VaR): "A statistical measure that estimates the potential loss in value of a portfolio over a given time period with a given confidence level" (Hull, 2018). For instance, 1-day 95% VaR of \$10 million implies that there is 95% confidence that losses will not exceed \$10 million in 1 day.

Non-Performing Loan (NPL): "A loan that is not being serviced in accordance with the terms agreed to in the contract, meaning the borrower is in default and has not made payments of principal and interest in the specified time period, usually 90 days or more" (International Monetary Fund, 2023). High levels of NPLs may also suggest deteriorating asset quality, which might lead to unstable financial systems within the country.

Default: "Failure to perform the legal obligations of the loan, which means the borrower is not making the required payments, among other things" (Basel Committee on Banking Supervision, 2006). Even if a default occurs, legal actions are typically taken for recovering losses (Hull, 2018).

Accuracy: The ratio of correct predictions, including true positives and true negatives, to the total number of predictions (James et al., 2013). Formula: $(TP + TN) / (TP + TN + FP + FN)$. Accuracy is a commonly used measure, although it is subject to certain pitfalls, as discussed in Chawla et al. (2002).

Precision: The ratio of true positive predictions to the total number of positive predictions (Pedregosa et al., 2011). Precision is a measure of the number of predictions of defaulters who have indeed defaulted. Formula: $TP / (TP + FP)$. Precision is an important metric as it reduces false alarms and risk aversion (Baesens et al., 2003).

Recall (Sensitivity): It is defined as the ratio of positive predictions to the actual number of positive instances. Recall is defined as the ratio of the number of positive predictions to the total number of positive instances (James et al., 2013). Recall signifies the number of actual defaulters that are predicted correctly. It is extremely important from the point of view of risk management because, from this measure, false alarms are minimized, i.e., actual defaulters (Lessmann et al., 2015).

F1 Score: This is the harmonic mean of both the precision and recall values. This is essential because we want to get both the precision and the recall into one formula. This becomes even more significant when we are dealing with unbalanced classes, meaning there is no uniformity. $F1 \text{ Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$

AUC-ROC: A measure of model performance for classification problems, which measures how well the model can discriminate between classes for all possible classification thresholds. Its value varies between 0 and 1, where 1 means that all instances are classified correctly. Any value above 0.8 implies that the model is performing well.

Backtesting: Backtesting is a process that involves checking how well a model would have performed in the past using data that was not used in model development. Backtesting is important, especially for validating model performance for risk models, such as Value at Risk, as required by regulators.

Stress Testing: This is where the reaction of the model/portfolio is analyzed to extreme yet reasonable bad scenarios, which could be a financial recession or stock exchanges crashing, as proposed by the Basel Committee on Banking Supervision (2022). This is a vital method to assess the robustness of the model during economically challenging periods, as demonstrated by Moscatelli et al. (2020).

Feature Engineering: This can be described as the art of creating new features from the data using a process that involves the effectiveness of the machine learning algorithms working with the new versions of the data (Hastie et al., 2009). This can be described as the most labor-intensive, however arguably essential part of the entire process of creating effective machine learning models (Gu, Kelly, & Xiu, 2020).

Overfitting: This is usually a problem in modeling, whereby the model is too “accurate” for the information it was trained on, which includes noise and outliers, and therefore does not perform well when new information is introduced (James et al., 2013). This is usually in complex modeling when little information is available, as controlled in the model by cross-validation and regularization (Hastie et al., 2009).

Hyperparameter Tuning: The task of optimizing parameters controlling the learning process itself (as opposed to the parameters learned during the process), like learning rate, number of trees, or depth of the network (Chen & Guestrin, 2016). Usually done via grid search or random search with cross-validation (Pedregosa et al., 2011).

Cross-Validation: A resampling method for assessing machine learning models by splitting data into several subsets, training on some and testing on others, then rotating the subset used for testing (James et al., 2013). k-fold cross-validation is the most popular variant, giving reliable estimates of performance (Hastie et al., 2009).

SMOTE (Synthetic Minority Over-sampling Technique): A method for dealing with imbalanced datasets by synthesizing examples of the minority class (e.g., defaults in credit data) to balance the training dataset (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). Deals with the issue of rare events being underrepresented in the training dataset.

Basel Accords: International banking regulations set by the Basel Committee on Banking Supervision that establish the minimum capital requirements for banking institutions (Basel Committee on Banking Supervision, 2017). The recent version of the Basel III accords contains stronger banking risk management rules and was introduced as a reaction to the 2008 financial crisis.

Capital Adequacy Ratio (CAR): The ratio of the capital of the bank to its weighted assets, which is maintained for the safety of depositors and the stability of financial systems (Basel Committee on Banking Supervision, 2017). The Basel III accords stipulate that the minimum CAR for banking institutions should be 10.5%, including the capital conservation buffer.

Model Governance: The set of rules and procedures that ensure the development and use of models are appropriate (Office of the Comptroller of the Currency, 2021). This is important for regulatory compliance and for managing banking risks (Financial Conduct Authority, 2022).

Explainable AI (XAI): The techniques and methods that allow humans to understand decisions made by AI models (Molnar, 2022). This includes SHAP (SHapley Additive exPlanations) (Lundberg & Lee, 2017) and LIME (Local Interpretable Model-agnostic Explanations) (Ribeiro, Singh, & Guestrin, 2016). Becoming increasingly mandatory for regulators who are concerned about “black box” models (European Banking Authority, 2020).

Technology Readiness: An organization's readiness for new technology and its capabilities in terms of its existing infrastructure and technology capabilities (Deloitte, 2019). An important factor for successful implementation of AI/ML technologies in financial institutions (PwC, 2021).

Change Management: The process for managing change in an organization and helping individuals and teams adapt from their current state to their desired state (McKinsey & Company, 2020). An important process for successful technology adoption and change management in an organization, which covers people and cultural dimensions as well as technology itself (Capgemini & Efma, 2021).

Data Governance: This is considered the overall management of data availability, usability, integrity, and security in an organization (European Banking Authority, 2020). This includes data policies, procedures, standards, and metrics that support the use of data assets in an organization (Monetary Authority of Singapore, 2021).

Regulatory Sandbox: A structure or framework which permits businesses to conduct experiments using the venture of new products and/or services and/or new business units in a controlled and/or monitored setting with reduced regulatory requirements (Financial Conduct Authority, 2022). The current trend among financial sector regulators and/or actors concerning the facilitation and/or development of financial technology, and/or the monitoring of risk (Monetary Authority of Singapore, 2021).

Digital Transformation: This is defined as a process of fully integrating digital technology into all areas of a business, hence transforming a business and how it delivers value to its customers and other concerned actors (McKinsey & Company, 2020). This, therefore, includes modernization of foundational technology platforms, channel digitalization, and technology-intelligent advanced analytics for financial institutions (Accenture, 2020).

Fintech (Financial Technology): Adoption of technology with the intention of creating an enabling environment for the advancement of financial services with the potential of creating new business, applications, processes, and products (Financial Stability Board, 2017). Fintech is considered the major disruptor and transformer of the world's banking landscape (Deloitte, 2022).

8. EMPIRICAL STUDY

Table 8.1: Dataset Characteristics

Data Category	Observations	Time Period	Variables	Missing Data	Source
Credit Portfolio	125,487 loans	2015-2024	47 features	3.2%	10 Iraqi Banks
Default Indicators	125,487 loans	2015-2024	Binary flag, time to default	0.8%	Banks B01-B10
Market Data	2,520 daily obs	2015-2024	Prices, rates, volatility	0.3%	Iraq Stock Exchange, CBI
Interview Data	38 interviews	2025	Qualitative responses	N/A	Stakeholders
Training Set	87,841 (70%)	2015-2022	79 features	Preprocessed	Model training
Validation Set	12,549 (10%)	2023	79 features	Preprocessed	Hyperparameter tuning
Test Set	25,097 (20%)	2024	79 features	Preprocessed	Final evaluation

Source: Primary data collected from 10 Iraqi banks (B01 to B10) during the field study in 2025, Market data from Iraq Stock Exchange (2024), Central Bank of Iraq statistical bulletins (2015 – 2024), Prepared by researcher based on James et al. (2013) and Hastie et al. (2009) methodological guidelines.

Table 8.2: Credit Portfolio Distribution

Characteristic	Category	Loans (n)	% of Total	Avg Default Rate
Loan Size	Small (< \$50K)	68,250	54.4%	8.2%
	Medium (\$50K-\$500K)	42,180	33.6%	11.5%
	Large (> \$500K)	15,057	12.0%	15.7%
Borrower Type	Corporate	32,450	25.9%	14.3%
	SME	48,120	38.3%	12.8%
	Consumer/Retail	44,917	35.8%	7.9%
Status	Performing	107,820	85.9%	N/A
	30-90 days past due	6,850	5.5%	N/A
	90+ days past due	5,420	4.3%	N/A
	Defaulted	5,397	4.3%	N/A
TOTAL		125,487	100%	11.8%

Source: Primary data collection from loan portfolio data of 10 participating Iraqi banks (2015-2024); Analysis done by researcher according to the credit risk classification standards set by the Basel Committee on Banking Supervision (2006) and Hull (2018).

Table 8.3: Model Performance Comparison - Credit Risk

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Training Time
M1: Current Bank Practice	68.4%	31.2%	42.8%	36.1%	0.642	Manual (days)
M0: Logistic Regression (Baseline)	72.8%	38.5%	54.2%	45.1%	0.723	3 min
M2: Random Forest	81.3%	52.7%	68.4%	59.5%	0.847	45 min
M3: XGBoost	84.2%	58.3%	72.1%	64.5%	0.881	38 min
M4: Neural Network	82.7%	54.9%	70.3%	61.6%	0.862	82 min
M5: LSTM	81.9%	53.8%	69.1%	60.5%	0.854	156 min
M6: Ensemble (M2+M3+M4)	84.6%	59.1%	71.8%	64.9%	0.885	5 min

Key Finding: Best ML Model (Ensemble) shows 16.2 percentage points increase in accuracy and 29.0 percentage points improvement in default detection than current practice.

Source: Results from the empirical testing done using Python 3.11, Scikit-learn library (Pedregosa et al., 2011), on the bank data from Iraq by the researcher; Models' performance metrics were validated using the benchmarking methodologies outlined in Baesens et al. (2003) and Lessmann et al. (2015); XGBoost algorithm implementation using Chen & Guestrin (201

Table 8.4: Economic Impact Analysis

Cost Category	Current Practice	ML Models	Improvement	Annual Savings
False Negatives (Bad loans approved)	7,150 loans	3,520 loans	-50.8%	\$209.4M
False Negative Cost	\$412.8M	\$203.4M	-50.7%	
False Positives (Good loans rejected)	4,820 loans	2,450 loans	-49.2%	\$17.8M
False Positive Cost	\$36.2M	\$18.4M	-49.2%	
Total Error Cost	\$449.0M	\$221.8M	-50.6%	\$227.2M
Loan Processing Time	18.5 days	0.5 days	-97.3%	Efficiency gain
Processing Cost per Loan	\$425	\$180	-57.6%	\$245 per loan

Source: Cost calculation using actual data for loans drawn from 10 Iraqi banks; Cost of False Positive & False Negative accounting for data sourced from financial reports of banks; Methodology mostly drawn from Khandani et al. (2010); Brown & Mues (2012).

Table 8.5: Performance During Economic Stress

Period	Economic Condition	Traditional AUC	ML Model AUC	Performance Gap	Actual Default Rate
2015	Stable (pre-oil shock)	0.695	0.892	+0.197	9.2%
2016	Severe stress (oil crash)	0.582	0.821	+0.239	18.7%
2017-2019	Recovery	0.658	0.878	+0.220	11.4%
2020	COVID-19 pandemic	0.611	0.843	+0.232	16.3%
2021-2022	Post-COVID recovery	0.672	0.886	+0.214	10.8%
2023-2024	Relative stability	0.708	0.891	+0.183	9.5%
Average		0.654	0.869	+0.215	12.2%

Key Finding: ML model sustains high performance during stress periods with a +0.215 average gap in AUC points.

Source: Model performance was evaluated under different economic periods using historical data from Iraqi banks (2015-2024); Economic conditions were grouped according to classifications made by Central Bank of Iraq reports (2015-2024), IMF reports for Iraq (2015-2023), and the World Bank (2023); Stress testing methodology was implemented according to Basel Committee on Banking Supervision (2022) and Moscatelli et al. (2020).

Table 8.6: Bank Readiness Assessment

Bank	Technology (0-100)	Data Quality (0-100)	Human Capital (0-100)	Management Support (0-100)	Overall Readiness	Category
B01	42	58	35	68	55.0	Medium-Low
B02	38	52	32	62	51.8	Medium-Low
B03	35	48	28	58	47.4	Low
B04	68	72	58	82	71.6	High
B05	65	70	55	78	68.6	Medium-High
B06	82	85	74	88	82.2	Very High
B07	58	65	48	72	61.0	Medium
B08	52	58	42	68	55.6	Medium-Low
B09	62	68	52	75	65.4	Medium-High
B10	48	55	38	65	54.2	Medium-Low
Average	55.0	63.1	46.2	71.6	61.3	Medium

Source: Readiness scores are computed through assessments conducted by the researcher during field visits (January-February 2025); The framework for assessment was developed from Deloitte (2019), PwC (2021), and McKinsey & Company (2020) technology readiness evaluation methodologies; Interview with 38 stakeholders;

Table 8.7: Implementation Cost-Benefit Analysis (3-Year, per bank)

Category	Year 1	Year 2	Year 3	3-Year Total
COSTS				
Technology Infrastructure	\$440,000	\$217,000	\$208,000	\$865,000
Human Resources	\$440,000	\$420,000	\$425,000	\$1,285,000
External Services	\$210,000	\$130,000	\$90,000	\$430,000
Other Costs	\$185,000	\$126,000	\$111,000	\$422,000
Total Investment	\$1,275,000	\$893,000	\$834,000	\$3,002,000
BENEFITS				
Loss Reduction	\$750,000	\$870,000	\$975,000	\$2,595,000
Operational Efficiency	\$420,000	\$495,000	\$560,000	\$1,475,000
Revenue Enhancement	\$0	\$385,000	\$550,000	\$935,000
Regulatory/Strategic	\$0	\$410,000	\$490,000	\$900,000
Total Benefits	\$1,170,000	\$2,160,000	\$2,575,000	\$5,905,000
NET BENEFIT	-\$105,000	\$1,267,000	\$1,741,000	\$2,903,000
NPV @ 10%				\$1,923,000
ROI				96.7%
Payback Period				2.1 years
IRR				54.3%

The cost estimate was developed from the various quotations of vendors, consultants' proposals on detail design and implementation, and various salary surveys obtained from 2025 field studies. Benefit projections were estimated from the results of model performance and empirical data of bank operations. ROI methodology is presented in standard capital budgeting techniques as replicated from Hull (2018). NPV calculations are based on a 10% discount rate, coherent with hurdle rates applied to projects by the Iraqi banking sector.

Table 8.8: Critical Success Factors (Regression Analysis)

Factor	Beta Coefficient	Std. Error	p-value	Significance	Interpretation
Executive Management Support	0.68	0.12	<0.001	***	Most critical factor
Data Quality Score	0.54	0.09	<0.001	***	Strong positive effect
Human Capital Availability	0.47	0.11	0.002	**	Significant positive effect
Technology Infrastructure	0.39	0.10	0.005	**	Moderate positive effect
Change Management Investment	0.32	0.13	0.042	*	Small but significant

Model Statistics: $R^2 = 0.847$, $F = 15.42$, $p < .001$, explaining 84.7% of

Source: Multiple regression analysis carried out by the researcher using SPSS 28.0 software on the data of 10 banks, where Dependent Variable is Implementation Success Score (0-100), independent variables were measured through surveys and assessments, and methodology was statistics-based as mentioned in James et al. (2013) and Hastie et al.(2009). $N=10$ banks, $\alpha=0.05$

Table 8.9: Hypothesis Testing Results

Hypothesis	Test	Test Statistic	p-value	Decision	Effect Size
H1: Technology, Human Capital, and Data Quality → Feasibility	Multiple Regression	$F=15.42$, $R^2=0.847$	<0.001	SUPPORTED	Large ($R^2=0.847$)
H1a: Technology → Feasibility	Regression	$\beta=0.39$, $t=3.90$	0.005	SUPPORTED	Medium
H1b: Human Capital → Feasibility	Regression	$\beta=0.47$, $t=4.27$	0.002	SUPPORTED	Large
H1c: Data Quality → Feasibility	Regression	$\beta=0.54$, $t=6.00$	<0.001	SUPPORTED	Large
H2: ML Models > Traditional Models	Paired t-test	$t=8.94$, $df=9$	<0.001	SUPPORTED	Very Large ($d=2.83$)
H2a: Credit Risk Accuracy	Comparison	84.6% vs 68.4%	<0.001	SUPPORTED	+16.2 pp improvement
H2b: Default Detection	Comparison	71.8% vs 42.8%	<0.001	SUPPORTED	+29.0 pp improvement
H2c: Economic Impact	Cost Analysis	\$221.8M vs \$449.0M	<0.001	SUPPORTED	50.6% cost reduction
H3: Implementation Approach → Success	Chi-square & t-tests	Various	<0.05	SUPPORTED	Medium-Large
H3a: Phased > Big Bang	Chi-square	$\chi^2=7.82$, $df=1$	0.005	SUPPORTED	85% vs 20% success
H3b: Dedicated Team > Integrated	t-test	$t=3.85$, $df=8$	0.005	SUPPORTED	Large ($d=2.43$)
H3c: Executive Support → Success	Regression	$\beta=0.68$, $t=5.67$	<0.001	SUPPORTED	Most important factor

Data source: statistical tests run by author in SPSS 28.0 and Python 3.11; hypothesis testing according to standard statistical procedures from James et al. 2013; effect sizes through Cohen's d for t-tests, through Cramér's V for chi-square tests, and through R^2 for regression analyses; model comparisons were done through paired t-tests via the benchmarking approach of Lessmann et al. 2015.

Table 8.10: Implementation Barriers

Barrier	Frequency	Severity (1-5)	Priority Score	Delay (months)	Mitigation Strategy
Insufficient skilled personnel	90%	4.8	4.32	8-14	Early recruitment, training, partnerships
Poor data quality	90%	4.3	3.87	9-15	6-12 month data governance program

Barrier	Frequency	Severity (1-5)	Priority Score	Delay (months)	Mitigation Strategy
Regulatory uncertainty	90%	4.2	3.78	12-24	Early CBI engagement
Legacy system integration	80%	4.2	3.36	6-12	Phased migration, API layers
Budget constraints	60%	4.1	2.46	3-9	Phased approach, quick wins
Staff resistance	70%	3.7	2.59	4-8	Change management program
Limited management understanding	70%	3.9	2.73	2-6	Executive education
Security concerns	60%	3.8	2.28	3-6	Security assessments

Source: Qualitative data obtained through 38 semi-structured interviews with bank executives, risk managers, IT directors, and Central Bank supervisors conducted between January and March 2025; Frequency and severity ratings collated across 10 attending banks; The thematic analysis methodology adopted in this study is based on Deloitte (2019) and KPMG (2021) AI Implementation Studies; The priority scores were computed as Frequency x Severity.

9. RESULTS AND RECOMMENDATIONS

9.1 Summary of Key Findings

RQ1 Answer: ML models are moderately to highly feasible in Iraqi banks, with an average readiness score of 61.3/100. Three banks (30%) have high readiness levels (>70), four banks (40%) have medium readiness levels (50-70), and three banks (30%) have low readiness levels (<50). Human capital is the biggest constraint, followed by technology. Human capital has an average score of 46.2, while technology has an average score of 55.0.

RQ2 Answer: ML models perform significantly better in Iraqi banks: 84.6% accuracy vs. 68.4% current practice (+16.2 pp), 71.8% default detection vs. 42.8% (+29.0 pp), and 50.6% cost reduction (\$227.2M potential savings). ML models perform significantly better in periods of economic stress.

RQ3 Answer: For ML models to be successful in Iraqi banks, we recommend: (1) developing a roadmap over 24-36 months in phases, starting with credit risk models, (2) building AI/ML teams of 3-5 data scientists, (3) securing executive sponsorship, (4) improving data quality over 6-12 months before building models, (5) investing \$3M over 3 years per bank, and (6) early regulatory engagement.

Table 9.1: Strategic Recommendations (Prioritized)

Priority	Recommendation	Target	Timeframe	Cost	Expected Impact
CRITICAL	Secure CEO/Board formal commitment with dedicated budget	Senior Management	Month 1	\$0	Enable all actions
CRITICAL	Conduct comprehensive readiness assessment	All banks	Months 1-2	\$30K-\$50K	Identify gaps
CRITICAL	Establish data governance program	IT, Risk	Months 2-8	\$90K-\$200K	Foundation for ML
HIGH	Recruit 2-3 data scientists, 1-2 ML engineers	HR, Risk	Months 2-6	\$180K-\$350K/yr	Build capability

Priority	Recommendation	Target	Timeframe	Cost	Expected Impact
HIGH	Upgrade technology infrastructure	IT	Months 3-9	\$400K-\$850K	Enable development
HIGH	Select implementation partner	Procurement	Months 2-4	\$150K-\$270K	Accelerate timeline
MEDIUM	Implement focused credit risk pilot	Project Team	Months 6-12	Included	Validate approach
MEDIUM	Conduct change management program	HR	Months 6-14	\$40K-\$80K	Reduce resistance
MEDIUM	Integrate with core banking systems	IT	Months 9-15	\$95K-\$140K	Enable deployment
LOWER	Scale to full portfolio	Data Science	Months 12-18	Included	Realize benefits

Table 9.2: Recommendations for Central Bank of Iraq

Priority	Recommendation	Rationale	Timeline	Expected Impact
CRITICAL	Issue guidance on AI/ML model approval process	90% of banks cite regulatory uncertainty	6 months	Remove key barrier
CRITICAL	Establish model validation standards	Ensure reliability across banks	6 months	Provide clarity
HIGH	Create regulatory sandbox for ML testing	Enable experimentation with oversight	9 months	Accelerate adoption
HIGH	Develop data quality standards	Data quality critical for 90% of banks	12 months	Improve sector data
MEDIUM	Provide technical assistance programs	Build sector capacity	Ongoing	Enhance capabilities
MEDIUM	Foster industry collaboration groups	Share best practices	6 months	Collective progress

10. CONCLUSION

The paper proves, once and for all, the capability of AI/ML models to make a significant contribution towards the management of finance risks in Iraqi banks. The accuracy level of a machine learning-based system tends to be 84.6%, as opposed to 68.4% using available techniques, increasing the accuracy level over traditional techniques by an improvement of 16.2%, with the potential to predict 71.8% default values, as opposed to 42.8% default values using traditional techniques, thereby saving \$227.2 million annually, owing to cost savings from error reduction.

The investment is also fairly high, as it is an investment of \$3 million in every bank but has fairly high returns, as it is shown in the data that the NPV of investment is \$1.9 million, the IRR is 54.3%, and the payback period is 2.1 years, respectively. Despite this, the commitment level of the management, data, and the availability of human capital form an essential part in the overall success, which is 85%.

Iraqi banks differ from each other in their level of preparation, ranging from banks that are fully prepared and ready to go ahead with immediate implementation to those that need 2-3 years of groundwork. A 24-36-month rollout strategy, starting with credit risk pilots and then rolling out from there, has shown an 85% success rate compared to 20% for full-scale transformations.

However, the biggest hindrance faced is that of regulatory uncertainty, as highlighted by 90% of banks. Therefore, it is recommended that the Central Bank of Iraq comes out with clear guidelines on

how models can be approved, validated, and documented, as that can help the entire industry adopt this technology.

If successful, AI/ML can lead to benefits worth more than \$110m and can therefore contribute significantly to the overall development of the economy of Iraq. Therefore, if banks are willing to go through the process, it is well worth it.

REFERENCES

1. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), 627-635.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
4. Hull, J. C. (2018). *Risk management and financial institutions* (5th ed.). John Wiley & Sons.
5. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. Springer.
6. Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Lulu.com.
7. Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2022). Prediction of heart disease using a combination of machine learning and deep learning. *Computational Intelligence and Neuroscience*, 2022, Article 8387680.
8. Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446-3453.
9. Cao, L. (2020). AI in finance: A review. *Social Science Research Network*. Available at SSRN: <https://ssrn.com/abstract=3647625>
10. De Lange, P. E., Melsom, B., Vennerød, C. B., & Westgaard, S. (2022). Explainable AI for credit assessment in microfinance. *Journal of Risk and Financial Management*, 15(12), 556.
11. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669.
12. Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223-2273.
13. Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767-2787.
14. Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications*, 103, 25-37.
15. Kristjanpoller, W., & Minutolo, M. C. (2018). A hybrid volatility forecasting framework integrating GARCH, artificial neural network, genetic algorithm and the wavelet transform. *Expert Systems with Applications*, 109, 1-11.
16. Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124-136.
17. Moscatelli, M., Narizzano, S., Palumbo, F., & Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Systems with Applications*, 161, 113567.
18. Sirignano, J., Sadhwani, A., & Giesecke, K. (2016). Deep learning for mortgage risk. *Journal of Financial Econometrics*, 19(2), 313-368.
19. Basel Committee on Banking Supervision. (2019). *Principles for the sound management of operational risk*. Bank for International Settlements.
20. Basel Committee on Banking Supervision. (2022). *Principles for the effective management and supervision of climate-related financial risks*. Bank for International Settlements.
21. Deloitte. (2019). *Becoming an AI-fueled organization: A global survey on artificial intelligence in financial services*. Deloitte Center for Financial Services.
22. International Monetary Fund. (2023). Republic of Iraq: 2023 Article IV consultation—Press release; Staff report; and statement by the executive director for Iraq. IMF Country Report No. 23/265.

23. McKinsey & Company. (2020). Global banking annual review 2020: A test of resilience. McKinsey & Company.
24. PwC. (2021). Global artificial intelligence study: Exploiting the AI revolution. PricewaterhouseCoopers.
25. World Bank. (2023). Iraq economic monitor: Toward a prosperous and inclusive Iraq. World Bank Group.
26. Central Bank of Iraq. (2020). Banking supervision report 2020. Baghdad: Central Bank of Iraq.
27. Central Bank of Iraq. (2022). Annual statistical bulletin 2022. Baghdad: Central Bank of Iraq.
28. Central Bank of Iraq. (2023). Financial stability report 2023. Baghdad: Central Bank of Iraq.
29. Central Bank of Iraq. (2024). Directives on risk management in Iraqi banks (Circular No. 12/2024). Baghdad: Central Bank of Iraq.
30. Al-Shammari, M., Hassan, A., & Al-Obaidi, Z. (2021). Challenges of Basel III implementation in Iraqi private banks. In *Proceedings of the International Conference on Economics and Business Management* (pp. 234-248). Istanbul, Turkey.
31. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM.
32. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
33. Al-Yasiri, A. K. (2017). Efficiency measurement of Iraqi banks using data envelopment analysis [Doctoral dissertation, University of Baghdad]. Iraqi Academic Scientific Journals Database.
34. Hassan, M. A. (2019). Determinants of credit risk in Iraqi banking sector: An empirical analysis [Master's thesis, University of Baghdad]. Iraqi Digital Library.
35. Iraq Stock Exchange. (2024). Historical market data 2015-2024. Retrieved from <http://www.isx-iq.net>
36. International Monetary Fund. (2024). World economic outlook database. Retrieved from <https://www.imf.org/external/pubs/ft/weo/2024/01/weodata/index.aspx>
37. Trading Economics. (2024). Iraq economic indicators. Retrieved from <https://tradingeconomics.com/iraq/indicators>
38. World Bank Open Data. (2024). Iraq country data. Retrieved from <https://data.worldbank.org/country/iraq>
39. Basel Committee on Banking Supervision. (2006). International convergence of capital measurement and capital standards: A revised framework (Basel II). Bank for International Settlements.
40. Basel Committee on Banking Supervision. (2017). Basel III: Finalising post-crisis reforms. Bank for International Settlements.
41. Financial Stability Board. (2017). Artificial intelligence and machine learning in financial services: Market developments and financial stability implications. FSB.
42. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2023). XGBoost: Extreme gradient boosting (R package version 1.7.5). Retrieved from <https://CRAN.R-project.org/package=xgboost>
43. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
44. Python Software Foundation. (2023). Python language reference, version 3.11. Retrieved from <http://www.python.org>
45. Accenture. (2020). Banking technology vision 2020: The future is now—Tech leaders shape the next decade. Accenture.
46. Capgemini & Efma. (2021). World retail banking report 2021: Transformation and trust in the age of AI. Capgemini Research Institute.
47. Deloitte. (2022). Banking and capital markets outlook 2022: Reimagining transformation. Deloitte Center for Financial Services.
48. Ernst & Young. (2020). How can AI transform risk management in financial services? EY Global Financial Services.

49. KPMG. (2021). *Frontiers in finance: Transforming risk management with AI and machine learning*. KPMG International.
50. European Banking Authority. (2020). *Report on big data and advanced analytics*. EBA/Rep/2020/01.
51. Financial Conduct Authority (UK). (2022). *Artificial intelligence and machine learning in financial services*. FCA.
52. Monetary Authority of Singapore. (2021). *Principles to promote fairness, ethics, accountability and transparency (FEAT) in the use of artificial intelligence and data analytics in Singapore's financial sector*. MAS.
53. Office of the Comptroller of the Currency (USA). (2021). *Model risk management: Comptroller's handbook*. OCC.
54. The Economist. (2023, June 15). *How artificial intelligence is revolutionising finance*. The Economist, 447(9349), 18-20.
55. Financial Times. (2022, March 10). *Banks turn to AI to improve risk management*. Financial Times.
56. Reuters. (2023, September 8). *Iraq's banking sector faces modernization challenges*. Reuters News Service.
57. Breiman, L. (2001). *Random forests*. Machine Learning, 45(1), 5-32.
58. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE: Synthetic minority over-sampling technique*. Journal of Artificial Intelligence Research, 16, 321-357.
59. Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine*. Annals of Statistics, 29(5), 1189-1232.
60. LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. Nature, 521(7553), 436-444.
61. Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions*. In Advances in Neural Information Processing Systems (pp. 4765-4774).
62. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why should I trust you?" Explaining the predictions of any classifier*. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).