

ArthaX An Explainable AI System for Loan and Insurance

Sumit Roy¹, Ansh Thakur², Ajay Gaur³, Avani Srivastava⁴

¹Sumit Roy Computer Engineering & Pillai College of Engineering

²Ansh Thakur Computer Engineering & Pillai College of Engineering

³Ajay Gaur Computer Engineering & Pillai College of Engineering

⁴Avani Srivastava Computer Engineering & Pillai College of Engineering

Guide: Prof. Leenata Parab

Abstract—Machine learning is increasingly being used by financial institutions for approving loans, forecasting credit-risk, for underwriting insurance, and in fraud detection, among other applications, but the machine learning systems the institutions use are often opaque to the people whose lives they touch and affect: the loan applicant, the loan auditor and the financial supervisor. A consolidated survey of twenty recent studies on how to implement Explainable Artificial Intelligence (XAI) in financial decision making is presented, grouped into four research themes: loan approval and financial/credit risk assessment, insurance and personalization, financial and insurance fraud detection, and know your customer (KYC) identity verification and financial-document intelligence. The survey is able to benchmark the algorithms including machine learning models, explainability techniques, data sets, and the results reported over the literature, but does not suggest a "one fits all" approach to comparing performance across different studies. Even though every pipeline stage - identity verification, document processing, fraud screening, predictive decisioning, post-hoc explanation - is well represented on its own, nothing from the literature has been found of an integrated system that underpins all these stages in a user-centric workflow for the benefit of the applicant. Given this disintegration, the paper proposes a data-consent based all-in-one explainable artificial intelligence system for loans and insurance decision making, which consists of a combination of biometric KYC verification, OCR-based document intelligence, anomaly detection based on fraud screening, gradient boosting based on risk prediction, explanation based on SHAP and a data access layer behind a consent. ArthaX places itself throughout as a research direction, not an experimentally proven system, and the survey concludes with identifying the fairness, explanation-stability, dataset-imbalance, regulatory and privacy challenges that remain open for trustworthy and end-to-end explainable financial decision systems.

Index Terms—Credit risk assessment, explainable artificial intelligence (XAI), financial document intelligence, fraud detection, insurance underwriting, know-your-customer (KYC) verification, LIME, loan approval, SHAP, trustworthy AI.

I. INTRODUCTION

Machine learning has been at the helm of many of the critical decisions that financial institutions make automatically, including loan approval, credit-risk scoring, insurance underwriting and fraud detection. The reasoning behind these models is not always obvious to those who are the subject of their decisions, such as credit or insurance applicants, as regulators demand to audit decisions and justify them, or as auditors want to confirm actions are non-discriminatory, and for those whose decisions are denying or granting an insurance policy or deal. In the past few years, there was a considerable amount of research focused on Explainability in AI, and specifically post-hoc model agnostic methods like SHapley Additive explanations (SHAP) and LIME (Local Interpretable Model-agnostic Explanations).

The goal of this survey is to advance the implementation of ArthaX, a proposed explainable AI system for integrated loan and insurance decision making. The ArthaX architecture consists of a few integrations: FaceNet style computer-vision KYC check, OCR-based document intelligence, Isolation etc. Forest-based fraud screening, XGBoost fraud prediction, SHAP-based customer-facing fraud explanation, personalized recommendation layer and consent-based framework on receiving financial data from an Account Aggregator (AA). Again, ArthaX is not introduced with the intent of making any performance claims about ArthaX itself, because it has not yet been experimentally evaluated; it is merely presented here as motivation for the survey and as a lens to identify research gaps.

The current paper's value lies in a critical and synoptic evaluation of twenty studies related to the loan approval process, credit-risk assessment, the insurance

underwriting process, the insurance and financial-forgery detection process, e-KYC (Know Your Customer) identity verification and financial document processing. How are ML/DL models used, and what are the possible applications of XAI techniques and how so, specifically: (1) which ML/DL models are widely utilized; (2) which XAI techniques are utilized and how; (3) which datasets are available and what evaluation protocols are used for these studies; and (4) The efficacy of the reported outcome – and how comparable it is across various studies; (5) the recurring limitations reported in the literature; (6) how loan, insurance, fraud, KYC and document processing issues are handled individually; and (7) how an integrated architecture like ArthaX could potentially bridge those gaps.

The rest of the paper is laid out as follows. Section II will provide an overview of the background concepts of financial ML and XAI. The methodology of the survey is described in section III. The literature reviewed is classified in Section IV. These five sections, V-VIII, are literature reviews/operationalization of the concepts of loan/credit-risk, insurance, fraud and KYC/documentation intelligence risk, respectively. A comparative analysis is provided in Section IX with tables of summarizing data. Section X can be used to document gaps in research. ArthaX is introduced as a proposed integrated architecture in Section XI. The remaining sections (XII-XIV) are on the limitations, future scope and cross cutting themes followed by a Conclusion (Section XV).

II. BACKGROUND AND FUNDAMENTALS

A. Machine Learning Models in Financial Decisioning

The literature review has resulted in a variety of existing models being used to financial decisioning. Therefore, while linear and rule-based models like Logistic Regression and Belief Rule Base systems are inherently interpretable, the models found in the loan-approval and credit-risk literature are—Random Forest, Gradient Boosting, XGBoost and LightGBM—models based on trees and ensembles of trees. Facial identity verification (KYC) is supported by convolutional and super-resolution networks and the most commonly used approach for fraud screening is called isolation forest. Some more recent approaches focus on agentic and multi-agent large language models (LLMs) to facilitate autonomous decision making with an ability to explain reasoning [9],[10].

B. Explainable AI Concepts

There are broadly speaking two kinds of XAI techniques: intrinsic - which are models specifically

designed to be interpretable by design, like decision trees or Belief Rule Base systems - and post-hoc, which develop explanations after the model is trained and deployed. There are also post-hoc techniques that explain overall model behaviour (global) and also local/explain an individual prediction (local). Against the background of cooperative game theory, SHAP offers explanations in the global as well as in the local dimension, and is the most common method in the literature that was reviewed. LIME explains predictions of the complex model in terms of the interpretable surrogate model in the neighborhood of each prediction. Other methods seen are the Partial Dependence Plots (PDP), Accumulated Local Effects (ALE), counterfactual explanations, rule extraction, natural-language explanations, and confidence estimation approaches like Monte Carlo Dropout.

C. Overview of the ArthaX Framework

ArthaX vision is to provide an end-to-end solution that will first use a FaceNet-based KYC module to verify an applicant's identity, followed by an OCR and a document-intelligence module to process supporting documents; then an Isolation Forest-based module to screen for fraud indicators; then an XGBoost model to predict loan or insurance approval, and risk; then the generation of explanations to provide to the applicant; then a Financial Product Recommendation algorithm that proposes safer and more eligible financial products; and finally an Account Aggregator framework to govern secure access to the financial data of an applicant throughout the process. This architecture is also utilized as the lens to highlight areas of existing research where and where not there is a direct connection. (Section IX - XI)

III. SURVEY METHODOLOGY

The present survey compiles and summarizes 20 papers and reports in recent years (2023-2026) on explainable financial decision-making, including the following applications: loan approval, credit-risk assessment, insurance underwriting, fraud detection, e-KYC, and financial-document intelligence. The literature was gathered as scoping for the ArthaX project. Due to the lack of a formal set of bibliographic databases, search strings, and a pre-registered inclusion/exclusion protocol, this survey can be best characterized as a structured, category-based scoping and comparative review as opposed to a systematic literature review (SLR) in the PRISMA sense. The categorization in Section IV and the comparative tables in Section IX should not be viewed as a comprehensive list of the literature, but rather as an analysis and summarization of what was found.

All twenty of the studies spoken of in this survey are presented with complete bibliographic information, models, datasets, and reported results, with the sole exception of those studies that only reported results in qualitative terms (e.g., "the best performing" without specifying a numerical figure).

IV. CLASSIFICATION OF EXISTING RESEARCH

The literature considered is divided into four groups according to the sections of the proposed ArthaX architecture: (1) loan approval and credit risk, (2) insurance underwriting and personalization (3) financial and insurance fraud detection, (4) KYC, identity verification, financial-document intelligence. A paper can be in multiple categories, and is discussed in the main category and referenced in the other categories. Of the twenty papers, eleven focus on loan approval and credit risk, two on insurance, three on fraud detection (and one paper on cybersecurity that is domain focused but for the methodological relevance for explanation reliability in this context), and one is a stakeholder-oriented paper that crossed sections of all domains related to e-KYC and document intelligence.

V. EXPLAINABLE AI FOR LOAN AND CREDIT-RISK DECISIONS

The biggest body of literature on which we set out reflects the interest in channeling loan applications and in forecasting credit risks with tree-based ensembles and SHAP and/or LIME. AkteR et al. [2] employ SHAP not just for post-hoc explanation, but as a feature selection mechanism before modelling on a Kaggle credit card loan-approval dataset – their Logistic Regression model accuracy increases to 99.06% once they apply LIME to provide local explanations for each person in the dataset. Most of the essential parts of SHAP (global) and LIME (local) were merged according to the idea of Li et al. [3].

On the massive LendingClub dataset (about 2.93 million loans), LightGBM showed the most promising performance among seven different reported models, with an AUC of 0.94 and an accuracy of 87%, and provided explanations on the loan record level, including loan amount, interest rate, and debt-to-income ratio, for default phenomena at the applicant level.

The second one focuses on systematic comparison between classical models and SHAP, LIME and PDP. Ravi et al. [4] provides a comparative study of the five classifiers proposed using loan-prediction data provided by the Kaggle website with a classification accuracy of 99% for Random Forest, and point out some drawbacks:

overfitting and infrastructure. In another recent analysis of XAI in credit-risk [5] the same classical models are examined as successful models and SHAP and LIME are used in the analysis, achieving good accuracy but explicitly stating the need for larger and more real-world banks data sets. Nag et al. [6] explain how to interpret the results of deep learning methods, such as the Decision Tree and Random Forest, with a focus on regulatory trust, along with visualizing the results using the SHAP global and LIME local approaches to understanding.

Kakkar [7] advocates a more regulations-driven approach, suggesting that the top-performing model amongst five classifiers does not necessarily translate to the most deployable in a Basel III / GDPR framework of banking regulations, and sees SHAP as a linkage between complex ensembles and regulatory requirements for transparency. Madhu et al. [8] came up with a Belief Rule Base (BRB) framework for loan underwriting, which is more interpretable than Random Forest and Decision Tree models, achieving 95% accuracy for loan approval and 90% accuracy for explaining loan rejection reasoning in terms of the rules that are activated.

In addition to classical supervised learning, two studies on agent-based architecture are presented. Kubam [9] introduces a multi-agent, reinforcement-learning approach for making autonomous real-time credit-risk decisions with 94.2% accuracy and close to zero latency while continually learning; however, reliance on high-quality streaming data, and vulnerability to model drift, respectively, cause an issue for these types of systems. The multi-agent LLM system (with both role-specialized agents and FinBERT/MiniLM) proposed by Shi et al. [10] explains corporate credit ratings (not consumer) on a Kaggle corporate-credit dataset with 72.6% accuracy and 72.7% F1-score using traceable, automated reports instead of SHAP/LIME outputs, without the use of KYC, OCR or fraud screening.

Two additional studies support this group. Mujo et al. [19] find a clear accuracy ordering of the models for transparency for the task of loan-decision, namely Logistic Regression (accuracy 72.40%), Decision Tree (75.10%), Random Forest (81.60%), XGBoost (84.20%) and a Deep Neural Network (86.70%) which establishes that the higher the accuracy, the greater the computational cost, and the harder it is to sustain the explainability in real time. Acharya et al. [20] study the connections between the aspects of explainability, fairness, and predictive power for credit and real-estate machine-learning models (over 80% Accurate) and find that their key attribute—computational cost of Shapley-value

calculations at scale—echoes that of some of the studies that have been mentioned so far.

VI. EXPLAINABLE AI FOR INSURANCE UNDERWRITING

Work on insurance is generally less in quantity but deals with slightly different issues: personalisation, static/dynamic policy type and data quality issues identification by means of explanation. Gummadi [11] examines some of the most recent developments in gradient-boosted trees, Bayesian networks, constrained neural networks, and fuzzy logic as means to enable underwriting automation with explanation and a comprehensive toolkit of explainability tools: SHAP, LIME, Accumulated Local Effects, counterfactuals, rule extraction, and natural-language explanations. The improvements are significant, including reduced processing time and better policy structures, and crucially, the ability to generate explainability that reveals data-quality problems and explain predictions – but also that the major issues are a lack of an openly available benchmark dataset, and that static policy structures are still used for annual policies.

Kapse et al. [12] investigate the personalization of health insurance premiums using some of the classical models, Gradient Boosting model is the best model, local interpretation and global interpretation with SHAP. One of the major limitations of the study is the few numbers of the respondents who participated in the survey ($n = 145$), which limits the validity and applicability of the study findings.

VII. EXPLAINABLE AI FOR FINANCIAL AND INSURANCE FRAUD DETECTION

The reviewed set contains the studies related to the fraud-detection research from conventional anomaly detection, deep learning, to systematic review. Another set of authors, Tripathi et al. [14] employ an unsupervised approach to anomaly detection on the Kaggle credit-card-fraud dataset (284,807 transactions) and achieve 99.67% accuracy but also do not provide any layer of explainability and do not test their solution in real-time or with multiple domains of fraud. The three-stage convex-to-non-convex neural network with an improved focal-loss function for auto-insurance fraud detection and performed feature-importance analysis with SHAP is proposed by Boabang and Gyamerah [13] with qualitative results reporting the best balance of accuracy, precision, recall, F1-score, and AUC among the compared configurations, though the dataset is proprietary and domain-specific.

A systematic review of 108 peer-reviewed publications (from 2019 to 2024) on deep learning financial fraud detection, Chen et al. [15] evaluate the deep learning variants used to integrate four prominent deep learning techniques—CNN, LSTM, transformer, GNNs, and GANs—into deep-learning financial fraud detection and propose a deep-learning-sector-governance framework to align the type of deep learning with sector-specific regulatory requirements. The review identifies that while credit-card and banking fraud are comparatively well studied, the same cannot be said for loan and mortgage fraud, and that the concept of explainability is not consistently, in the same terms, mentioned in the studies reviewed.

A related, cybersecurity domain study is provided for its methodological relevance with explanation reliability. Fernández-Morales et al. [16] use Decision Tree, Random Forest, LSTM and two types of MLP with SHAP and LIME for IoT intrusion.

Random Forest had the best accuracy of 98.96% and F1-score on the NF-UQ-NIDS-v2 (over 76 million records). What makes it special is that it also provides, for the first time, an analysis of the stability of the prediction's explanation, by measuring the coefficient of variation of the SHAP values that explain it from one prediction to the next, showing that any financial system (including ArthaX) wanting to offer explanations for applicants predictions will find this measure useful.

VIII. KYC, IDENTITY VERIFICATION, AND FINANCIAL DOCUMENT INTELLIGENCE

Ketcham et al. [17] have analysed the performance of CNN, SRCNN, VDSR and DRRN architectures on the task of facial identity verification for electronic KYC, alongside with MobileNetV2, ResNet50 and FaceNet, and replaced SHAP/LIME with Monte Carlo Dropout to perform uncertainty-aware and confidence-based prediction. The optimal configuration of CASIA-WebFace is shown to reach 99.7% accuracy, 99.2% precision, 99.3% recall and 99.5% AUC, and it is found that the preprocessing of super-resolution can help improve recognition accuracy for degraded images; although the held-out test set is small (70 subjects) and the problem may have demographic-bias, the primary disadvantage is the limited size of the held-out test set.

For financial-document intelligence, Jin et al. [18] directly tackles the problem by jointly utilizing PaddleOCR and PP-LCNet for the text detection/recognition task and a compact vision-language

model (MiniCPM-o-2.6-8B) for multi-stage field extraction from 92 confidential financial reports with a larger Qwen2.5-VL-72B model. The system processes around 1522 documents per hour on an A100 GPU and returns the pages/regions of the document to the original source (as opposed to a SHAP/LIME-style attribution). Other limitations of the dataset are the confidentiality and ambiguity of the terms and age of the data, and the non-entered fields do not enter a decision for downstream approval.

IX. COMPARATIVE ANALYSIS

Table I summarizes the domain, techniques, best reported result, and key limitation of each of the twenty reviewed studies. Table II compares the principal categories of ML/DL models observed across the literature in terms of typical application, strengths, limitations, and compatibility with explainability techniques.

TABLE I

Comparative Summary of Reviewed Studies

Ref.	Year / Domain	Techniques Used	Best Reported Result	Key Limitation
[1]	2025 - Stakeholder review of XAI in finance	N/A (conceptual/industry report); discusses SHAP, LIME, counterfactuals, rule-based, PDP/ICE, evaluative & neurosymbolic AI	N/A - conceptual analysis, no ML experiment	No original model or dataset; qualitative synthesis only
[2]	2026 - Loan approval	SHAP-guided feature selection + LIME (8 classifiers)	Logistic Regression: 99.06% acc., 98.87% prec., 99.62% rec., 99.24% F1	Single public dataset; generalization across institutions not verified
[3]	2025 - Loan default (LendingClub, ~2.93M records)	LightGBM + SHAP (global) + LIME (local) + Recursive Feature Elimination	LightGBM: 0.94 AUC, 87% acc., 0.88 F1	Residual class imbalance in a 13-year, real-world dataset
[4]	2025 - Credit scoring / loan approval	RF, LR, SVM, DT, KNN + SHAP, LIME, PDP	Random Forest: 99% acc., 98% prec., 97% rec., 98% F1	Overfitting risk with RF; infrastructure and expertise requirements
[5]	2024 - Credit risk analysis	Scikit-learn classifiers + SHAP + LIME	High classification accuracy (exact figures not disclosed in source)	Needs larger, real-world banking datasets for validation
[6]	2024 - Credit risk & decision support	Decision Tree and Random Forest + SHAP/LIME (force, summary, dependency, decision plots)	Random Forest: 94% acc.; Decision Tree: 90% acc.	Evaluated on public data only; limited real-time applicability
[7]	2025 - Credit risk scoring (regulatory)	Comparison of 5 ML models + SHAP	Gradient Boosting highest accuracy (exact value not disclosed)	Highest-accuracy model may not be the most deployable under Basel III/GDPR

[8]	2025 - Loan underwriting (Belief Rule Base)	Belief Rule Base + Random Forest/Decision Tree + SHAP	95% acc. (approval); 90% acc. (rejection-reasoning)	Dataset not publicly disclosed; rule engineering overhead
[9]	2024 - Autonomous, real-time credit risk	Multi-agent system + reinforcement learning + dynamic risk thresholds	94.2% acc., 91.5% precision, 92.3% recall, near-zero latency	Needs high-quality streaming data; model drift and regulatory uncertainty
[10]	2025 - Corporate credit rating (5,403 firm-years)	Multi-agent LLM system (BRA/FRA/GRA/CRA/CAA) + FinBERT/MiniLM	72.6% acc., 72.7% F1 (no SHAP/LIME; automated reports instead)	Corporate, not consumer, loans; no KYC/OCR/fraud/insurance integration
[11]	2025 - Insurance underwriting automation	GBDT, Bayesian Networks, Constrained NN, Fuzzy Logic + SHAP, LIME, PDP, ALE, counterfactuals, NL explanations	Qualitative gains in speed and personalization; no numeric accuracy/F1/AUC disclosed	No benchmark dataset disclosed; limited to static annual policies
[12]	2025 - Health insurance premium prediction	MLR, Decision Tree, RF, Gradient Boosting, XGBoost + SHAP	Gradient Boosting: highest RZ among evaluated models	Very small, self-collected sample (145 responses)
[13]	2025 - Auto-insurance fraud detection	3-stage convex→non-convex neural network with enhanced focal loss + SHAP	Best balance of accuracy/precision/recall/F1/AUC among compared configurations (exact values not disclosed)	Proprietary dataset; qualitative-only SHAP reporting; auto-insurance domain only
[14]	2023 - Credit-card fraud detection	Isolation Forest (unsupervised anomaly detection)	99.67% accuracy on Kaggle credit-card fraud dataset (284,807 transactions)	No explainability layer; single fraud type; no real-time/loan/insurance integration
[15]	2025 - Financial fraud detection (systematic review, 108 studies)	CNN, LSTM, RNN, Transformer, GNN, GAN, RF, XGBoost (surveyed); SHAP/LIME discussed broadly	N/A - review paper proposing a deep-learning-sector-governance framework	No original experiment; loan/mortgage fraud comparatively understudied
[16]	2025 - IoT intrusion detection (smart rural)	Decision Tree, Random Forest, LSTM, two MLP variants + SHAP + LIME + explanation-stability analysis	Random Forest: 98.96% acc./F1 on NF-UQ-NIDS-v2 (76M+ records)	Cybersecurity, not financial, domain; not human-centered XAI

[17]	2025 - e-KYC / facial identity verification	CNN, SRCNN, VDSR, DRRN, Monte Carlo Dropout (vs. MobileNetV2, ResNet50, FaceNet)	99.7% acc., 99.2% prec., 99.3% rec., 99.5% AUC on CASIA-WebFace	Small test subset (70 subjects); demographic-bias risk; degrades under extreme distortion
[18]	2025 - Financial-document field extraction (OCR)	PaddleOCR, PP-LCNet, BM25, MiniCPM-o-2.6-8B (vs. Qwen2.5-VL-72B)	80.87% field accuracy; 1,522 documents/hour on an A100 GPU	Small, confidential dataset; terminology/currency ambiguity; no downstream decision layer
[19]	2025 - Loan-decision transparency	Logistic Regression, Decision Tree, Random Forest, XGBoost, Deep Neural Network + SHAP	DNN best at 86.70% acc. (LR 72.40%, DT 75.10%, RF 81.60%, XGBoost 84.20%)	Rising computational requirements make real-time explainability difficult
[20]	2024 - Fairness in financial and real-estate ML	XGBoost + SHAP + Scikit-learn	Over 80% accuracy; income and credit history identified as leading decision drivers	Shapley-value computation is expensive for very large or real-time datasets

TABLE II

Comparison of Machine-Learning and Deep-Learning Model Categories

Model Category	Typical Application	Strengths	Limitations	XAI Compatibility
Logistic Regression	Loan approval, credit scoring [2], [4],[6]	Intrinsically interpretable; efficient; regulator-friendly	Limited capacity for complex, non-linear interactions	High - inherently interpretable, also usable with SHAP
Decision Tree	Loan approval, credit risk [4],[6], [8],[16]	Human-readable rule paths; fast to train	Prone to overfitting; unstable with small data changes	High - path-based intrinsic explanation
Random Forest	Loan approval, credit risk, e-KYC baseline [4]–[6],[8],[16]	Strong accuracy; robust to overfitting vs. single trees	Reduced transparency vs. single trees; heavier compute	Moderate–High - SHAP TreeExplainer widely used
Gradient Boosting / XGBoost / LightGBM	Loan default prediction, credit scoring, premium prediction [3],[7], [8],[12],[19],[20]	State-of-the-art tabular performance; handles non-linearities	Less transparent than linear/rule models; hyperparameter-sensitive	Moderate–High - SHAP TreeExplainer, feature importance
SVM / KNN / Naive Bayes	Comparative baselines [3],[4]	Effective on smaller, well-structured datasets	Poor scalability; limited native interpretability	Low–Moderate - requires post-hoc explainers such as LIME

Isolation Forest	Credit-card and insurance fraud detection [13],[14]	Effective unsupervised anomaly detection; no labeled fraud needed	No native explainability; single-domain focus in reviewed studies	Low - rarely paired with XAI in the reviewed literature
Belief Rule Base (BRB)	Loan approval with intrinsic transparency [8]	Inherently interpretable; encodes expert knowledge with ML	Requires rule engineering; less flexible than data-driven ensembles	High - intrinsically interpretable
CNN / Super-Resolution Networks	Facial identity verification for e-KYC [17]	High accuracy on degraded/low-resolution images	Demographic-bias risk; degrades under extreme distortion	Low via SHAP/LIME; addressed instead via uncertainty estimation
Deep Learning (CNN/LSTM/Transformer/GNN/GAN)	Financial fraud detection [13],[15]; intrusion detection [16]	Captures complex temporal/relational patterns; strong on large data	High compute cost; limited native interpretability; data-hungry	Low-Moderate - requires post-hoc SHAP/LIME
Multi-Agent / LLM-based Agents	Autonomous credit decisions, corporate credit rating [9],[10]	Modular reasoning; fuses structured and unstructured data	Emerging paradigm; limited benchmarking against SHAP/LIME	Moderate - achieved via agent reasoning/report generation

A. Performance Analysis

The reported accuracy levels of the loan/credit-risk studies reviewed range from about 72% [10],[19] to more than 99% [2],[4], [17], while fraud-detection accuracy reaches 98.96%–99.67% [14], [16]. These numbers are only to be compared with great caution: the samples investigated differ (145 self-report answers [12] versus > 76 million network-traffic records [16]), the class distributions are different, and so are the methods of handling imbalance. A model with 99% accuracy over a small, relatively clean, public Kaggle dataset [4] is not comparable to one which reports an accuracy of 87% over a large, real-world thirteen-year loan-origination dataset [3] or a corporate-credit model with 72.6% accuracy on a much smaller target much harder to predict [10]. Therefore, the best result reported for each paper is for its own experimental setting and is not necessarily the most accurate, but rather the most significant in that study, compared with simpler baselines; the results of the different studies can only be compared qualitatively, for example, that gradient-boosting and Random Forest models have been consistently outperforming simpler models in a specific study.

B. SHAP, LIME, and the Explainability–Trust Distinction

From the literature reviewed, the most consistent use of XAI across all applications is for global feature importance ranking, local applicant-level explanation, and - in at least one instance - as a feature-selection mechanism before model training, with SHAP being the most frequently applied method. It is based on Shapley values, which provides it with a desirable consistency property in regulatory situations [7]. LIME is most commonly combined with SHAP, to provide fast, locally faithful surrogates to individual predictions [3],[6],[8]. There are studies that go beyond SHAP/LIME to richer explanation types, such as counterfactual explanations, which offer "what would need to change" information in an applicant-facing way [11]; Monte Carlo Dropout, which provides information about the model's confidence, not just its reason, [17]; and agent-based reasoning, which gives structured, human-readable explanations, rather than a numerical attribution [9],[10].

One common theme is the difference between the predictive performance of a model and trust in the model when making decisions: a highly accurate model might not be suitable for deployment because an applicant may

not be able to understand a denial, or because an auditor may not be able to verify the non-discrimination of the model, or because a regulator may not be able to certify compliance with regulatory frameworks, such as Basel III or GDPR [1],[7]. This is supported by [16] which demonstrates that explanations may be reliable in some ways but not others, that is, the explanation technique may be present, but may be inconsistent across predictions, which translates to a lack of trust. This indicates that explanation stability and applicant comprehensibility (i.e., ease of understanding) of a system like ArthaX – which explains their proposals to those who are directly applying for loans and insurance – should be considered first-class design features.

X. RESEARCH GAPS

Synthesizing the twenty reviewed papers reveals a set of recurring, evidence-supported gaps directly relevant to the design of an integrated system such as ArthaX.

- Fragmentation of financial AI solutions: each reviewed paper addresses at most one or two pipeline stages - for example, loan prediction [2],[3] or fraud detection [13], [14] in isolation - with no paper combining identity verification, document processing, fraud screening, and approval prediction end-to-end.
- Limited integration of loan and insurance decisioning: loan-focused studies [2]–[10],[19],[20] and insurance-focused studies [11],[12] form separate literatures, with no reviewed paper addressing both.
- Limited integration of KYC with financial prediction: the e-KYC study [17] does not connect identity verification to downstream loan or insurance decisioning, and none of the loan/credit-risk papers incorporate biometric identity verification.
- Limited integration of OCR/document intelligence with decision models: [18] extracts structured financial fields with high accuracy and traceability but stops short of feeding those fields into an explainable approval decision.
- Fraud detection as a separate module: both reviewed fraud papers [13],[14] and the broader review [15] confirm that fraud screening is rarely embedded within a unified approval pipeline; [15] also finds that loan and mortgage fraud are comparatively under-studied relative to credit-card and banking fraud.
- Post-hoc explainability dominates over explainability-by-design: most studies apply

SHAP/LIME after training a black-box model, while only [8] and [9],[10] pursue explanation mechanisms more tightly integrated into the modeling or reasoning process itself.

- Limited applicant-centered explanation: [1] and [11] argue explicitly that customer-facing explanations must be simpler and more actionable than the technical SHAP/LIME outputs most other reviewed studies produce for data-scientist audiences.
- Dataset imbalance and small-sample limitations: severe class imbalance is explicitly addressed via resampling or focal-loss variants in fraud-detection studies [13],[15], while [12] relies on a small, self-collected sample of 145 responses.
- Limited external validation: several studies note results confined to a single dataset or platform, not yet validated across institutions or geographies [3],[4],[6],[17].
- Fairness and demographic bias: [20] is the only reviewed study to center fairness explicitly, and even there, formal fairness auditing beyond feature-importance analysis is not the primary focus; bias is flagged but not deeply evaluated in [8],[17].
- Explanation stability: [16] is the only reviewed paper to quantify the consistency of SHAP explanations across predictions, indicating this is an under-explored dimension of financial XAI.
- Regulatory compliance: only [1] and [7] explicitly connect XAI design choices to named regulatory frameworks (GDPR, Basel III), pointing to a broader gap in regulation-aware model design.
- Privacy-preserving financial data access: none of the reviewed papers implement a consent-based data-access mechanism analogous to an Account Aggregator framework, though [15] discusses federated learning and privacy-preserving techniques conceptually.
- Personalized financial recommendation: only [11] and, to a lesser extent, [12] move beyond a binary approval/denial decision toward personalized product recommendation.

XI. ARTHAX: A PROPOSED INTEGRATED EXPLAINABLE AI ARCHITECTURE

ArthaX suggests a holistic, explainable decision-support workflow for loans and insurance approval based on the identified gaps in Section X. ArthaX is not trying to be novel in any individual direction of research, but

rather a porting of such directions, singly seen in the reviewed literature, but not previously seen in a single pipeline for an Applicant.

A. System Architecture

Organized around the core processing layer of a KYC Verification module, a Document Analysis module, a Fraud Detection module, and a Credit Approval and Risk Prediction module, all of which provide input into an Explainable AI Engine that serves users (borrowers/policyholders, loan and insurance underwriters, and administrators/compliance auditors). External financial-information providers are connected to a set of services and APIs - such as identity, tax, and financial-affairs registries - through an Account

Aggregator (AA) framework to gain access to the encrypted data.

Provide evidence to back up their data. The applicant and staff are interacting with the system via a web/mobile portal, and a Decision Output is generated for each automated decision, along with its explanation based on SHAP. Table III shows the proposed module for each function and to the specific reviewed study that motivates it.

TABLE III

ArthaX Modules and Their Grounding in the Reviewed Literature

Module	Function	Grounded in Reviewed Literature
KYC Verification (FaceNet)	Compares a live image against identity-proof documents to confirm applicant identity and reduce fraudulent profiles.	Motivated by the confidence-aware, resolution-robust facial-verification approach in [17].
Document Analysis (OCR)	Extracts structured fields (income, employer, account balances) from salary slips, bank statements, and reports.	Motivated by the multi-stage OCR/vision-language extraction pipeline in [18].
Fraud Detection (Isolation Forest)	Screens applications and transactions for anomalous patterns before a financial decision is finalized.	Motivated by [14] and informed by the governance considerations raised in the fraud-detection review [15].
Credit Approval & Risk Prediction (XGBoost)	Predicts loan or insurance approval outcomes from applicant and financial information.	Draws on the gradient-boosting approaches shown effective in [3],[7],[8],[12],[19].
Explainable AI Engine (SHAP)	Generates applicant-facing explanations that identify which factors drove an approval or rejection.	Informed by the stakeholder-oriented explainability principles in [1] and the SHAP/LIME/counterfactual toolkit in [11].
Recommendation Engine	Suggests safer or more eligible financial products rather than returning a bare approval/rejection.	Extends the premium-personalization concept demonstrated in [12] and the recommendation logic in [11].
Account Aggregator (AA) Framework	Governs secure, consent-based access to the applicant's financial data across banks and institutions.	Not observed in any reviewed study; identified in this survey as an open gap (Section X).

B. Proposed Methodology

The first step in the conceptual workflow is the collection of data, which involves the collection of applicant information, uploaded documents, and financial information from which to verify and assess risk. KYC

verification finally uses a FaceNet module to compare the live image with the identity-proof documents, inspired by the concept of confidence-aware identity-verification in [17]. Document processing uses OCR to process salary slips, bank statements and reports due to the long pipeline

of OCR/vision-language extraction in [18]. Motivated by [14] and with guidance from the governance concerns identified in [15] fraud detection uses Isolation Forest to detect suspicious or abnormal patterns prior to arriving at a decision. Risk prediction uses XGBoost to predict loan or insurance approval responses, inspired by the gradient-boosting methods that proved to be effective in [3],[7],[8],[12], [19]. Explainability provided through SHAP, that outlines what drove the approval or denial decision and that translates a prediction result into a decision that can be explained to the applicant according to the stakeholder principles.

The explanation toolkit of [1] and the wider explanation toolkit in [11]. Last but not least, recommendation generation suggests safer or more suitable

The distinction over personalization in [12] is extended to financial options—apart from an approval/rejection result, a user can receive a list of available options.

C. Applications

Loan Approval - ArthaX can support automated loan assessment by processing applicant information, evaluating risk, predicting approval, and explaining the resulting decision to the applicant.

Insurance Underwriting - The system can assist insurance-related decisions by combining financial information with explainable machine learning to support personalized, transparent policy recommendations.

Fraud Detection - Anomaly-detection integration can help identify suspicious profiles and abnormal financial patterns before a decision is completed, reducing exposure to the fraud types documented in [13]–[15].

Financial Inclusion - By reducing manual processing and providing understandable decision explanations, the proper system aims to make financial services more accessible as well as more transparent.

Personalized Financial Guidance - Rather than a binary approval or rejection outcome, ArthaX is designed to suggest safer and more suitable financial options, extending the personalization direction found in [11],[12].

D. Positioning ArthaX Relative to the Reviewed Literature

It is imperative to clearly differentiate what is currently being done (research) and what is proposed here (ArthaX integration). Every step of the ArthaX pipeline is driven by a particular aspect of the literature they have reviewed, but no paper examined in this literature review performs more than two or three of the steps in sequence

and no performance measure in Table I is valid for the ArthaX pipeline. This survey neither claims to be accurate, precise, recall, F1-score, nor has ArthaX been experimentally evaluated for its fraud-reduction rate, or any other performance metric. ArthaX does not contribute to any single algorithmic component (FaceNet-based KYC, OCR-based document intelligence, Isolation Forest-based fraud detection, XGBoost-based prediction, SHAP-based explanation, and recommendation), but rather, the proposed combination of components in a single, consent-governed loan-and-insurance decision-support workflow. Therefore, ArthaX is not presented as a system that is shown to be superior to the individually reviewed systems, but rather as a direction of research and system integration to be validated in future experimental work.

XII. LIMITATIONS AND CHALLENGES

Data Quality - Machine-learning models depend on the quality and representativeness of training data. Poor or biased datasets can produce unreliable predictions, and several reviewed studies rely on small or non-public samples [12],[13].

Computational Requirements - Advanced machine learning and explainability techniques, especially Shapley-value computation, can require substantial compute at scale [5],[6],[20], which is in tension with real-time explainability [7],[19].

Privacy - Financial and biometric information is sensitive; a system such as ArthaX therefore needs strong data-protection mechanisms, an area only discussed conceptually in the reviewed literature [15].

Explainability Trade-off - A technically accurate explanation may still be difficult for non-technical users to understand; explanations need to be converted into simple, actionable language, as argued in [1] and [11].

Model Drift - Financial behavior and economic conditions change over time, and models trained on historical data may require ongoing monitoring and updating, a concern raised but not empirically evaluated in [7],[9].

Regulatory Compliance - Financial AI systems must operate within applicable legal and regulatory frameworks; only [1] and [7] connect XAI design choices to named frameworks such as GDPR or Basel III.

XIII. FUTURE SCOPE

- Real-Time Financial Data: integrating continuously updated financial information could improve risk assessment.
- Advanced Counterfactual Explanations: providing specific, actionable recommendations for improving future approval probability.
- Fairness Monitoring: integrating formal fairness metrics beyond the feature-importance analysis used in [20] to identify potentially discriminatory outcomes.
- Multimodal AI: jointly processing text, financial records, images, and other information sources.
- Continuous Model Monitoring: applying drift detection to maintain model performance over time, extending the concerns raised in [7],[9].
- Natural-Language Explanations: converting SHAP results into explanations understandable to non-technical applicants, in the spirit of [1] and [11].
- Privacy-Preserving AI: investigating techniques such as federated learning and privacy-aware processing, as discussed conceptually in [15].
- Dynamic Insurance Recommendations: supporting continuously updated rather than only static policy recommendations, extending [11].
- Scalable Explainability: developing efficient approximation methods to reduce the computational cost of Shapley-value calculations, a limitation noted in [5], [6],[20].

XIV. DISCUSSION

The literature suggests that explainability is being more frequently expected, not just an optional feature, in financial decision making using AI. Across various sensitive financial applications, predictive models like XGBoost, Random Forest, Gradient Boosting, and deep neural networks offer great predictive power [3],[4],[19] but the demand for transparency remains. There is no single solution able to solve all the needs: SHAP and LIME can be used for addressing model interpretation, but cannot be used as a standalone tool to solve fraud issues related to identity and document processing, and biometric verification can help securing identities but cannot explain a financial risk prediction.

This disintegration is what drives the integrated architecture offered by ArthaX: verification is performed using FaceNet, financial document processing is performed using OCR, anomaly detection is performed using Isolation Forest, explanation and risk prediction are

performed using XGBoost, and the recommendation is performed using a layer that converts the recommendation into a user-oriented financial recommendation. This is useful because there are multiple steps that go into making a financial decision, and a good system needs to go through the following steps: first, identify the applicant; second, understand what financial information is being presented; third, identify potentially alarming behavior; fourth, assess financial risk; fifth, communicate why the model has concluded its decision; and sixth, communicate what the applicant can do next. Empirical validation of the integrated pipeline should be a priority in future ArthaX work, including benchmarking the individual components against the reviewed studies, formal evaluation for fairness and explanation-stability and pilot testing with the end user (applicant, not data scientist) of explanations that are built using the SHAP-based pipeline.

XV. CONCLUSION

This survey has analyzed twenty recent papers pertaining to explainable AI in loan approval, credit-risk assessment, insurance underwriting, fraud detection, KYC, and financial-document intelligence, summarizing their models, XAI techniques, data, and results. The evidence suggests that the most dominant and well-validated method for combining tree models with explainability is gradient-boosting and ensemble tree models; the explainability research of insurance, fraud, KYC, and document-intelligence is more fragmented and less consistently associated with tree models. Importantly, no reviewed study brings together identity verification, document processing, fraud screening, predictive decisioning and explanation to the applicant in an end-to-end system. This gap inspires ArthaX, an integrated explainable AI (XAI) direction made up of six modules: FaceNet-based KYC, OCR-based document intelligence, Isolation Forest-based fraud detection, XGBoost-based prediction, SHAP-based explanation, personalized recommendation, and Account Aggregator-based data access. The presented approach of ArthaX is not empirical but rather a research direction based on reviewed literature and future research. More generally, this survey indicates that the future of financial XAI research should shift from post hoc and post financial towards integrated and financial applicant-centered and verified consistent explainable decision systems.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Prof. Leenata Parab (Project Guide) for guidance, support and encouragement during the development of the ArthaX project. Finally, the authors would like to express their gratitude to all of the colleagues and members of their faculty and institution for their support as they went about doing this work.

REFERENCES

- [1] C. -A. Wilson, "Explainable AI in Finance: Addressing the Needs of Diverse Stakeholders," CFA Institute Research and Policy Center, 2025.
- [2] R. Akter, R. K. Halder, M. Rahman, Md. A. Uddin, A. Khraisat, and M. N. Uddin, "Prediction with SHAP-guided feature selection and LIME-based explainability for loan approval," *International Journal of Intelligent Systems*, 2026.
- [3] L.-H. Li, A. K. Sharma, and S.-T. Cheng, "Explainable AI-based LightGBM prediction model for default borrowers on social lending platforms," *Intelligent Systems with Applications*, vol. 26, Art. no. 200514, 2025.
- [4] V. Ravi, S. R. Mani Sekhar, and S. Siddesh, "Explainable AI (XAI) for credit scoring and loan approvals," *Journal of Financial Data Science*, 2025.
- [5] M. N. Sowmiya, Jaya Sri S., Deepshikha S., and Hanushya Devi G., "Credit risk analysis using explainable artificial intelligence," *Journal of Soft Computing Paradigm*, vol. 6, no. 3, 2024.
- [6] M. K. Nag et al., "Credit risk assessment and financial decision support using XAI," *Risks (MDPI)*, vol. 12, Art. no. 164, 2024.
- [7] S. Kakkar, "Explainable AI models for credit risk scoring in banking: Balancing accuracy and regulatory transparency," *International Journal of Financial Data Science (IJFDS)*, vol. 3, no. 2, pp. 1–6, 2025.
- [8] K. Madhu, V. Shirisha, A. Sonya, and L. R. Chandra, "Future of loan approvals using explainable AI," *International Journal of Scientific Research in Engineering Trends (IJSRET)*, vol. 11, no. 3, 2025.
- [9] C. S. Kubam, "Agentic AI for autonomous, explainable, and real-time credit risk decision-making," *International Journal of Intelligent Systems and Applications in Engineering (IJISAE)*, vol. 12, no. 23s, pp. 3669–3676, 2024.
- [10] Y. Shi, Z. Yang, Y. Wang, and L. Zhou, "CreditXAI: A multi-agent system for explainable corporate credit rating," *arXiv:2510.22222*, 2025.
- [11] H. S. B. Gummadi, "Explainable AI-enhanced underwriting automation for personalized insurance policy recommendations," *European Journal of Computer Science and Information Technology*, vol. 13, no. 19, pp. 24–40, 2025.
- [12] M. Kapse, V. Sharma, R. Vidhale, and V. Vellanki, "Customization of health insurance premiums using machine learning and explainable AI," *International Journal of Information Management Data Insights*, vol. 5, Art. no. 100328, 2025.
- [13] F. Boabang and S. A. Gyamerah, "An enhanced focal loss function to mitigate class imbalance in auto insurance fraud detection with explainable AI," *arXiv:2508.02283*, 2025.
- [14] V. Tripathi, D. Mishra, D. Gupta, and P. Tiwari, "Fraud detection using isolation forest algorithm in credit card," *International Journal of Humanities and Social Science & Management (IJHSSM)*, 2023.
- [15] Y. Chen, C. Zhao, Y. Xu, C. Nie, and Y. Zhang, "Deep learning in financial fraud detection: Innovations, challenges, and applications," *Data Science and Management*, 2025.
- [16] E. Fernández-Morales, L. Tobarra, A. Robles-Gómez, R. Pastor-Vargas, R. Hernández, and J. Sarraipa, "eXplicable AI (XAI) for attack detection toward smart rural applications," *Internet of Things*, vol. 34, Art. no. 101804, 2025.
- [17] M. Ketcham, P. Boonyopakorn, and T. Ganokratanaa, "Resolution-aware deep learning with feature space optimization for reliable identity verification in electronic know your customer processes," *Mathematics*, vol. 13, Art. no. 1726, 2025.
- [18] Y. Jin, Y. Wang, Q. Zhong, K. C. Jin-Chun, K. Z. Ke, and D. MacDonald, "Multi-stage field extraction of financial documents with OCR and compact vision-language models," *arXiv:2510.23066*, 2025.
- [19] A. Mujo, S. Nikolla, E. Hoxha, and E. Pelivani, "Explainable AI in credit scoring: Improving transparency in loan decisions," *Journal of*

Information Systems Engineering and
Management, 2025, doi: 10.55220/2576-
6821.v9.709.

- [20] D. B. Acharya, B. Divya, and K. Kuppan,
"Explainable and fair AI: Balancing performance
in financial and real estate machine learni
models," IEEE Access, vol. 12, 2024, doi:
10.1109/ACCESS.2024.3484409.