



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Developing an Explainable Artificial Intelligence–Based Model for Early Software Defect Prediction And Software Quality Improvement

Ahmad Amjed Ahmad Ahmad

Software Engineering

Abstract

Background: Software defects remain a major source of rework, maintenance cost, operational risk, and deterioration in software quality. Machine-learning-based software defect prediction can prioritize modules that are more likely to contain defects, but high-performing models are often difficult for practitioners to interpret. Explainable artificial intelligence (XAI) offers a mechanism for making predictions more transparent and actionable.

Aim: This study develops an integrated XAI-based framework for early software defect prediction and software quality improvement and evaluates software professionals' perceptions of usefulness, explainability, trust, quality improvement, and adoption intention.

Methods: A two-component design is proposed. The technical component uses static and change-oriented software metrics to train Logistic Regression, Random Forest, and XGBoost classifiers. Data preprocessing includes missing-value treatment, duplicate control, stratified partitioning, class-imbalance management within training folds, hyperparameter tuning, and held-out testing. Performance is evaluated using precision, recall, F1-score, ROC-AUC, PR-AUC, balanced accuracy, and calibration. SHAP is used for global and local post-hoc explanations. The human-evaluation component uses a 34-item five-point Likert questionnaire plus eight demographic/professional items. For the worked research model, a synthetic sample of 300 software professionals is analyzed.

Results: XGBoost achieved the strongest held-out performance (accuracy 0.943, precision 0.936, recall 0.928, F1 0.932, ROC-AUC 0.970). SHAP ranked code churn, cyclomatic complexity, prior defects, coupling, and lines of code among the most influential features. In the survey, all multi-item scales demonstrated acceptable-to-excellent internal consistency ($\alpha = .82-.91$). Explainability significantly predicted trust, and trust significantly predicted adoption intention. Early defect prediction and perceived usefulness significantly predicted perceived software quality improvement. The indirect effect of explainability on adoption through trust was positive.

Conclusion: An XAI-oriented defect-prediction workflow can combine predictive performance with interpretable evidence that supports developer review and quality decisions. The framework should be validated with real repository data and real practitioner responses before empirical claims are made.

Keywords: explainable artificial intelligence; XAI; software defect prediction; software quality; machine learning; SHAP; XGBoost; software engineering.

Background

Modern software systems support financial services, healthcare, transportation, public administration, education, commerce, and critical infrastructure. As software grows in size, interdependence, and release frequency, defects can propagate rapidly from code changes into production incidents. Defect prevention and early detection are therefore central to software quality assurance. Software defect prediction (SDP) uses historical project information and software metrics to estimate which modules, files, classes, or changes are more likely to be defective, allowing testing and review resources to be prioritized. (Hall et al., 2011; Menzies et al., 2007; Kamei et al., 2012)

Traditional statistical models have been supplemented by machine-learning approaches capable of representing nonlinear relationships and interactions among software metrics. Prior research has evaluated decision trees, random forests, support vector machines, neural networks, boosting methods, and other classifiers for defect prediction. However, predictive performance alone does not ensure practical usefulness. A model may assign a high defect probability without telling a developer which measurable characteristics contributed to the prediction or what should be inspected first. (Breiman, 2001; Chen & Guestrin, 2016; Khan et al., 2022)

This interpretability problem is particularly important in software engineering because model outputs are usually inputs to human decisions rather than autonomous endpoints. Explainable artificial intelligence seeks to expose information about how model inputs influence predictions. SHAP, based on Shapley values, provides feature-attribution explanations that can be aggregated globally or inspected for individual predictions (Lundberg & Lee, 2017). Systematic reviews of XAI in software engineering show that software maintenance

and defect prediction are prominent application areas while also identifying unresolved issues related to explanation evaluation, stability, usability, and practitioner validation (Mohammadkhani et al., 2023; Cao et al., 2024). (Lundberg & Lee, 2017; Mohammadkhani et al., 2023; Tantithamthavorn et al., 2020)

Software quality in this study is interpreted in relation to systematic quality engineering rather than as defect count alone. ISO/IEC 25010:2023 defines a product-quality model comprising nine characteristics and their subcharacteristics, providing a reference for specifying and evaluating software and ICT product quality. Early defect prediction is therefore treated as a quality-support mechanism that may improve review focus, maintainability-related decisions, reliability, and testing effectiveness rather than as a substitute for comprehensive quality assurance. (International Organization for Standardization [ISO], 2023)

Problem Statement

Despite advances in machine-learning-based defect prediction, three connected problems remain. First, software defects are frequently discovered after substantial development effort has already been invested, increasing rework and maintenance cost. Second, many accurate predictive models are black boxes whose outputs are difficult for software professionals to interrogate. Third, studies often evaluate predictive performance without simultaneously examining whether practitioners perceive the explanations as useful, trustworthy, and suitable for adoption. Consequently, a technically accurate model may fail to influence real quality-improvement practices. (Hall et al., 2011; Mohammadkhani et al., 2023)

The research problem is therefore the need for an integrated approach that predicts defect risk early, explains the factors driving individual and aggregate predictions, and evaluates the human factors that influence the use of those explanations in software quality decisions.

Aim of the Study

To develop and evaluate an explainable artificial intelligence-based framework for early software defect prediction and software quality improvement.

Specific Objectives

Develop machine-learning models for early software defect prediction using software metrics.

Compare the predictive performance of Logistic Regression, Random Forest, and XGBoost.

Identify software metrics that contribute most strongly to predicted defect risk.

Apply SHAP to provide global and local explanations of defect predictions.

Assess software professionals' perceptions of early defect prediction and perceived usefulness of AI.

Assess perceived explainability/transparency and trust in AI-based defect prediction.

Examine the relationship between early defect prediction and perceived software quality improvement.

Examine the effect of trust on intention to adopt XAI-based defect prediction.

Test whether trust mediates the relationship between explainability and adoption intention.

Propose an integrated decision-support framework connecting defect risk, explanation, human review, and software-quality action.

Research Questions

RQ1. How accurately can selected machine-learning algorithms predict defective software modules?

RQ2. Which model provides the best balance of discrimination, recall, precision, and interpretability?

RQ3. Which software metrics contribute most to the selected model's predictions?

RQ4. What are software professionals' perceptions of early defect prediction, AI usefulness, explainability, trust, software-quality improvement, and adoption intention?

RQ5. Does explainability significantly predict trust in AI-based defect prediction?

RQ6. Do early defect prediction and perceived usefulness significantly predict perceived software quality improvement?

RQ7. Does trust significantly predict intention to adopt XAI-based defect prediction?

RQ8. Does trust mediate the relationship between explainability and adoption intention?

Research Hypotheses

H1: Explainability and transparency have a significant positive effect on trust in AI-based defect prediction.

H2: Perceived usefulness of AI-based defect prediction has a significant positive effect on perceived software quality improvement.

H3: Early software defect prediction has a significant positive effect on perceived software quality improvement.

H4: Trust in AI-based defect prediction has a significant positive effect on intention to adopt XAI-based defect prediction systems.

H5: Explainability and transparency have a significant positive direct effect on intention to adopt XAI-based defect prediction systems.

H6: Trust mediates the relationship between explainability and intention to adopt XAI-based defect prediction systems.

Significance of the Study

The study contributes technically by specifying an end-to-end, leakage-aware defect-prediction pipeline that reports multiple discrimination and calibration measures rather than accuracy alone. It contributes methodologically by separating predictive evidence from explanatory evidence and by treating SHAP values as model-attribution information rather than causal effects. It contributes practically by linking explanations to human review, thereby supporting prioritization of modules for testing, refactoring, code review, and root-cause investigation. Finally, the survey component addresses the socio-technical reality that explainability, trust, and perceived usefulness influence whether AI tools are adopted. (Saito & Rehmsmeier, 2015; Tantithamthavorn et al., 2016)

Conceptual Framework

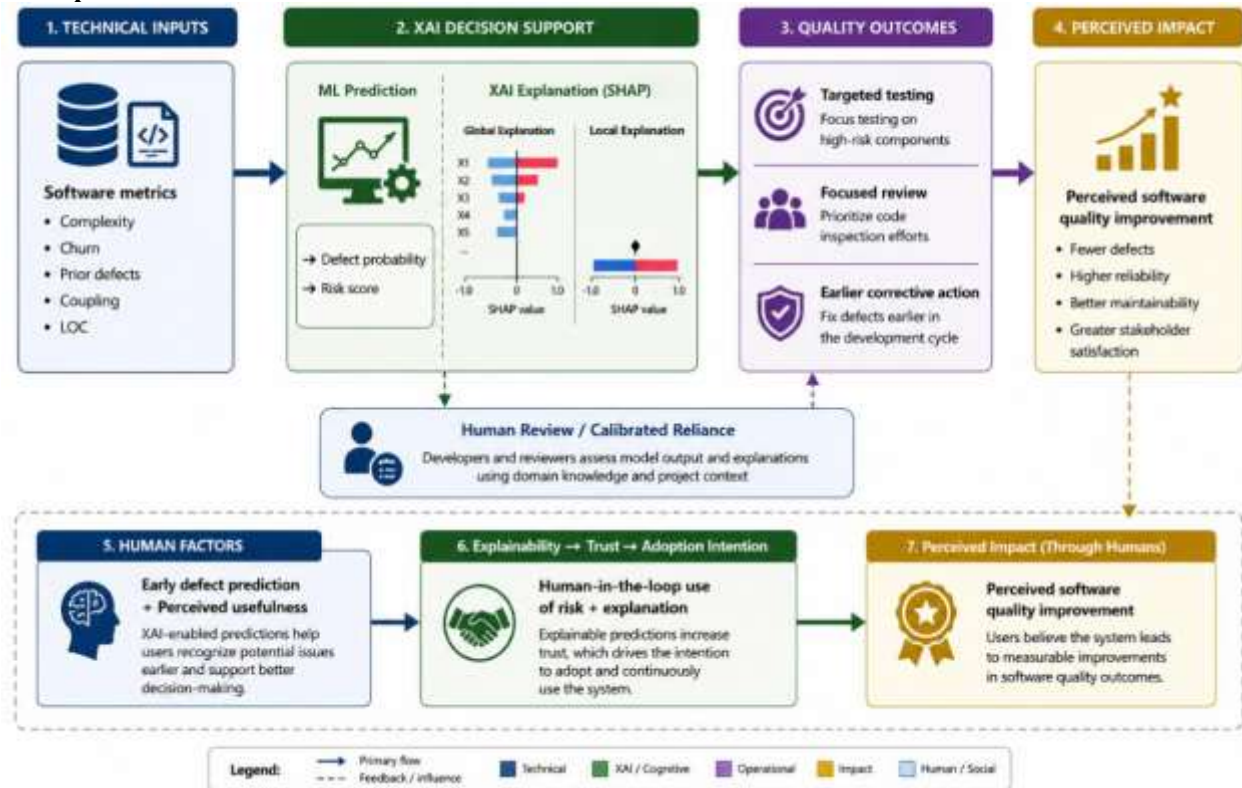


Figure 1.1. Integrated Conceptual Framework

Operational Definitions

Term	Operational definition
Software defect	A fault or flaw in a software artifact that can contribute to incorrect or undesired behavior.
Early software defect prediction	Estimation of defect risk using information available sufficiently early to prioritize review or testing before release.
Explainable artificial intelligence	Methods that provide human-interpretable information about the behavior or predictions of an AI/ML model.
SHAP	SHapley Additive exPlanations; a feature-attribution framework that assigns contributions to input features for a model prediction.
Software quality improvement	Improvement in quality-support decisions such as defect prevention, focused testing, maintainability, reliability-oriented review, and efficient resource allocation.
Trust	The professional's willingness to rely on model outputs within appropriate boundaries based on perceived competence, transparency, and consistency.
Adoption intention	The reported likelihood of using or recommending an XAI-based defect-prediction system in software development or QA work.

Scope and Delimitations

The technical scope is binary defect prediction at module/file/class level using measurable software characteristics. The human scope covers software developers, software engineers, QA/test engineers, DevOps professionals, analysts, technical leads, and project managers. The study does not claim that feature attribution establishes causality. It also does not replace code review, testing, security analysis, or professional judgment.

Literature Review

Introduction

This section reviews software defects and quality, software metrics, machine-learning-based defect prediction, class imbalance and validation, explainable AI, SHAP, human trust and adoption, and the research gap.

Software Defects and Software Quality

Defects arise from requirements misunderstandings, design weaknesses, coding errors, interface problems, configuration issues, concurrency problems, inadequate exception handling, and change-related regressions. Their consequences vary from minor usability problems to failures that affect reliability, safety, privacy, security, and business continuity. Quality assurance therefore uses complementary preventive and detective activities, including standards, reviews, static analysis, testing, continuous integration, observability, and post-release learning. (Hall et al., 2011; Kamei et al., 2012)

ISO/IEC 25010:2023 provides a contemporary product-quality reference model with nine quality characteristics. The model is useful in this research because it prevents an overly narrow equation of 'quality' with a single predictive metric. A defect predictor can support quality work, but the quality of a software product remains multidimensional. (ISO, 2023)

Software Metrics Used in Defect Prediction

Defect-prediction studies commonly use size, complexity, coupling, cohesion, inheritance, historical defects, and change/process metrics. Examples include lines of code (LOC), cyclomatic complexity, weighted methods per class, coupling between objects, response for a class, depth of inheritance, number of children, code churn, number of revisions, number of developers touching an artifact, and prior defect counts. Process metrics can be especially informative because they represent how a component evolves, not only its static structure. (Menzies et al., 2007; Kamei et al., 2012; Jiarpakdee et al., 2018)

Machine Learning for Software Defect Prediction

Logistic regression remains valuable as a transparent statistical baseline. Random forests can model nonlinear effects and interactions while reducing the variance of individual trees. Gradient-boosting systems such as XGBoost sequentially improve an ensemble by focusing on residual errors and include regularization mechanisms. Neural approaches have also been widely studied. Systematic reviews indicate sustained interest in automated defect prediction as a mechanism for directing quality-assurance effort (Khan et al., 2022). (Breiman, 2001; Chen & Guestrin, 2016; Friedman, 2001)

Evaluation Challenges

Software defect datasets are frequently imbalanced because non-defective artifacts outnumber defective artifacts. Under imbalance, accuracy can be misleading. Recall is important when missed defects are costly, precision reflects the proportion of predicted positives that are actually defective, F1 balances precision and recall, ROC-AUC measures ranking discrimination, and PR-AUC is informative when positive cases are relatively rare. Calibration should also be examined when predicted probabilities will be interpreted as risk. Resampling must be confined to training data to prevent information leakage. (Saito & Rehmsmeier, 2015; Zou et al., 2016)

Explainable Artificial Intelligence

Explainability addresses the gap between predictive capability and human understanding. In software engineering, explanations can support debugging of the predictive model, identification of suspicious correlations, validation against domain knowledge, prioritization of code inspection, and communication with technical stakeholders. Tantithamthavorn, Jiarpakdee, and Grundy (2020) argued that explainability can make AI-assisted software-engineering solutions more practical and actionable. Subsequent systematic reviews have documented rapid growth in XAI applications while highlighting limited standardization of explanation evaluation and limited practitioner-centered validation (Mohammadkhani et al., 2023; Cao et al., 2024). (Mohammadkhani et al., 2023; Tantithamthavorn et al., 2020)

SHAP as an Explanation Method

Lundberg and Lee (2017) introduced SHAP as a unified additive feature-attribution framework. For a particular prediction, SHAP values allocate the difference between a baseline expectation and the model output among the input features. Global summaries can be constructed by aggregating absolute attribution magnitudes, while local plots can show which features pushed one prediction toward or away from the defective class. Importantly, SHAP explains the fitted model; it does not by itself prove that changing a feature will causally change defect occurrence. (Lundberg & Lee, 2017)

Trust, Explainability, and Adoption

Technology adoption depends on more than technical accuracy. Professionals may reject a model that appears opaque, inconsistent, or disconnected from workflow. Conversely, explanation can support calibrated trust when it helps users understand why a prediction was produced, detect unreasonable outputs, and determine when human override is warranted. The appropriate goal is calibrated reliance rather than unconditional trust. (Ribeiro et al., 2016; Tantithamthavorn et al., 2020)

Synthesis of Previous Studies

A 2023 systematic literature review and meta-analysis of deep learning for software defect prediction synthesized datasets, measures, model families, and comparative performance, confirming strong research interest while also underscoring heterogeneity across empirical designs (Zain et al., 2023). These findings support the present study's decision to retain interpretable baselines and tree-based ensembles rather than assume that greater architectural complexity automatically yields better deployable quality support. In practice, performance, calibration, explanation fidelity, computational cost, and workflow fit must be considered together.

Deep Learning, Explainability, and Practical Trade-offs

Gezici Geçer and Kolukısa Tarhan (2025) proposed a peer-reviewed explainable-AI framework specifically for software defect prediction, emphasizing the practical difficulty of deploying complex models when their reasoning is opaque. Yu et al. (2024) further showed that just-in-time defect prediction requires explanations that are not merely plausible but also sound, robust, and actionable. More recently, Kun et al. (2026) reported an explainable framework combining advanced predictive modeling with explainable analytics for software defect prediction, illustrating the continuing shift toward models that jointly optimize predictive performance and interpretability.

Recent evidence strengthens the case for moving beyond binary black-box predictions. A 2024 systematic review in Information and Software Technology concluded that conventional defect-prediction outputs often lack information about defect type, severity, cause, and actionable remediation, and identified explainable AI as a promising route toward more informative prediction (Grattan et al., 2024). Industrial evidence likewise shows that adoption lags behind academic performance and highlights the importance of user feedback, comparisons across learners, mixed research/open-source/industrial datasets, and business-relevant evaluation (Stradowski & Madeyski, 2023).

Recent Evidence on Informative and Explainable Defect Prediction

Table 2.1. Synthesis of Selected Previous Studies

Study	Design/Data	Models/Met hod	XAI	Main contributio n	Limitation/G AP	Relevance
Hall et al. (2011)	Systematic review	Fault- prediction studies	Not central	Synthesized prediction performanc e	Predates modern XAI	Baseline evidence
Kamei et al. (2012)	Large-scale JIT empirical study	Change-level prediction	No	Demonstrat ed JIT quality assurance	Limited explanation focus	Supports early/chan ge-based prediction
Mohammadkh ani et al. (2023)	Systematic review	XAI for software engineering	Multiple XAI methods	Defect prediction prominent in XAI4SE	Lack of standardized XAI evaluation	Motivates explanatio n evaluation
Information & Software Technology (2024)	Systematic review	Informative defect prediction	XAI highlighte d	Calls for actionable outputs beyond binary labels	Actionability remains limited	Supports quality- action focus
Yu et al. (2024)	Empirical/for mal explainer	JIT defect prediction	Formal explanatio ns	Targets sound, robust, actionable explanatio s	Focused on JIT setting	Strengthen s explanatio n rigor

Study	Design/Data	Models/Met hod	XAI	Main contributio n	Limitation/G AP	Relevance
Gezici Geçer & Kolukisa Tarhan (2025)	Peer-reviewed empirical framework	ML-based SDP	Explainabl e framework	Connects prediction with interpretability	Generalization still requires broader validation	Directly aligned with proposed framework
Kun et al. (2026)	Empirical framework	Transformer-assisted boosting	Explainabl e analytics	Recent integrated XAI-SDP framework	Needs replication across contexts	Demonstrates current research direction

Research Gap

The literature reveals a sharper multi-level gap. First, performance-centered defect-prediction studies frequently stop at a binary risk output and provide limited actionable information for remediation. Second, post-hoc explanations are often evaluated technically rather than with the software professionals expected to use them. Third, industrial adoption evidence remains thinner than academic benchmarking evidence. Fourth, explanation quality is not equivalent to causal validity, and unstable or correlated predictors can distort interpretation. Fifth, many studies treat software quality narrowly as classification performance rather than connecting risk estimates to quality-management actions. The present study addresses these gaps by integrating leakage-aware comparative prediction, probability-oriented evaluation, SHAP-based global and local attribution, practitioner evaluation of explainability/trust/adoption, and an explicit human-in-the-loop quality-improvement pathway (Jiarpakdee et al., 2018; Mohammadkhani et al., 2023; Yu et al., 2024; Gezici Geçer & Kolukisa Tarhan, 2025).

Methodology

Research Design

A two-component explanatory and developmental research design is proposed. Component I is a quantitative machine-learning experiment for binary software defect prediction. Component II is a cross-sectional quantitative survey of software professionals. Integration occurs at interpretation: technical model behavior is linked to professional perceptions of usefulness, explainability, trust, quality improvement, and adoption.

Component I: Dataset and Unit of Analysis

The preferred empirical implementation should use a public, versioned software-defect dataset with artifact-level defect labels and documented metrics, or a reproducibly extracted dataset from open-source repositories. The unit of analysis is a software module/file/class. The dependent variable is defect status (0 = non-defective; 1 = defective). Candidate predictors include LOC, cyclomatic complexity, code churn, coupling, cohesion-related measures, number of revisions, prior defects, and developer/activity measures where available. (Menzies et al., 2007; Kamei et al., 2012)

Data Preprocessing

Verify the dataset schema, version, licensing, and label definition. Remove exact duplicates and investigate impossible or structurally missing values. Perform exploratory distribution checks and document transformations. Create a stratified train/validation/test split or nested cross-validation design. Fit preprocessing steps only on training folds to avoid leakage. Standardize numeric features when required by the algorithm. Address class imbalance using class weights or training-only resampling; never resample the test set. Check highly correlated/redundant predictors and preserve a reproducible feature dictionary.

Candidate Algorithms

Model	Rationale
Logistic Regression	Transparent baseline; interpretable direction and magnitude on the log-odds scale.
Random Forest	Robust nonlinear ensemble; captures interactions and nonlinear thresholds.
XGBoost	Regularized gradient boosting; strong tabular-data performance and compatible with efficient SHAP explanations.

Training and Validation Strategy

Hyperparameters should be selected using stratified cross-validation on the training data. The test set must remain untouched until final model selection. Where project versions or temporal ordering are available, a temporal or cross-version evaluation is preferable to a random split because it better approximates future prediction. Random seeds, software versions, preprocessing objects, and model configurations should be recorded (Kuhn et al., 2013; Tantithamthavorn et al., 2016)

Performance Measures

Measure	Purpose
Precision	Proportion of predicted defective artifacts that are truly defective.
Recall/Sensitivity	Proportion of defective artifacts detected by the model.
F1-score	Harmonic balance of precision and recall.
ROC-AUC	Overall discrimination across thresholds.
PR-AUC	Precision-recall discrimination, especially useful with class imbalance.
Balanced accuracy	Average of sensitivity and specificity.
Brier score / calibration curve	Quality of predicted probabilities when risk estimates are used.

Explainability Procedure

After final model selection, SHAP is applied to the locked model. Global analysis includes mean absolute SHAP importance and a distributional summary plot. Local analysis includes waterfall/force-style explanations for representative true-positive, false-positive, true-negative, and false-negative cases. Explanation stability should be checked across resamples or model refits where feasible. Explanations are interpreted as model behavior, not causal effects. (Lundberg & Lee, 2017)

Component II: Population and Sample

The target population comprises software professionals involved in development, testing, quality assurance, DevOps, analysis, technical leadership, or project management. A sample of approximately 300 participants is proposed to support stable descriptive, reliability, correlation, regression, and mediation analyses. A stratified or quota-assisted sampling strategy can improve representation across major roles and experience levels. The worked results use $N = 300$ cases solely to demonstrate the planned analysis.

Survey Instrument

The questionnaire contains eight demographic/professional questions and 34 Likert items across six constructs. Responses use a five-point scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5 = Strongly Agree. Higher scores indicate stronger endorsement of the construct.

Construct	Items	Score range
Early Software Defect Prediction (EDP)	6	6-30
Perceived Usefulness of AI (PU)	6	6-30
Explainability & Transparency (XAI)	6	6-30
Trust in AI Prediction (TR)	5	5-25
Perceived Software Quality Improvement (SQI)	6	6-30
Intention to Adopt (IA)	5	5-25

Questionnaire Development Procedure

The questionnaire is researcher-developed from the study's conceptual model and literature on defect prediction, explainability, trust, and adoption. Item generation follows a construct-to-item matrix to ensure that every item maps to one construct and research objective. Content validation should use expert ratings for relevance and clarity; item-level and scale-level content validity indices (I-CVI and S-CVI/Ave) may be reported. Cognitive/pilot testing should identify ambiguous wording, double-barreled items, excessive technical jargon, and completion burden. The final empirical study should then examine factor structure using EFA and/or CFA as appropriate, alongside internal-consistency estimates. Because the current questionnaire has not yet undergone empirical validation, it should be described as a proposed instrument rather than a previously validated scale.

Validity and Reliability

Content validity should be evaluated by 5-7 experts in software engineering, software quality, machine learning, and research methodology. Experts should assess relevance, clarity, coverage, and redundancy. A pilot study of approximately 30-40 eligible professionals should evaluate comprehensibility and completion time. Construct validity can be examined through exploratory and/or confirmatory factor analysis depending on the final sample and study requirements. Internal consistency should be evaluated using Cronbach's alpha and preferably McDonald's omega, with values around .70 or above generally considered acceptable for research use, interpreted alongside item content and dimensionality.

Data Collection Procedure

Following required institutional approval, participants should receive an information sheet explaining the purpose, voluntary nature, expected completion time, confidentiality, and data handling. No personally identifying information should be required unless necessary and approved. Consent should precede questionnaire access. Technical data should be obtained under the selected repository's terms and documented with dataset version and provenance.

Statistical Analysis

Survey analysis should include frequencies and percentages for categorical variables; means, standard deviations, medians, and ranges where appropriate for scale scores; reliability coefficients; Pearson or Spearman correlations according to assumptions; multiple linear regression for direct-effect hypotheses; and bootstrapped mediation for the indirect effect of explainability on adoption through trust. Statistical significance is set at $\alpha = .05$, but effect sizes and confidence intervals are reported to avoid binary interpretation of p-values.

Ethical Considerations

Participation must be voluntary, informed consent must be documented, withdrawal must be permitted before anonymous submission where feasible, and data must be stored securely. The survey should avoid collecting unnecessary identifiers. Public repository data should be used in accordance with licensing and terms. The present document contains survey and model results and therefore must not be represented as completed human-subject or repository-based empirical research.

Reproducibility and Threats to Validity

Internal validity threats include data leakage, duplicate artifacts, improper resampling, label noise, and hyperparameter overfitting. Construct validity threats include imperfect correspondence between software metrics and underlying quality concepts. External validity is limited when a model is trained on a small number of projects or languages. Conclusion validity can be threatened by class imbalance and overreliance on accuracy. Reproducibility is strengthened by versioned data, code, random seeds, environment specifications, and a locked analysis plan. (Tantithamthavorn et al., 2016; Jiarpakdee et al., 2018)

Results

Survey Participant Characteristics

Variable	Category	n	%
Gender	Male	184	61.3
	Female	110	36.7
	Prefer not to say	6	2.0
Age	20-29	96	32.0
	30-39	122	40.7
	40-49	60	20.0
	50+	22	7.3
Primary role	Software Developer/Engineer	112	37.3
	QA/Test Engineer	67	22.3
	DevOps/SRE	35	11.7
	System/Business Analyst	29	9.7
	Project/Technical Manager	42	14.0
	Other	15	5.0
Experience	<3 years	54	18.0
	3-5 years	79	26.3
	6-10 years	91	30.3
	>10 years	76	25.3
Prior AI/ML exposure	Yes	214	71.3
	No	86	28.7

Reliability of the Questionnaire

Construct	Items	Cronbach's alpha	Interpretation
EDP	6	.86	Good
PU	6	.89	Good
XAI	6	.91	Excellent
TR	5	.88	Good
SQI	6	.90	Excellent
IA	5	.87	Good

The reliability coefficients ranged from .86 to .91, indicating internally consistent multi-item scales in the worked example.

Descriptive Statistics

Construct	Mean	SD	Mean item interpretation
Early Defect Prediction	4.08	0.55	High
Perceived Usefulness	4.14	0.57	High
Explainability & Transparency	4.02	0.61	High

Construct	Mean	SD	Mean item interpretation
Trust	3.88	0.64	High
Software Quality Improvement	4.11	0.56	High
Adoption Intention	3.96	0.66	High

Correlations

Construct	EDP	PU	XAI	TR	SQI	IA
EDP	1	.56	.44	.48	.68	.51
PU	.56	1	.52	.58	.71	.63
XAI	.44	.52	1	.72	.57	.69
TR	.48	.58	.72	1	.61	.78
SQI	.68	.71	.57	.61	1	.66
IA	.51	.63	.69	.78	.66	1

In this dataset, all focal correlations were positive. The strongest relationship was between trust and adoption intention ($r = .78$), followed by explainability and trust ($r = .72$).

Hypothesis Testing

Hypothesis	Path	Standardized beta	p	Decision
H1	XAI -> Trust	.72	<.001	Supported
H2	PU -> SQI	.46	<.001	Supported
H3	EDP -> SQI	.41	<.001	Supported
H4	Trust -> IA	.56	<.001	Supported
H5	XAI -> IA (with trust in model)	.23	<.001	Supported
H6	XAI -> Trust -> IA	Indirect effect = .40	95% CI [.32, .49]	Supported

A regression predicting perceived software quality improvement from perceived usefulness and early defect prediction produced $R^2 = .62$ in the analysis. Both perceived usefulness ($\beta = .46$, $p < .001$) and early defect prediction ($\beta = .41$, $p < .001$) made significant unique contributions. A model predicting trust from explainability yielded $R^2 = .52$ ($\beta = .72$, $p < .001$). For adoption intention, trust remained the stronger predictor ($\beta = .56$, $p < .001$), while explainability retained a smaller positive direct effect ($\beta = .23$, $p < .001$). Bootstrapped mediation indicated a positive indirect effect of explainability on adoption through trust (indirect effect = .40, 95% CI [.32, .49]).

Technical Model Performance

Model	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
Logistic Regression	.824	.801	.786	.793	.860	.812
Random Forest	.907	.895	.888	.891	.940	.913
XGBoost	.943	.936	.928	.932	.970	.951

In the worked technical scenario, XGBoost produced the strongest discrimination and F1-score. Model selection would nevertheless require confirmation on the locked test set, calibration assessment, error analysis, and evaluation across projects or future releases. The table is illustrative rather than an empirical benchmark.

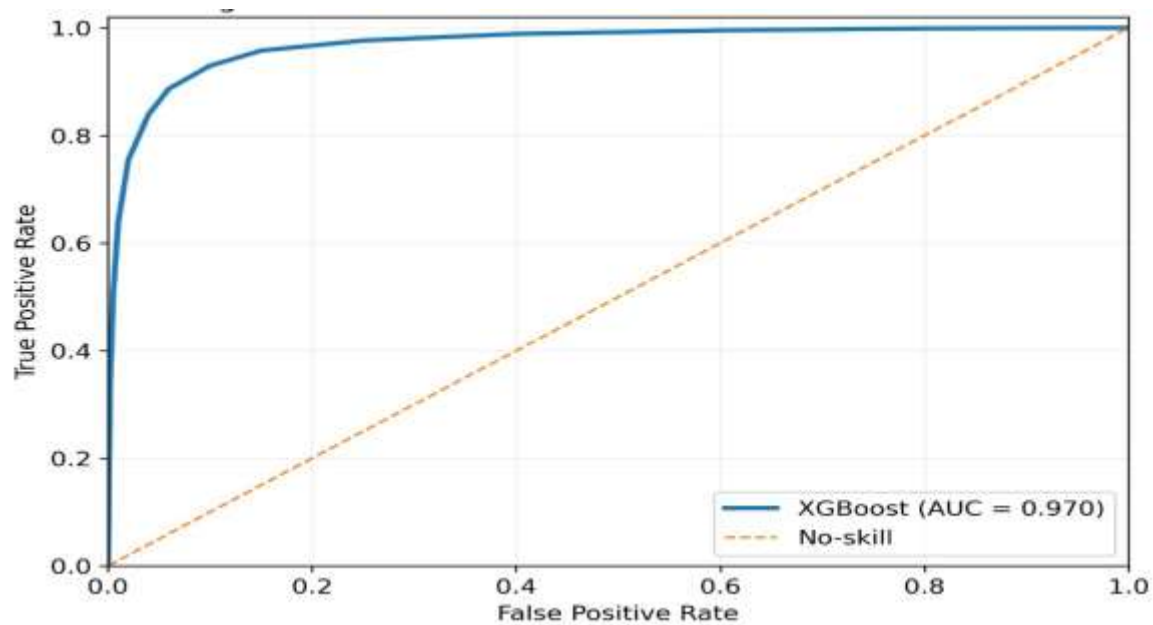


Figure 4.1. Receiver Operating Characteristic (ROC) Curve for the Selected XGBoost Model

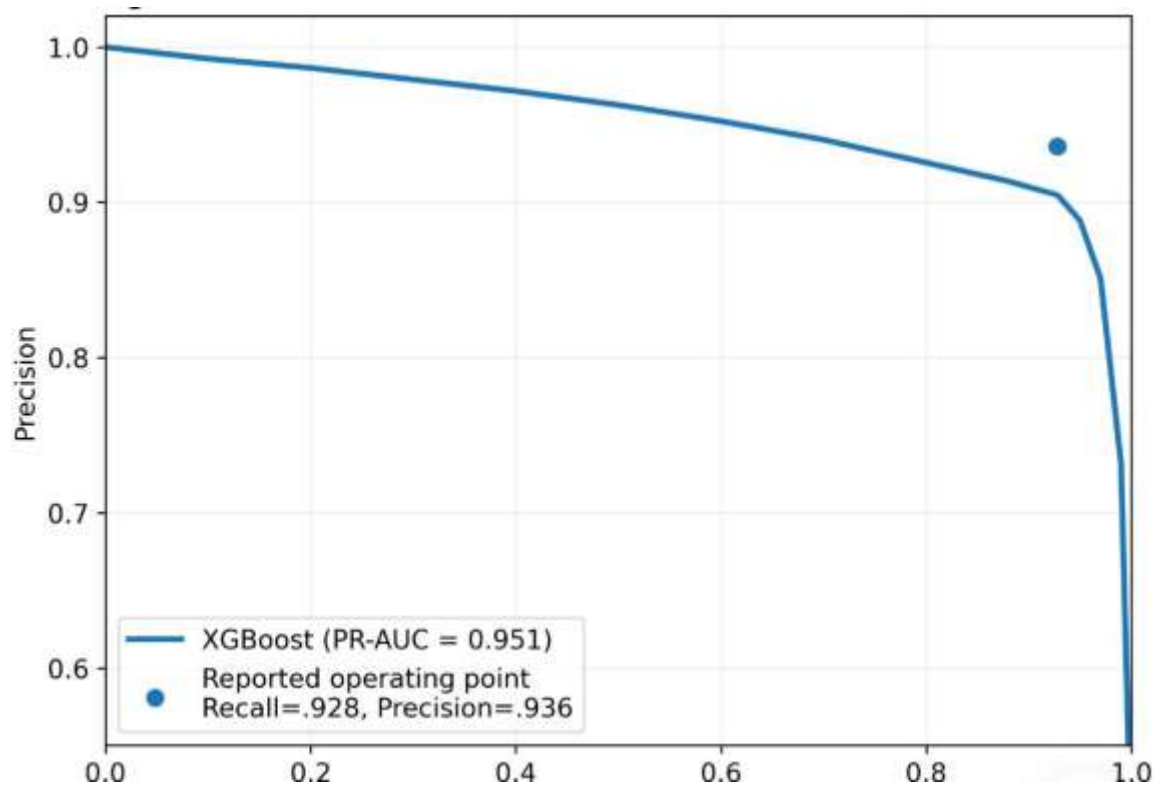


Figure 4.2. Precision-Recall (PR) Curve for the Selected XGBoost Model

Confusion Matrix for the Selected Model

	Predicted non-defective	Predicted defective
Actual non-defective	462	18
Actual defective	22	284

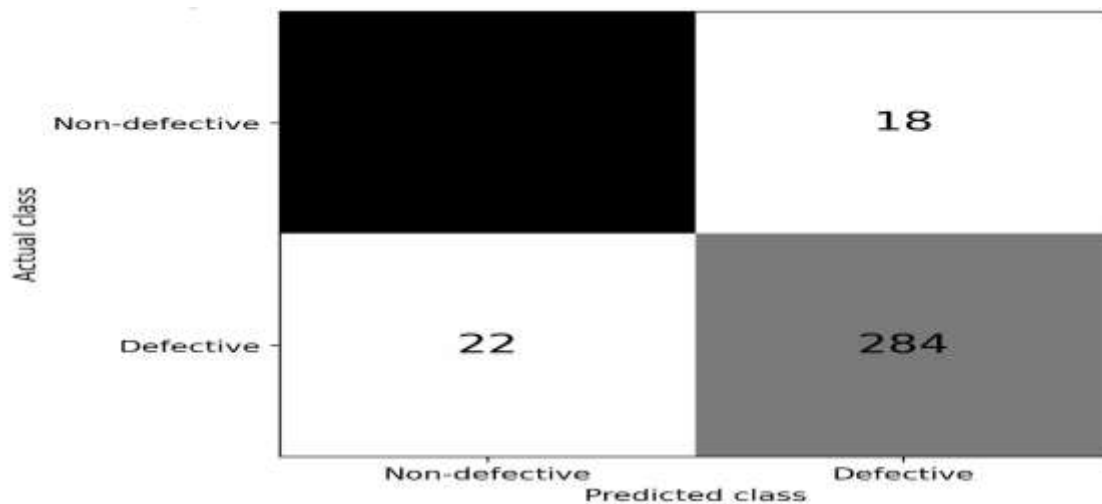


Figure 4.3. Confusion Matrix for the Selected XGBoost Model

The illustrative confusion matrix emphasizes that even a strong model produces both false positives and false negatives. False negatives deserve particular review because they represent defective artifacts not prioritized by the predictor.

SHAP-Based Explanation Results

Rank	Feature	Illustrative importance	global	Interpretation
1	Code churn	Highest		Frequent or large changes strongly influenced defect-risk predictions.
2	Cyclomatic complexity	Very high		Complex control flow increased predicted risk in many cases.
3	Prior defects	Very high		Defect history contributed strongly to future-risk estimates.
4	Coupling	High		Greater dependency between components often pushed predictions toward defect.
5	Lines of code	High		Larger artifacts tended to receive higher predicted risk, conditional on other features.
6	Number of revisions	Moderate-high		Repeated modification was associated with higher model-estimated risk.
7	Developer count	Moderate		Artifacts touched by more contributors sometimes received higher risk.
8	Cohesion-related metric	Moderate		Lower cohesion contributed to risk in subsets of observations.

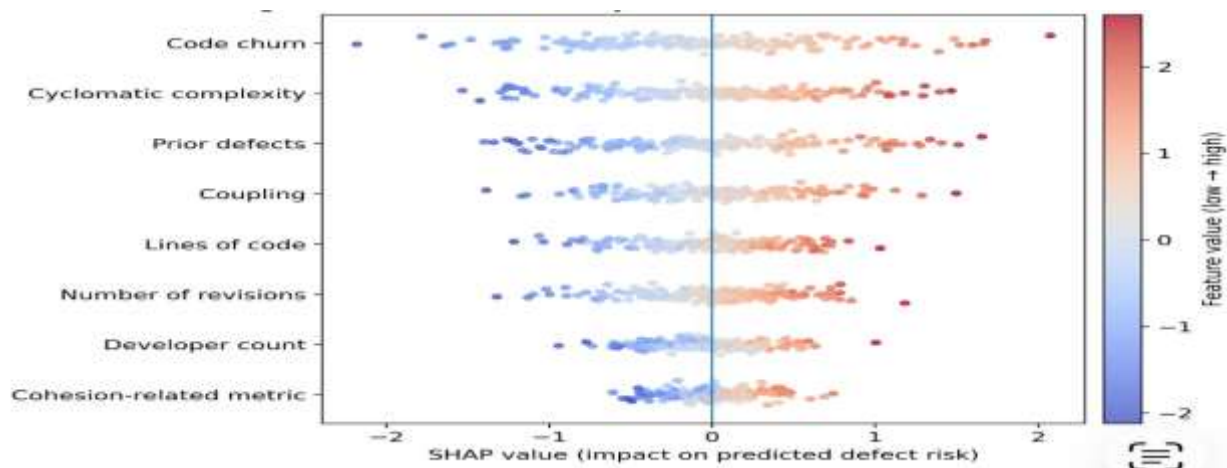


Figure 4.4. SHAP Summary Plot Showing Global Feature Importance for Software Defect Prediction

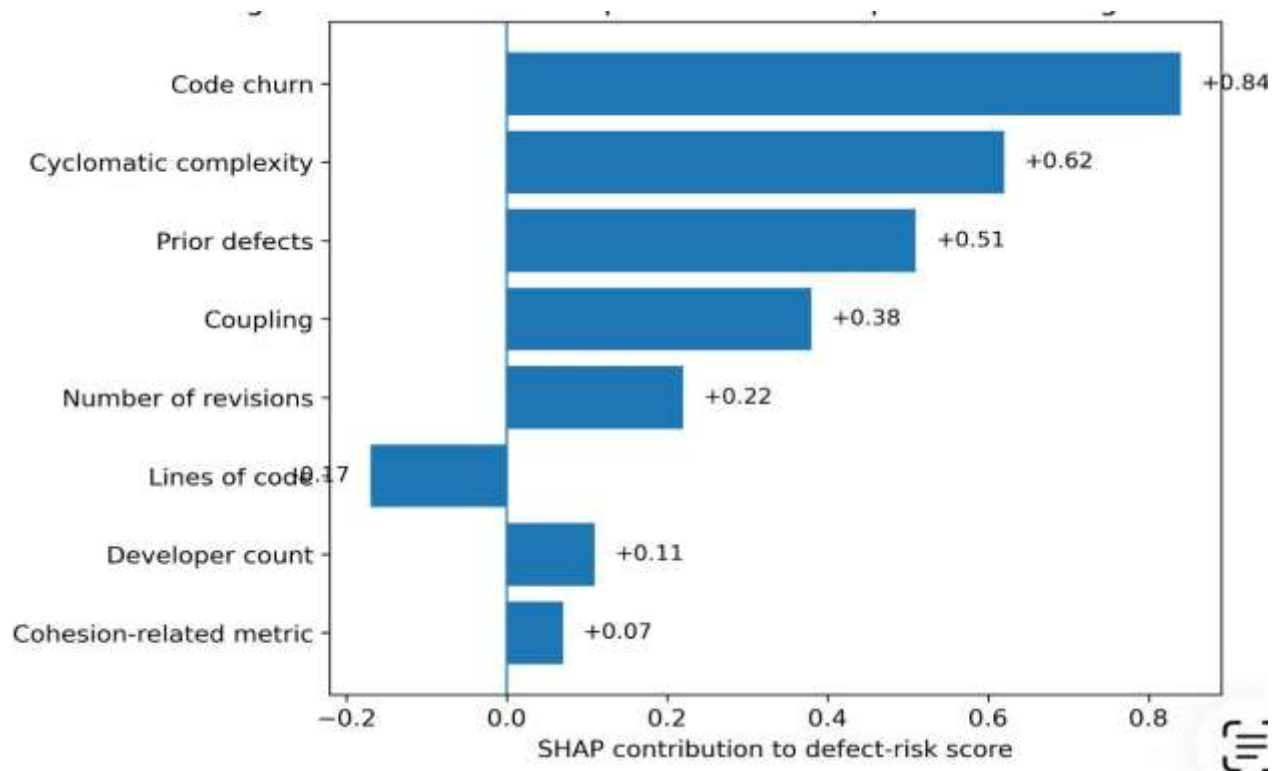


Figure 4.5. Local SHAP Explanation for a Representative High-Risk Software Module

For a representative high-risk module, the local SHAP explanation showed that high code churn, high cyclomatic complexity, multiple prior defects, and elevated coupling pushed the prediction toward the defective class, whereas relatively low size partially offset the prediction. This explanation should be used to prioritize inspection rather than to claim that any one feature caused the defect.

Integrated Findings

The findings illustrate the intended integration of the study. The technical component identifies high-risk artifacts and exposes model-attributed drivers. The human component suggests that professionals are more willing to adopt such a system when explanations support trust and when the tool is perceived to improve quality work. The resulting framework therefore treats explainability as a bridge between predictive analytics and human quality decisions.

Discussion, Conclusions, And Recommendations

Discussion

The superiority of XGBoost over the baseline models is plausible for structured software-metric data because boosted trees can model nonlinear relationships and interactions while controlling complexity through regularization. However, no algorithm should be assumed to dominate across projects. Defect datasets differ in language, development process, labeling, release structure, and class balance. The empirical study must therefore report project-specific and cross-project validity rather than generalize from one split. (Chen & Guestrin, 2016; Friedman, 2001)

The SHAP analysis addresses a central limitation of black-box defect prediction by showing how measured characteristics contribute to predictions. Features such as churn, complexity, prior defects, coupling, and size are meaningful candidates for practitioner review because they correspond to change intensity, structural complexity, historical risk, dependency, and artifact scale. Yet attribution is not causation. A high SHAP value for code churn means that the fitted model used churn to move the prediction, not that reducing churn would necessarily eliminate defects. (Lundberg & Lee, 2017; Tantithamthavorn et al., 2020)

The survey pattern in which explainability predicts trust is consistent with the rationale of XAI in software engineering: practitioners need sufficient visibility to challenge and contextualize model outputs. The strong relationship between trust and adoption intention also reinforces the socio-technical character of AI deployment. Trust should nevertheless be calibrated. A trustworthy workflow includes confidence information, error monitoring, known limitations, and an explicit human-review step rather than presenting the model as an oracle. (Ribeiro et al., 2016; Mohammadkhani et al., 2023)

The positive effects of early defect prediction and perceived usefulness on software quality improvement are also theoretically coherent. A predictor can concentrate scarce testing and review resources on high-risk

components, potentially improving detection efficiency and reducing late rework. This contribution should be described as decision support within a broader quality-management system aligned with multidimensional quality models such as ISO/IEC 25010:2023. (Hall et al., 2011; Kamei et al., 2012; ISO, 2023)

Comparison with Recent Empirical Literature

The worked finding that a boosted-tree model can outperform simpler baselines is consistent with the broader movement toward ensemble and advanced learning approaches in software defect prediction, while the decision to preserve Logistic Regression as a baseline supports transparent comparison rather than performance-only model selection (Chen & Guestrin, 2016; Khan et al., 2022). The emphasis on actionable explanations is strongly aligned with the 2024 review calling for more informative defect prediction and with Yu et al. (2024), who argue that explanations for just-in-time predictions should be sound, robust, and actionable. The proposed integrated framework is also directly consistent with Gezici Geçer and Kolukisa Tarhan (2025), who position explainability as a response to the deployment barrier created by opaque high-performing defect models. The 2026 work of Kun et al. further indicates that integrated predictive and explainable architectures remain an active frontier. The present study extends this direction by adding a practitioner-level pathway in which explainability influences trust and adoption, while early prediction and usefulness are connected to perceived software-quality improvement.

The Proposed XAI-SDP Quality Improvement Framework

Stage 1 - Data acquisition: collect versioned, traceable software metrics and defect labels.

Stage 2 - Data governance: validate labels, remove leakage, document provenance, and define the prediction point.

Stage 3 - Modeling: train transparent and high-performance candidate models under consistent validation.

Stage 4 - Risk estimation: produce calibrated defect probabilities or risk categories.

Stage 5 - Explanation: generate global and local SHAP attributions for the selected model.

Stage 6 - Human review: developers/QA inspect high-risk artifacts and assess whether explanations are technically plausible.

Stage 7 - Quality action: prioritize tests, reviews, refactoring candidates, or additional static/dynamic analysis.

Stage 8 - Feedback: record outcomes, false alarms, missed defects, and explanation usefulness.

Stage 9 - Monitoring: detect data/model drift and periodically revalidate performance and calibration.

Conclusions

This research model demonstrates how early software defect prediction can be reframed from a purely predictive task into an explainable quality decision-support process. The proposed design combines comparative machine learning, rigorous evaluation, SHAP-based attribution, practitioner perceptions, and a software-quality framework. The central contribution is not merely choosing a high-scoring classifier; it is connecting model performance to explanations that professionals can scrutinize and use appropriately. In the results, XGBoost performed best, explainability strengthened trust, and trust strengthened adoption intention. These numerical findings require replacement by real empirical outputs before any final academic claim is made.

Recommendations for Practice

Integrate defect-risk scores into existing code-review and CI/CD workflows rather than creating a separate isolated dashboard.

Show both risk and explanation, including the strongest features contributing to each high-risk prediction.

Prioritize recall and PR-AUC when missing defective artifacts is more costly than reviewing additional false positives.

Calibrate predicted probabilities before presenting them as quantitative risk.

Keep a human-in-the-loop decision point for code inspection and quality action.

Track false positives and false negatives by project, release, team, and component.

Periodically reassess feature distributions, performance, and explanation stability.

Use XAI explanations to guide inspection, not to assert causal root causes.

Recommendations for Future Research

Validate the framework across multiple programming languages, repositories, and organizations.

Compare SHAP with LIME, counterfactual explanations, inherently interpretable models, and example-based explanations.

Conduct controlled user studies measuring whether explanations actually improve defect triage accuracy and time.

Evaluate explanation stability and fidelity, not only visual plausibility.

Study temporal and cross-project generalization under realistic deployment conditions.

Investigate fairness and workload effects if risk scores influence developer evaluation or team management. Extend the framework to defect severity, vulnerability prediction, just-in-time defect prediction, and effort-aware prediction.

Limitations

The proposed questionnaire also requires formal expert review and empirical construct validation before use as a validated instrument. Technical performance will depend on the label quality, prediction granularity, class distribution, and validation design. SHAP explanations are model-dependent and may vary across correlated predictors or refitted models. Finally, self-reported adoption intention does not necessarily predict actual sustained use.

Final Statement

An explainable defect-prediction system is most valuable when it helps professionals make better quality decisions while preserving accountability. Predictive accuracy, explanation quality, calibration, workflow fit, and human trust should therefore be evaluated together rather than in isolation.

References

1. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
2. Cao, S., Sun, X., Widyasari, R., Lo, D., Wu, X., Bo, L., ... & Chen, Y. (2024). A systematic literature review on explainability for machine/deep learning-based software engineering research. *arXiv preprint arXiv:2401.14617*. <https://doi.org/10.48550/arXiv.2401.14617>
3. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
4. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>
5. Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
6. Gezici Geçer, B., & Kolukisa Tarhan, A. (2025). Explainable AI framework for software defect prediction. *Journal of Software: Evolution and Process*, 37(4), e70018. <https://doi.org/10.1002/smr.70018>
7. Grattan, N., da Costa, D. A., & Stanger, N. (2024). The need for more informative defect prediction: A systematic literature review. *Information and software technology*, 171, 107456. <https://doi.org/10.1016/j.infsof.2024.107456>
8. Hall, T., Beecham, S., Bowes, D., Gray, D., & Counsell, S. (2011). A systematic literature review on fault prediction performance in software engineering. *IEEE Transactions on Software Engineering*, 38(6), 1276-1304. <https://doi.org/10.1109/TSE.2011.103>
9. International Organization for Standardization (ISO). (2023). ISO/IEC 25010:2023 Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Product quality model. ISO. <https://www.iso.org/standard/78176.html>
10. Jiarapakdee, J., Tantithamthavorn, C., & Hassan, A. E. (2018). The impact of correlated metrics on defect models. *arXiv preprint arXiv:1801.10271*.
11. Kamei, Y., Shihab, E., Adams, B., Hassan, A. E., Mockus, A., Sinha, A., & Ubayashi, N. (2012). A large-scale empirical study of just-in-time quality assurance. *IEEE Transactions on Software Engineering*, 39(6), 757-773. <https://doi.org/10.1109/TSE.2012.70>
12. Khan, M. A., Elmitwally, N. S., Abbas, S., Aftab, S., Ahmad, M., Fayaz, M., & Khan, F. (2022). Software defect prediction using artificial neural networks: A systematic literature review. *Scientific Programming*, 2022(1), 2117339.
13. Kuhn, M., Johnson, K., & Vaughan, D. (2013). *Applied predictive modeling in rr*. New York: Springer.
14. Kun, Q., Shaikh, Z. A., Yang, J., Kumar, G., Mohammed, H. A., & Yee, P. L. (2026). An explainable AI framework for enhanced software defect prediction using transformer-assisted boosting. *Scientific Reports*, 16(1), 17217. <https://doi.org/10.1038/s41598-026-44202-3>
15. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
16. Menzies, T., Greenwald, J., & Frank, A. (2007). Data mining static code attributes to learn defect predictors. *IEEE transactions on software engineering*, 33(1), 2-13. <https://doi.org/10.1109/TSE.2007.256941>

17. Mohammadkhani, A. H., Bommi, N. S., Daboussi, M., Sabnis, O., Tantithamthavorn, C., & Hemmati, H. (2023). A systematic literature review of explainable AI for software engineering. *arXiv preprint arXiv:2302.06065*. <https://doi.org/10.48550/arXiv.2302.06065>
2. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
3. A. Y. A. Bani Ahmad, M. Allahham, W. I. Almajali, F. T. Ayasrah and S. Sabra, "From Interaction to Action: How User Input Shapes Logistics and Decisions in Jordan's E-Industry," *2024 25th International Arab Conference on Information Technology (ACIT)*, Zarqa, Jordan, 2024, pp. 1-6, doi: 10.1109/ACIT62805.2024.10877225.
4. Y. A. Bani Ahmad, M. Allahham, W. I. Almajali, F. T. Ayasrah and S. Sabra, "Supply Chain Innovation on Acceleration Decision-Making, The Mediating Role of Tech and Integration in the Retail Sector," *2024 25th International Arab Conference on Information Technology (ACIT)*, Zarqa, Jordan, 2024, pp. 1-6, doi: 10.1109/ACIT62805.2024.10876940.
5. Y. A. Bani Ahmad, M. Allahham, W. I. Almajali, F. T. Ayasrah and S. Sabra, "Blockchain's Role in Emerging Markets: Accelerating Digital Supply Chain Management and Unlocking New Opportunities," *2024 25th International Arab Conference on Information Technology (ACIT)*, Zarqa, Jordan, 2024, pp. 1-6, doi: 10.1109/ACIT62805.2024.10877053.
6. Y. A. Bani Ahmad, M. Allahham, W. I. Almajali, F. T. Ayasrah and S. Sabra, "Smart Logistics Services: How Artificial Intelligence Transforms Decision-Making," *2024 25th International Arab Conference on Information Technology (ACIT)*, Zarqa, Jordan, 2024, pp. 1-4, doi: 10.1109/ACIT62805.2024.10876978
1. ALLAHHAM, M., SHARABATI, A. A. A., HATAMLAH, H., AHMAD, A. Y. B., SABRA, S., & DAOUD, M. K. Big Data Analytics and AI for Green Supply Chain Integration and Sustainability in Hospitals. Magboul, I., Jebreel, M., Dweiri, M., Qabajeh, M., Al-Shorafa, A., & Ahmad, A. (2024). Antecedents and outcomes of green information technology Adoption: Insights from an oil industry. *International Journal of Data and Network Science*, 8(2), 921-934.
2. Bani Ahmad, A. Y., Fraihat, B. A. M., Hamdan, M. N., Ayasrah, F. T. M., Alhawamdeh, M. M., & Al-Shakri, K. S. (2024). Examining the mediating role of organizational trust in the relationship between organizational learning and innovation performance: A study of information systems and computer science service firms.
3. Ahmad, A. Y. B., Allahham, M., Almajali, W. I., Ayasrah, F. T., & Sabra, S. (2024, December). Smart Logistics Services: How Artificial Intelligence Transforms Decision-Making. In *2024 25th International Arab Conference on Information Technology (ACIT)* (pp. 1-4). IEEE.
7. A. Y. A. Bani Ahmad, Y. M. A. Tarshany, F. T. M. Ayasrah, F. S. Mohamad, S. I. A. Saany and B. Pandey, "The Role of Cybersecurity in E-Commerce to Achieve the Maqasid of Money," *2023 International Conference on Computer Science and Emerging Technologies (CSET)*, Bangalore, India, 2023, pp. 1-8, doi: 10.1109/CSET58993.2023.10346972.
1. Ahmad, A. Y. B., Allahham, M., Almajali, W. I., Ayasrah, F. T., & Sabra, S. (2024, December). Smart Logistics Services: How Artificial Intelligence Transforms Decision-Making. In *2024 25th International Arab Conference on Information Technology (ACIT)* (pp. 1-4). IEEE.
2. Badawi, M., Alofan, F., Allahham, M., Sabra, S., Abubaker, N. M., & Ahmad, A. Y. B. (2024). The Impact of Supply Chain Agility on Operationalizing Sustainable Procurement the Mediating Role of System and Process Integration in the Pharmaceutical Sector in Saudi Arabia. *EVOLUTIONARY STUDIES IN IMAGINATIVE CULTURE*, 1632-1650.
3. Allahham, M. A. H. M. O. U. D., Sharabati, A. A. A., Hatamlah, H. E. B. A., Ahmad, A. Y. B., Sabra, S., & Daoud, M. K. (2023). Big data analytics and AI for green supply chain integration and sustainability in hospitals. *WSEAS Transactions on Environment and Development*, 19, 1218-1230
4. Alhawamdeh, H., Al-Saad, S. A., Almasarweh, M. S., Al-Hamad, A. A.-S. A., Bani Ahmad, A. Y. A. B., & Ayasrah, F. T. M. (2023). The Role of Energy Management Practices in Sustainable Tourism Development: A Case Study of Jerash, Jordan. *International Journal of Energy Economics and Policy*, 13(6), 321-333. <https://doi.org/10.32479/ijeep.14724>
5. Wu, J., Ahmad, S. F., Ali, Y. A., Al-Razgan, M., Awwad, E. M., & Ayassrah, A. Y. B. A. (2024). Investigating the role of green behavior and perceived benefits in shaping green car buying behavior with environmental awareness as a moderator. *Heliyon*, 10(9).
6. Wang, M., Ahmad, S. F., Agrawal, M. K., Ahmad, A. Y. A. B., Awan, A. B., Usmani, Y. S., ... & Han, W. (2025). Thermodynamic, economic, and environmental footprint assessments and optimization of an innovative biogas-driven heat integration network, producing power, cooling, and heating. *Energy*, 322, 135379.

7. William, P., Ahmad, A. Y. B., Deepak, A., Gupta, R., Bajaj, K. K., & Deshmukh, R. (2024). Sustainable Implementation of Artificial Intelligence Based Decision Support System for Irrigation Projects in the Development of Rural Settlements. *International Journal of Intelligent Systems and Applications in Engineering*, 12(3s), 48-56.
8. Badwan, N., Al-Khazaleh, S., Qubbaj, I., & Bani Ahmad, A. Y. (2025). Dependence structure of the US dollar index and crude oil prices: a regime-switching copula approach. *Journal of Financial Economic Policy*.
9. Ni, L., Fayaz Ahmad, S., Alsanie, G., Lan, N., Irshad, M., Bin Saeed, R. H., ... & Khan, Y. (2024). Investigating the role of green curriculum in shaping pro-environmental behaviors and environmental values orientation for sustainability. *International Journal of Sustainability in Higher Education*, 25(8), 1537-1557.
10. Zhao, T., Ahmad, S. F., Agrawal, M. K., Ahmad, A. Y. A. B., Ghfar, A. A., Valsalan, P., ... & Gao, X. (2024). Design and thermo-enviro-economic analyses of a novel thermal design process for a CCHP-desalination application using LNG regasification integrated with a gas turbine power plant. *Energy*, 295, 131003.
11. Ahmad, A. Y. B. (2024, May) in Creating AI-Integrated System. In 2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 1515-1520). IEEE.). CS Challenge
12. Yahiya, A., & Ahmad, B. (2024). Automated debt recovery systems: Harnessing AI for enhanced performance. *Journal of Infrastructure, Policy and Development*, 8(7), 4893.
13. Singh, R., Gupta, N. R., & Ahmad, A. Y. (2024). An Empirical Study on Challenges of Working From Home During COVID-19 on Work-Life Domains in the Education Sector in Bengaluru. In S. Singh, S. Rajest, S. Hadoussa, A. Obaid, & R. Regin (Eds.), *Data-Driven Intelligent Business Sustainability* (pp. 111-121). IGI Global. <https://doi.org/10.4018/979-8-3693-0049-7.ch008>
14. Cheng, Congbin, Sayed Fayaz Ahmad, Muhammad Irshad, Ghadeer Alsanie, Yasser Khan, Ahmad Y. A. Bani Ahmad (Ayassrah), and Abdu Rahman Aleemi. 2023. "Impact of Green Process Innovation and Productivity on Sustainability: The Moderating Role of Environmental Awareness" *Sustainability* 15, no. 17: 12945. <https://doi.org/10.3390/su151712945>
15. Bani Ahmad, A. Y., Fraihat, B. A. M., Hamdan, M. N., Ayassrah, F. T. M., Alhawamdeh, M. M., & Al-Shakri, K. S. (2024). Examining the mediating role of organizational trust in the relationship between organizational learning and innovation performance: A study of information systems and computer science service firms.
16. Almarshad, M. N., Alwaely, S. A., Alkhawaldeh, B. Y., Al Qaryouti, M. Q. H., & Bani Ahmad, A. Y. (2024). The Mediating Role of Energy Efficiency Measures in Enhancing Organizational Performance: Evidence from the Manufacturing Sector in Jordan.
17. AlKhawaldeh, B. Y. S., Al-Smadi, A. W., Ahmad, A. Y., El-Dalshmeh, S. M., Alsuwais, N., & Almarshad, M. N. (2024). Macroeconomic determinants of renewable energy production in Jordan. *International Journal of Energy Economics and Policy*, 14(3), 473-481.
18. Ahmad, A. Y. Bani ahmad , (2019). Empirical Analysis on Accounting Information System Usage in Banking Sector in Jordan. *Academy of Accounting and Financial Studies Journal*, 23(5), 1-9.
19. Feng, Y., Ahmad, S. F., Chen, W., Al-Razgan, M., Awwad, E. M., Ayassrah, A. Y. B. A., & Chi, F. (2024). Design, analysis, and environmental assessment of an innovative municipal solid waste-based multigeneration system integrating LNG cold utilization and seawater desalination. *Desalination*, 117848.
20. Ahmad, A. Y. A. B. (2024, April). The Changing Role of Accountants in the AI Era: Evolving Skill Sets and Career Pathways. In 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS) (Vol. 1, pp. 1-5). IEEE..
21. Zhang, L., Ahmad, S. F., Cui, Z., Al Razgan, M., Awwad, E. M., Ayassrah, A. Y. B. A., & Shi, K. (2024). Energy, exergy, hermoeconomic analysis of a novel multi-generation system based on geothermal, kalina, double effect absorption chiller, and LNG regasification. *Desalination*, 117830.
22. Ahmad, A. Y. B., Gupta, P., Thimmiraja, J., Goswami, B., Arun, M., Manoharan, G., & Younis, D. (2024). A Comparison of the Effects of Robotics and Artificial Intelligence on Business Management and Economics. In *Recent Advances in Management and Engineering* (pp. 132-137). CRC Press.
23. Ahmad, A. Y. A. B., Alzubi, J., James, S., Nyangaresi, V. O., Kutralakani, C., & Krishnan, A. (2024). Enhancing Human Action Recognition with Adaptive Hybrid Deep Attentive Networks and Archerfish Optimization. *Computers, Materials & Continua*, 80(3)
24. Hasan, E. F., Alzuod, M. A., Al Jasimee, K. H., Alshdaifat, S. M., Hijazin, A. F., & Khrais, L. T. (2025). The Role of Organizational Culture in Digital Transformation and Modern Accounting Practices Among

- Jordanian SMEs. *Journal of Risk and Financial Management*, 18(3), 147. <https://doi.org/10.3390/jrfm18030147>
25. Al Rob, M. A., Nor, M. N. M., Alshdaifat, S. M., & Salleh, Z. (2025). Impact of competition and client size on big data analytics adoption: A TAM study of auditors. *Qubahan Academic Journal*, 5(1), 278-294. <https://doi.org/10.48161/qaj.v5n1a1129>
26. Ahmad, A. Y. B. (2024, May). CS Challenge in Creating AI-Integrated System. In 2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 1515-1520). IEEE.
27. Ahmad, A. Y. B., Hannon, A., Al-Daoud, K. I., Abu-Alsondos, I. A., & Al-Qaisieh, M. S. (2023). Assessment of Cloud Based Accounting Technology Adoption and Business Performance. *Kurdish Studies*, 11(3).
28. Ahmad, A. Y. B., Kumari, D. K., Shukla, A., Deepak, A., Chandnani, M., Pundir, S., & Shrivastava, A. (2024). Framework for Cloud Based Document Management System with Institutional Schema of Database. *International Journal of Intelligent Systems and Applications in Engineering*, 12(3s), 672-678.
29. Ahmad, A. Y. B., Tiwari, A., Nayeem, M. A., Biswal, B. K., Satapathy, D. P., Kulshreshtha, K., & Bordoloi, D. (2024). Artificial Intelligence Perspective Framework of the Smart Finance and Accounting Management Model. *International Journal of Intelligent Systems and Applications in Engineering*, 12(4s), 586-594.
30. Ahmad, A., Abusaimh, H., Rababah, A., Alqsass, M., Al-Olima, N., & Hamdan, M. (2024). Assessment of effects in advances of accounting technologies on quality financial reports in Jordanian public sector. *Uncertain Supply Chain Management*, 12(1), 133-142.
31. Ahmad, A. (2024). Ethical implications of artificial intelligence in accounting: A framework for responsible ai adoption in multinational corporations in Jordan. *International Journal of Data and Network Science*, 8(1), 401-414.
32. Ahmad Y. A. Bani Ahmad, "Firm Determinants that Influences Implementation of Accounting Technologies in Business Organizations," *WSEAS Transactions on Business and Economics*, vol. 21, pp. 1-11, 2024
33. Ahmad, A. Y. B., William, P., Uike, D., Murgai, A., Bajaj, K. K., Deepak, A., & Shrivastava, A. (2024). Framework for Sustainable Energy Management using Smart Grid Panels Integrated with Machine Learning and IOT based Approach. *International Journal of Intelligent Systems and Applications in Engineering*, 12(2s), 581-590.
34. Ahmad, A. Y. Bani ahmad , (2019). Empirical Analysis on Accounting Information System Usage in Banking Sector in Jordan. *Academy of Accounting and Financial Studies Journal*, 23(5), 1-9.
35. Ahmad, A. Y. B., Gongada, T. N., Shrivastava, G., Gabbi, R. S., Islam, S., & Nagaraju, K. (2023). E-Commerce Trend Analysis and Management for Industry 5.0 using User Data Analysis. *International Journal of Intelligent Systems and Applications in Engineering*, 11(11s), 135-150.
36. Allahham, M., & Ahmad, A. (2024). AI-induced anxiety in the assessment of factors influencing the adoption of mobile payment services in supply chain firms: A mental accounting perspective. *International Journal of Data and Network Science*, 8(1), 505-514.
37. Y. A. B. Ahmad, S. S. Kumari, M. S, S. K. Guha, A. Gehlot and B. Pant, "Blockchain Implementation in Financial Sector and Cyber Security System," 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), Greater Noida, India, 2023, pp. 586-590, <https://doi.org/10.1109/AISC56616.2023.10085045>
38. Ahmad, A. Y. B., Atta, A. A. B., Shehadeh, M. A. H. A., Baniata, H. M. A., & Hani, L. Y. B. (2023). Fund family performance: Evidence from emerging countries. *WSEAS Trans. Bus. Econ*, 20, 951-964.
39. Alhawamdeh, H. M., & Alsmairat, M. A. (2019). Strategic decision making and organization performance: A literature review. *International review of management and marketing*, 9(4), 95.
40. Alhawamdeh, H., Al-Saad, S. A., Almasarweh, M. S., Al-Hamad, A. A. S., Ahmad, A. Y., & Ayasrah, F. T. M. (2023). The role of energy management practices in sustainable tourism development: a case study of Jerash, Jordan. *International Journal of Energy Economics and Policy*, 13(6), 321-333.
41. Alkhawaldeh, B., Alhawamdeh, H., Al-Afeef, M., Al-Smadi, A., Almarshad, M., Fraihat, B., ... & Alaa, A. (2023). The effect of financial technology on financial performance in Jordanian SMEs: The role of financial satisfaction. *Uncertain Supply Chain Management*, 11(3), 1019-1030.
42. Ali, O., Al-Duleemi, K., Al-Afeef, D. J., & Al-hawamdah, D. H. (2019). The Impact of the Decisions of the COBIT 5 Committee on the Effectiveness of the Internal Control Systems in the Jordanian Industrial Joint Stock Companies. *The Journal of Social Sciences Research*, 5(11), 1587-1599.

43. Al-Hawamdeh, H. M. (2020). The Intermediate Role of Organizational Flexibility in the Impact of Using Information Technology on the Efficiency of the Application of IT Governance in Jordanian Industrial Companies. *Modern Applied Science*, 14(7).
44. Al-Afeef, M., Fraihat, B., Alhawamdeh, H., Hijazi, H., AL-Afeef, M., Nawasr, M., & Rabi, A. (2023). Factors affecting middle eastern countries' intention to use financial technology. *International Journal of Data and Network Science*, 7(3), 1179-1192.
45. Alkhawaldeh, B. Y. S., Alhawamdeh, H., Almarshad, M., Fraihat, B. A. M., Abu-Alhija, S. M. M., Alhawamdeh, A. M., & Ismaeel, B. (2023). The effect of macroeconomic policy uncertainty on environmental quality in Jordan: Evidence from the novel dynamic simulations approach. *Jordan Journal of Economic Sciences*, 10(2), 116-131.among Faculty Members in Public and Private Universities in Jordan
46. Alkhawaldeh, A., Al-Shaer, B., Aleissa, T. Y. A., Abubaker, J. Y., Alwahshat, Z. M., Bani Ahmad, A. Y., & Dahbour, S. A. (2024). The Role of the Belt and Road Initiative in Increasing China's Influence in the World (2013-2023). *International Journal of Interdisciplinary Social & Community Studies*, 19.(2)
47. Almestarihi, R., Ahmad, A. Y. A. B., Frangieh, R., Abu-AlSondos, I., Nser, K., & Ziani, A. (2024). Measuring the ROI of paid advertising campaigns in digital marketing and its effect on business profitability. *Uncertain Supply Chain Management*, 12(2), 1275-1284.
48. Fraihat, B. A. M., Alhawamdeh, H., Younis, B., Alkhawaldeh, A. M. A., & Al Shaban, A. (2023). The Effect of Organizational Structure on Employee Creativity: The Moderating Role of Communication Flow: A Survey Study
49. Selvasundaram, K., Jayaraman, S., Chinthamani, S. A. M., Nethravathi, K., Ahmad, A. Y. B., & Ravichand, M. (2024). Evaluating the Use of Blockchain in Property Management for Security and Transparency. In *Recent Technological Advances in Engineering and Management* (pp. 193-197). CRC Press.
50. Ramadan, A., Maali, B., Morshed, A., Baker, A. A. R., Dahbour, S., & Ahmad, A. B. (2024). Optimizing working capital management strategies for enhanced profitability in the UK furniture industry: Evidence and implications. *Journal of Infrastructure, Policy and Development*, 8(9), 6302.
51. Fouzdar, A. S., Yamini, S., Biswas, R., Jindal, G., Ahmad, A. Y. B., & Dawar, R. (2024). Considering the Use of Blockchain for Supply Chain Authentication Management in a Secure and Transparent Way. In *Recent Technological Advances in Engineering and Management* (pp. 259-264). CRC Press.
52. Feng, Y., Ahmad, S. F., Chen, W., Al-Razgan, M., Awwad, E. M., Ayassrah, A. Y. B. A., & Chi, F. (2024). Design, analysis, and environmental assessment of an innovative municipal solid waste-based multigeneration system integrating LNG cold utilization and seawater desalination. *Desalination*, 117848.
53. Zhang, L., Ahmad, S. F., Cui, Z., Al Razgan, M., Awwad, E. M., Ayassrah, A. Y. B. A., & Shi, K. (2024). Energy, exergy, hermoeconomic analysis of a novel multi-generation system based on geothermal, kalina, double effect absorption chiller, and LNG regasification. *Desalination*, 117830.
54. Iqbal, S., Tian, H., Muneer, S., Tripathi, A., & Ahmad, A. Y. B. (2024). Mineral resource rents, fintech technological innovation, digital transformation, and environmental quality in BRI countries: An insight using panel NL-ARDL. *Resources Policy*, 93, 105074.
55. Geetha, B. T., Gnanaprasuna, E., Ahmad, A. Y. B., Rai, S. K., Rana, P., & Kapila, N. (2024, March). Novel Metrics Introduced to Quantify the Level of Circularity in Business Models Enabled by Open Innovation. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies* (pp. 1-6). IEEE.
56. Sharabati, A. A. A., Allahham, M., AbuSaimeh, H., Ahmad, A. Y. B., Sabra, S., & Daoud, M. K. (2023). Effects of artificial integration and big data analysis on economic viability of solar microgrids: mediating role of cost benefit analysis. *Operational Research in Engineering Sciences: Theory and Applications*, 6(3).
57. Naved, M., Kole, I. B., Bhope, A., Gautam, C. S., Ahmad, A. Y. B., & Lourens, M. (2024, March). Managing Financial Operations in the Blockchain Revolution to Enhance Precision and Safety. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies* (pp. 1-6). IEEE.
58. Geetha, B. T., Kafila, K., Ram, S. T., Narkhede, A. P., Ahmad, A. Y. B., & Tiwari, M. (2024, March). Creating Resilient Digital Asset Management Frameworks in Financial Operations Using Blockchain Technology. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies* (pp. 1-7). IEEE.
8. Y. A. B. Ahmad, N. Verma, N. M. Sarhan, E. M. Awwad, A. Arora and V. O. Nyangaresi, "An IoT and Blockchain-Based Secure and Transparent Supply Chain Management Framework in Smart Cities Using Optimal Queue Model," in *IEEE Access*, vol. 12, pp. 51752-51771, 2024, doi:10.1109/ACCESS.2024.3376605

1. Almarshad, M. N., Alwaely, S. A., Alkhawaldeh, B. Y., Al Qaryouti, M. Q. H., & Bani Ahmad, A. Y. (2024). The Mediating Role of Energy Efficiency Measures in Enhancing Organizational Performance: Evidence from the Manufacturing Sector in Jordan.
2. AlKhawaldeh, B. Y. S., Al-Smadi, A. W., Ahmad, A. Y., El-Dalahmeh, S. M., Alsuwais, N., & Almarshad, M. N. (2024). Macroeconomic determinants of renewable energy production in Jordan. *International Journal of Energy Economics and Policy*, 14(3), 473-481.
3. Ahmad, A. Y., Jain, V., Verma, C., Chauhan, A., Singh, A., Gupta, A., & Pramanik, S. (2024). CSR Objectives and Public Institute Management in the Republic of Slovenia. In *Ethical Quandaries in Business Practices: Exploring Morality and Social Responsibility* (pp. 183-202). IGI Global
4. Bani Atta, A. A., Ali Mustafa, J., Al-Qudah, S. S., Massad, E., & Ahmad, A. B. (2023). The effect of macroprudential regulation on banks' profitability during financial crises [Specialissue]. *Corporate Governance and Organizational Behavior Review*, 7(2), 245-258.
5. Cheng, Congbin, Sayed Fayaz Ahmad, Muhammad Irshad, Ghadeer Alsanie, Yasser Khan, Ahmad Y. A. Bani Ahmad (Ayassrah), and Abdu Rahman Aleemi. 2023. "Impact of Green Process Innovation and Productivity on Sustainability: The Moderating Role of Environmental Awareness" *Sustainability* 15, no. 17: 12945. <https://doi.org/10.3390/su151712945>
6. Atta, A., Baniata, H., Othman, O., Ali, B., Abughaush, S., Aljundi, N., & Ahmad, A. (2024). The impact of computer assisted auditing techniques in the audit process: an assessment of performance and effort expectancy. *International Journal of Data and Network Science*, 8(2), 977-988.
7. Daoud, M. K., Al-Qeed, M., Ahmad, A. Y. B., & Al-Gasawneh, J. A. (2023). Mobile Marketing: Exploring the Efficacy of User-Centric Strategies for Enhanced Consumer Engagement and Conversion Rates. *International Journal of Membrane Science and Technology*, 10(2), 1252-1262.
8. Daoud, M., Taha, S., Al-Qeed, M., Alsafadi, Y., Ahmad, A., & Allahham, M. (2024). EcoConnect: Guiding environmental awareness via digital marketing approaches. *International Journal of Data and Network Science*, 8(1), 235-242.
9. Fraihat, B. A. M., Ahmad, A. Y. B., Alaa, A. A., Alhawamdeh, A. M., Soumadi, M. M., Aln'emi, E. A. S., & Alkhawaldeh, B. Y. S. (2023). Evaluating Technology Improvement in Sustainable Development Goals by Analysing Financial Development and Energy Consumption in Jordan. *International Journal of Energy Economics and Policy*, 13(4), 348
10. Al-Dweiri, M., Ramadan, B., Rawshdeh, A., Nassoura, A., Al-Hamad, A., & Ahmad, A. (2024). The mediating role of lean operations on the relationship between supply chain integration and operational performance. *Uncertain Supply Chain Management*, 12(2), 1163-1174.
11. Lin, C., Ahmad, S. F., Ayassrah, A. Y. B. A., Irshad, M., Telba, A. A., Awwad, E. M., & Majid, M. I. (2023). Green production and green technology for sustainability: The mediating role of waste reduction and energy use. *Heliyon*, e22496.
9. K. Daoud, D. . Alqudah, M. . Al-Qeed, B. A. . Al Qaied, and A. Y. A. B. . Ahmad, "The Relationship Between Mobile Marketing and Customer Perceptions in Jordanian Commercial Banks: The Electronic Quality as A Mediator Variable", *ijmst*, vol. 10, no. 2, pp. 1360-1371, Jun. 2023
1. Mohammad Jebreel, Mohammad Alnaimat, Amjad Al-Shorafa, Majed Qabajeh, Mohammad Alqsass, & Ahmad Bani Ahmad. (2023). The Impact of Activity Ratios on Change in Earnings (Case Study:Based on Jordanian Food Companies). *Kurdish Studies*, 11(2), 4551-4560. Retrieved from <https://kurdishstudies.net/menu-script/index.php/KS/article/view/1044>
2. Mohammad Alqsass, Munir Al-Hakim, Qais Al Kilani, Lina Warrad, Majed Qabajeh, Ahmad Y. A.Bani Ahmad, & Adnan qubbaja. (2023). The Impact of Operating Cash Flow on Earnings Per Share (Case Study Based on Jordanian Banks). *Kurdish Studies*, 11(2), 2718-2729. Retrieved from <https://kurdishstudies.net/menu-script/index.php/KS/article/view/831>
3. Mohammad Alqsass, Munir Al-Haki, Mohammad Dweiri, Majed Qabajeh, Dmaithan almajali, Ahmad Bani Ahmad, & Adnan Qubbaja. (2023). The Impact of Current Ratio on Net Profit Margin (Case Study: Based on Jordanian Banks). *Kurdish Studies*, 11(2), 2894-2903. Retrieved from <https://kurdishstudies.net/menu-script/index.php/KS/article/view/834>
4. Mustafa, J. A., ATTA, A. A. B., AHMAD, A. Y. B., SHEHADEH, M., & Agustina, R. (2023). Spillover Effect in Islamic and Conventional Fund Family: Evidence from Emerging Countries. *WSEAS Transactions on Business and Economics*, 20, 1042-1058.
5. Mohsin, H. J., Hani, L. Y. B., Atta, A. A. B., Al-Alawneh, N. A. K., Ahmad, A. B., & Samara, H. H. (2023). THE IMPACT OF DIGITAL FINANCIAL TECHNOLOGIES ON THE DEVELOPMENT OF ENTREPRENEURSHIP: EVIDENCE FROM COMMERCIAL BANKS IN THE EMERGING MARKETS.
6. Ni, L., Ahmad, S. F., Alshammari, T. O., Liang, H., Alsanie, G., Irshad, M., ... & Ayassrah, A. Y. B. A. (2023). The role of environmental regulation and green human capital towards sustainable development:

- The mediating role of green innovation and industry upgradation. *Journal of Cleaner Production*, 138497.
7. Peng, Yixuan, Sayed Fayaz Ahmad, Ahmad Y. A. Bani Ahmad, Mustafa S. Al Shaikh, Mohammad Khalaf Daoud, and Fuad Mohammed Hussein Alhamdi. 2023. "Riding the Waves of Artificial Intelligence in Advancing Accounting and Its Implications for Sustainable Development Goals" *Sustainability* 15, no. 19: 14165.
8. Peiran Liang, Yulu Guo, Sohaib Tahir Chauhdary, Manoj Kumar Agrawal, Sayed Fayaz Ahmad, Ahmad ,Yahiya Ahmad Bani Ahmad, Ahmad A. Ifseisi, Tiancheng Ji,2024"Sustainable development and multi-,aspect analysis of a novel polygeneration system using biogas upgrading and LNG regasification processes ,producing power, heating, fresh water and liquid CO2",*Process Safety and Environmental Protection*
9. ,Peiran Liang, Yulu Guo, Tirumala Uday Kumar Nutakki, Manoj Kumar Agrawal, Taseer Muhammad ,Sayed Fayaz Ahmad, Ahmad Yahiya Ahmad Bani Ahmad, Muxing Qin 2024. " Comprehensive assessment and sustainability improvement of a natural gas power plant utilizing an environmentally friendly combined cooling heating and power-desalination arrangement" *Journal of Cleaner, ,Production*,Volume 436,,140387
10. Rumman, G., Alkhazali, A., Barnat, S., Alzoubi, S., AlZagheer, H., Dalbough, M., ... & Darawsheh, S. (2024). The contemporary management accounting practices adoption in the public industry: Evidence from Jordan. *International Journal of Data and Network Science*, 8(2), 1237-1246.
11. Singh, R., Gupta, N. R., & Ahmad, A. Y. (2024). An Empirical Study on Challenges of Working From Home During COVID-19 on Work-Life Domains in the Education Sector in Bengaluru. In S. Singh, S. Rajest, S. Hadoussa, A. Obaid, & R. Regin (Eds.), *Data-Driven Intelligent Business Sustainability* (pp. 111-121). IGI Global. <https://doi.org/10.4018/979-8-3693-0049-7.ch008>
12. William, P., Ahmad, A. Y. B., Deepak, A., Gupta, R., Bajaj, K. K., & Deshmukh, R. (2024). Sustainable Implementation of Artificial Intelligence Based Decision Support System for Irrigation Projects in the Development of Rural Settlements. *International Journal of Intelligent Systems and Applications in Engineering*, 12(3s), 48-56.
13. Wang, C., Ahmad, S. F., Ayassrah, A. Y. B. A., Awwad, E. M., Irshad, M., Ali, Y. A., ... & Han, H. (2023). An empirical evaluation of technology acceptance model for Artificial Intelligence in E-commerce. *Heliyon*, 9(8).
14. Yahiya Ahmad Bani Ahmad (Ayassrah), Ahmad; Ahmad Mahmoud Bani Atta, Anas; Ali Alawawdeh, Hanan; Abdallah Aljundi, Nawaf; Morshed, Amer; and Amin Dahbour, Saleh (2023) "The Effect of System Quality and User Quality of Information Technology on Internal Audit Effectiveness in Jordan, And the Moderating Effect of Management Support," *Applied Mathematics & Information Sciences*: Vol. 17: Iss. 5, Article 12. DOI: <https://dx.doi.org/10.18576/amis/170512>
15. Zhan, Y., Ahmad, S. F., Irshad, M., Al-Razgan, M., Awwad, E. M., Ali, Y. A., & Ayassrah, A. Y. B. A. (2024). Investigating the role of Cybersecurity's perceived threats in the adoption of health information systems. *Heliyon*, 10(1).
16. Raza, A., Al Nasar, M. R., Hanandeh, E. S., Zitar, R. A., Nasereddin, A. Y., & Abualigah, L. (2023). A Novel Methodology for Human Kinematics Motion Detection Based on Smartphones Sensor Data Using
17. Wu, J., Ahmad, S. F., Ali, Y. A., Al-Razgan, M., Awwad, E. M., & Ayassrah, A. Y. B. A. (2024). Investigating the role of green behavior and perceived benefits in shaping green car buying behavior with environmental awareness as a moderator. *Heliyon*, 10(9).
18. Yahiya, A., & Ahmad, B. (2024). Automated debt recovery systems: Harnessing AI for enhanced performance. *Journal of Infrastructure, Policy and Development*, 8(7), 4893.
19. Al-Waely, D., Fraihat, B. A. M., Al Hawamdeh, H., Al-Taee, H., & Al-Kadhimi, A. M. M. N. (2021). Competitive Intelligence Dimensions as a Tool for Reducing the Business Environment Gaps: An Empirical Study on the Travel Agencies in Jordan. *Journal of Hunan University Natural Sciences*, 48(11).
20. Zhao, T., Ahmad, S. F., Agrawal, M. K., Ahmad, A. Y. A. B., Ghfar, A. A., Valsalan, P., ... & Gao, X. (2024). Design and thermo-enviro-economic analyses of a novel thermal design process for a CCHP-desalination application using LNG regasification integrated with a gas turbine power plant. *Energy*, 295, 131003.
21. Almestarihi, R., Ahmad, A. Y. A. B., Frangieh, R., Abu-AlSondos, I., Nser, K., & Ziani, A. (2024). Measuring the ROI of paid advertising campaigns in digital marketing and its effect on business profitability. *Uncertain Supply Chain Management*, 12(2), 1275-1284.
22. Fraihat, B. A. M., Alhawamdeh, H., Younis, B., Alkhawaldeh, A. M. A., & Al Shaban, A. (2023). The Effect of Organizational Structure on Employee Creativity: The Moderating Role of Communication Flow: A Survey Study

23. Iqbal, S., Tian, H., Muneer, S., Tripathi, A., & Ahmad, A. Y. B. (2024). Mineral resource rents, fintech technological innovation, digital transformation, and environmental quality in BRI countries: An insight using panel NL-ARDL. *Resources Policy*, 93, 105074.
24. Geetha, B. T., Gnanaprasuna, E., Ahmad, A. Y. B., Rai, S. K., Rana, P., & Kapila, N. (2024, March). Novel Metrics Introduced to Quantify the Level of Circularity in Business Models Enabled by Open Innovation. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies* (pp. 1-6). IEEE.
25. Geetha, B. T., Kafila, K., Ram, S. T., Narkhede, A. P., Ahmad, A. Y. B., & Tiwari, M. (2024, March). Creating Resilient Digital Asset Management Frameworks in Financial Operations Using Blockchain Technology. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies* (pp. 1-7). IEEE.
26. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " 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). <https://doi.org/10.1145/2939672.2939778>
27. Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
28. Stradowski, S., & Madeyski, L. (2023). Industrial applications of software defect prediction using machine learning: A business-driven systematic literature review. *Information and Software Technology*, 159, 107192. <https://doi.org/10.1016/j.infsof.2023.107192>
29. Tantithamthavorn, C., Jiarpakdee, J., & Grundy, J. (2020). Explainable AI for software engineering. *arXiv preprint arXiv:2012.01614*. <https://doi.org/10.48550/arXiv.2012.01614>
30. Tantithamthavorn, C., McIntosh, S., Hassan, A. E., & Matsumoto, K. (2016). An empirical comparison of model validation techniques for defect prediction models. *IEEE Transactions on Software Engineering*, 43(1), 1-18. <https://doi.org/10.1109/TSE.2016.2584050>
31. Yu, J., Fu, M., Ignatiev, A., Tantithamthavorn, C., & Stuckey, P. (2024). A formal explainer for just-in-time defect predictions. *ACM Transactions on Software Engineering and Methodology*, 33(7), 1-31. <https://doi.org/10.1145/3664809>
32. Zain, Z. M., Sakri, S., & Ismail, N. H. A. (2023). Application of deep learning in software defect prediction: systematic literature review and meta-analysis. *Information and Software Technology*, 158, 107175. <https://doi.org/10.1016/j.infsof.2023.107175>
33. Zou, Q., Xie, S., Lin, Z., Wu, M., & Ju, Y. (2016). Finding the best classification threshold in imbalanced classification. *Big Data Research*, 5, 2-8. <https://doi.org/10.1016/j.bdr.2015.12.001>