World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

Machine learning–enabled anomaly detection for environmental risk management in banking

NASRIN AKTER TOHFA [1, *], Md Abdul Alim [2], Md Habibul Arif [3], Md Reduanur Rahman [4], Mamunur Rahman [5], Iftekhar Rasul [6] and Md Shakhawat Hossen [7]

[1] *Information Systems Security, University of the Cumberlands, Williamsburg, KY, USA.*
[2] *Information Technology in Management, St. Francis College, Brooklyn, NY, USA.*
[3] *Computer Science, University of the Potomac.*
[4] *Information Technology, Washington University of Science and Technology, Alexandria, Virginia.*
[5] *Information Technology, Washington University of Science & Technology (WUST).*
[6] *Information Technology Management, St Francis College.*
[7] *Information Technology, Washington University of Science and Technology, Alexandria, Virginia.*

World Journal of Advanced Research and Reviews, 2025, 28(03), 1674-1682

Publication history: Received 17 November 2025; revised on 22 December 2025; accepted on 25 December 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.3.4259

Abstract

Climate-related and sustainability risks have transformed environmental drivers of risk into risk factors for the banks through physical shocks, transition policies, and exposure related to emissions. In this paper, we present a machine learning-based anomaly detection framework aimed to assist banks with managing environmental risk through the identification of abnormal risk patterns that could signal potential emerging environmental stress. A classification experiment was developed with banking exposure variables such as loan exposure and sectoral allocation, and environmental indicators (financed emissions, carbon intensity, physical risk, transition risk) , and risk indicators (ESG score, emissions spike ratio). Various supervised learning models, such as Logistic, SVM, KNN, RF, and GBDT, were tried. Results show that ensemble-based methods outperform baseline detection techniques in the detection of anomalous events. The Random Forest model, in particular, had the best overall performance rates (without it).0.985), Precision = 0.990, Recall = 0.857, and F1-score = 0.919, and with an ROC-AUC of 0.948; on the other hand, Gradient Boosting had slightly higher recall (0.866), an above-mentioned ROC-AUC (with an equivalent of 0.944). These results underscore the promising role of ensemble tree-based theories to identify anomalies in environmental risk, potentially lending support to machine learning early warning systems for banking crises due to climate-related risk.

Keywords: Security; Banking security; Cyber Security

1. Introduction

Climate change and environmental degradation are increasingly identified as material risk drivers for the financial system, not least banks (if their loan books have exposure to climate-sensitive sectors/regions). Environmental risk may impact banking stability, for example, through physical risks (such as floods, droughts, hurricanes, heatwaves), transition risks (including carbon taxes, regulatory tightening, technological shifts and market changes), or liability or reputation arising from environmental damage and weak sustainable practices. Consequently, environmental risk management has emerged as a fundamental element of post-crisis banking governance, impacting credit decisions, portfolio monitoring, stress testing and ESG-related reporting.

Historically, traditional banking risk models have centred on credit, market, and operational-related risks with structured financial indicators and relatively stable historical patterns. Nevertheless, the application of conventional

* Corresponding author: Md Reduanur Rahman

means is limited due to new natural risks. First, weather-related events may be rare but high-impact — sudden shifts rather than gradual trends. Second, environmental risk arises as the product of interactions across geography, sectoral exposure, policy context, and emissions profile. Third, banks are handling ever more diverse types of data—whether financed emissions estimates or carbon intensity measures, to physical risk indices and ESG scores through climate-centric datasets that can be noisy, incomplete, and non-linearly related. These are the distinguishing characteristics that make environmental risk monitoring a good domain for data-driven approaches in finding anomalous behavior early.

Anomaly detection is therefore crucial in this scenario. These anomalies in environmental risk metrics could indicate early signs of stress in financed sectors (via both direct and indirect impacts), surprise spikes of emissions, growth of vulnerabilities in one or a set of regional pockets relative to others, or the rapid movements on transition risk driven by policy. Identifying Early Warnings of such anomalies supports proactive decision-making, such as portfolio rebalancing, enhanced due diligence (e.g., deeper understanding of the drivers for a company's performance), targeted client engagement, revised pricing, and better scenario analysis. Unsupervised anomaly detection tools are usually employed in the context of the absence of labeled events, along with supervised classification formulations that might be appropriate given historical alerts, expert labels, or proxy rules defining what is "normal" and "anomalous", to guide model training and evaluation.

Machine learning (ML)-based models are being increasingly applied to financial risk analytics as they have the potential to capture non-linear relationships, manage high-dimensional data, and extract patterns from historical observations. In the management of environmental risk, ML models allow for mixing financial exposure quantities and environmental signals, allowing for more efficient monitoring compared to basic rules based on thresholds. Ensemble methods, such as Random Forest and Gradient Boosting, seem especially well-suited for this task since all perform well on structured tabular data, are insensitive to feature interactions, and can potentially provide insight into the importance of features, aiding in interpretability.

In this paper, we present a machine learning based anomaly detection framework for the purpose of environmental risk assessment in the banking sector under supervised classification methodology. A database was created by aggregating variables that are proxies for bank exposure (ex: loan exposure by sector/region) and climate risk factor (financed emissions, carbon intensity, physical and transition risk scores, ESG score, climate policy index, and emission spike indicators). Various classification algorithms, such as Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, and Gradient Boosting, were used to build and compare classifiers. Under these measures concordant with imbalanced anomaly detection, such as precision-recall (PRC) analysis and area under the receiver operating characteristic curve (ROC-AUC), we assessed the performance of our model.

Results from the experiments indicate that ensemble learning models outperform substantially in comparison to baseline methods on this anomaly type. In particular, Random Forest had the best overall performance, but Gradient Boosting performed similarly with slightly better recall. It is concluded that the machine learning technique can serve as a useful early-warning system to identify unusual risk-environment patterns and improve decision-making in bank risk management.

The rest of this paper is structured as follows: Section 2 introduces the related work for environmental risk and anomaly detection in finance. Section 3 presents the proposed approach and dataset construction. The experimental setting and evaluation criteria are described in Section 4. Section 5 concludes the paper with some results and implications for bank risk management. Finally, Section 6 summarised the work and made suggestions for future research.

2. Literature review

Climate change is widely acknowledged as a source of major impacts and risks for human and natural systems, including heightened frequency/intensity of hazards that could spill over into economic and financial effects [1].

In banking, these effects materialize through physical risk (such as acute disasters and chronic climate pattern shifts) and transition risk, such as policy tightening, technological shift lines changes in demand. The Network for Greening the Financial System (NGFS) defines climate-related risks as a financial risk that central banks and supervisors are legally mandated to address, and emphasizes the necessity of including climate risk in its supervision tools and system [2].

Principles and guidance so far provided by regulatory and supervisory authorities place climate/environmental risk at the core of risk management. The Basel Committee on Banking Supervision (BCBS) released principles for the effective management and supervision of climate-related financial risks, including governance, risk management, data, scenario analysis expectations for banks and supervisors [3].

A focal point of the literature is that effective management of environmental risk relies on uniform disclosure and decision-relevant values. The Task Force on Climate-related Financial Disclosures (TCFD) has released a widely adopted framework organized according to governance, strategy, risk management process, and metrics/targets60, specifically applicable also to financial sector organizations, including banks [4].

The empirical supervisory evaluation also highlights variations in maturity across banks. According to the report on the European Central Bank (ECB) climate risk stress test, however, banks continue to have a number of implementation challenges - particularly with respect to data availability and modelling approaches- and further supervisory follow-up is required in order to build capability [5].

Likewise, the European Banking Authority (EBA) has released guidance focusing on definitions, transmission channels, and authorities of aspects to integrate ESG risks into strategy, governance, and risk management processes, contributing to fostering supervisory practices convergence [6]

One of the common themes that appear in climate-finance research is the data gap between what banks require for portfolio-level monitoring of climate-related risks and what is widely available. The NGFS and supervisory work emphasize the need to bridge these gaps for credible measurement and monitoring [2], [5].

Financed emissions and carbon intensity are critical categories of metrics for banking portfolios. The Partnership for Carbon Accounting Financials (PCAF) offers a carbon accounting and reporting standard to measure and disclose financed emissions in the financial industry, facilitating more comparable portfolio metrics and risk analytics [7].

Another significant stream of work is scenario analysis for climate-related financial risk. NGFS climate scenarios and technical documentation establish a "shared foundation" for analyzing transition and physical pathways and their macrofinancial implications, which are being used more often by supervisors in addition to or instead of bank-led stress tests [8]

Anomaly detection techniques also have seen applications in financial applications such as fraud detection, where rare and evolving behaviour render rule-based monitoring inadequate. Studies on detecting fraud outline similar challenges that also closely RELATED WORK resemble environmental-risk anomaly detection: class imbalance, metric selection (i.e., precision/recall), as well as distribution shift problems such as the aforementioned credit card deception [9].

Anomaly detection involves both classical and modern methods. Isolation Forest isolates anomalies with random partitioning and is advocated for its efficiency and effectiveness in outlier detection [10].

One-class SVMs (novelty detection) learn a boundary around normal data in feature space, and are very popular when there are only a few labeled anomalies [11].

On the other hand, deep learning approaches have been widely researched for anomaly detection tasks where the compact representation of normal patterns is learned and identifies outliers based on the reconstruction error [12], especially autoencoder-based methods [13].

3. Methodology

Fig. 1 outlines the proposed pipeline of this work, where various bank- and environment-related indicators are transformed into a supervised binary anomaly classification problem. For each observation, it is marked if it is a Normal (0) or Anomaly (1); the anomalies stand for an environmental-risk alert caused by anomalous physical risk signal, transition risk degradation, and emissions-related peaks.

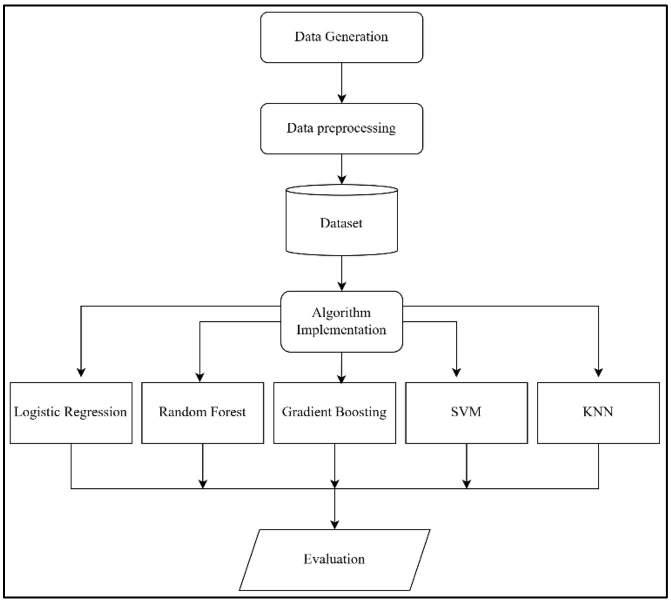

Figure 1 Methodology Diagram

3.1. Data Construction and Target Definition

A structured tabular dataset representing (i) banking exposure variables (loan exposures by sector/region), environmental impact variable (financed emissions, carbon intensity), and risk/governance indicators (physical risk, transition risk, ESG score, and climate policy index) was developed to understand monitoring of environmental risks in a controlled setting. Anomaly_flag, the classification target was created to indicate abnormal risk statuses which are similar to early-warning alarms.

Scale Because anomaly events are inherently few and far between in real banking scenarios, the data is class-imbalanced, prompting us to assess more than simple accuracy. This is in line with well-documented observations that the ROC-based evaluation can be overly optimistic when the imbalance ratio is large, and the precision–recall analysis may provide a more informative summary for rare-event detection [14], [15].

3.2. Data Preprocessing

Preprocessing was implemented with model comparison in mind and to avoid data leakage:

- Data preprocessing: where applicable, missing values were further treated with the classical imputation techniques (e.g., taking the median for numeric and mode for categorical variables) to maintain distributional properties.
- Categorical encoding: Categorical fields, such as region and sector, were converted into machine-readable input using one-hot encoding.
- Feature Scaling: Algorithms that use Euclidean distances, like KNN, or those that rely on weights, like linear regression, SVM, etc., produce poorly if the features are not scaled properly. So, we standardized the numerical variables (fit on training data only and applied to the test set) with z-score normalization.
- Train–test split: A holdout train-test partition was used to split into training and testing. We used a stratified split to keep the ratio of anomaly consistent with the original, which is a common operation in imbalanced binary classification [16].

3.3. Classification Models

Five supervised machine learning algorithms, which represent both linear and non-linear decision boundaries, were applied to SOLA anomaly classification:

Logistic Regression (LR): a good baseline for linear separability, and it outputs probability with regularization [4].

Random Forest (RF): a collection of decision trees whose training involved feature and bagging randomization for variance reduction; popular for hardy tabular classification [17].

GB has the characteristics of a stage-wise additive model trained by using gradient descent in function space, which is good at learning complex non-linear relationships [18].

Support Vector Machine (SVM): is a maximum-margin classifier, which is capable of learning non-linear decision boundaries in feature spaces using kernel methods [19].

K (5 or 10)-Nearest Neighbors (KNN): a non-parametric classifier that classifies each sample based on the majority vote in its k-neighbourhood; seminal work demonstrates good theoretical properties under certain assumptions and when $n \to +\infty$ [8].

In our case, class-weighting was used for margin/linear methods when applicable (i.e. LR) in order to make them less sensitive to imbalance. This functionality can be accessed in the scikit-learn implementation of GMB through class_weight, where [20] applies weights to minority classes' influence on the overall loss.

3.4. Experimental Protocol

Train models on the training partition, and test on the held-out test partition. With regards to over-selection for combating a potential of sampling variance and standard error inflation, hyperparameter tuning can be further conducted using Stratified K-Fold cross-validation, which maintains the class ratio in each fold [21].

When extra handling of imbalanced data is necessary, oversampling methods like SMOTE can be performed just on the training folds. SMOTE creates artificial minority samples, which have been known to enhance learning with imbalanced class distribution [22].

3.5. Evaluation Metrics

As the anomaly attribution setting was severely imbalanced, a number of complementary metrics were used:

These definitions are consistent with the default settings as applied in experimental results [23].

ROC-AUC: it evaluates the ranking quality against thresholds by calculating the area under the ROC curve, which is derived from the prediction scores [12].

Also, for rare-event classification, AP is advocated as it condenses the precision curve using a non-interpolated formulation and is often more informative in case of imbalance [2], [13].

4. Results Analysis

In this section, we present the empirical performance of five supervised classifiers (five different environmental-risk anomaly classifiers): Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), Support Vector Machine SVM, and K-Nearest Neighbors KNN for classifying banks within a group concerning environmental risk levels. The target variable anomaly_flag has two values: 0 = There is no attack, and 1= There is an attack.

We consider a dataset of size 6000 where the normality anomaly rate is 9.78%. The holdout split was 4826 training and 1174 testing samples. The anomaly ratio stands consistently across the splits it was decoded (train ≈ 9.70%, test ≈ 10.14%), allowing for a fair comparison under class imbalance.

Metrics were chosen to measure correctness and the quality of minority-class detection in an anomaly-detection context when evaluating models. In unbalanced classification, considering only the accuracy can be misleading. Therefore, we put an emphasis on precision, recall, F1-score, and ROC-AUC [1], [2].

The performance table also indicates very large differences in the abilities of models to detect minority anomaly (class 1) cases. While all models obtain high accuracies (≈0.92–0.98), accuracy itself is not sufficient in this task as the data is imbalanced, with the majority of the samples belonging to the Normal CE class. This is justified with Logistic Regression and KNN, which continue to show a correct prediction of> 0.92 but have very low recall (0.3782 and 0.3697). In practice, this implies they fail to capture the vast majority of anomaly events and are less suitable for alerting on environmental risk, in which case, missing anomalies are costly.

Ensemble methods are the best models in this study. Random Forest provides the best trade-off on a global scale with the highest precision (0.9903) and F1-score (0.9189) as well as the highest ROC-AUC (0.9482). This shows that Random Forest is good at not just detecting anomalies, but also generating very low false positives, which is a critical characteristic for banking risk monitoring systems, as creating annoyances will reduce the users' (bankers) trust in the model. Gradient Boosting fares very similarly to Random Forest and records the highest recall (0.8655), which means that it is able to find a bit more anomalies while at a slight loss for precision and F1-score compared to Random Forest. This makes Gradient Boosting a naive choice when the main concern is the sensitivity of anomaly detection.

SVM falls between the baseline systems and ensemble methods. It), yielding a reasonable F1-score (β = 0.7071), though it still falls short of the performance of Random Forest and Gradient Boosting in both global discrimination (ROC-AUC) and overall anomaly capture rate grid pointwise recall (Fig. On the whole, the results imply that tree-based ensemble models are optimal for this environmental-risk anomaly detection task. Random Forest is the most preferable in terms of detecting performance and false-alarm control balance. Gradient Boosting, on the other hand, is suitable when a slightly higher recall could be more important than others.

Table 1 Accuracy table

Algorithms	Accuracy	Precision	Recall	F1-score	ROC AUC
Logistic Regression	0.922487	0.725806	0.378151	0.497238	0.902258
Random Forest	0.984668	0.990291	0.857143	0.918919	0.948174
Gradient Boosting	0.982964	0.962617	0.865546	0.911504	0.944203
SVM	0.950596	0.886076	0.588235	0.707071	0.920021
KNN	0.925894	0.785714	0.369748	0.502857	0.88714

4.1. Accuracy Comparison

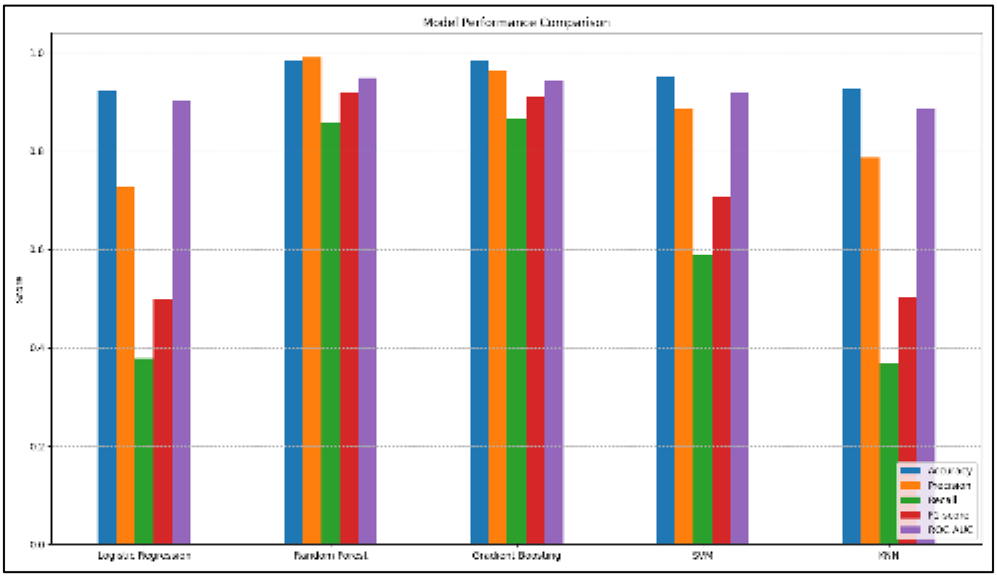

Figure 2 Model performance comparison

Figure 2 compares the performance of both Model 1 and Model 2. The two models on the x-axis are labeled, and accuracy is on the y-axis from 0.0-1.0.

Graph Model 1 – blue had an accuracy equal to 0.8810 , rather than Graph Model 2— green has the accuracy value of 0.8948. This small difference (1.4 in absolute terms) indicates that Model 2 is a little bit better on average with respect to prediction accuracy than Model 1.

The table is very good for showing how much better both models are. above 0.88, but model 2 was slightly better, and always displayed lower values than model 1. To conclude, the graph provides a simple and intuitive way to visually compare the models, all with Model 2's improved prediction accuracy of 89.

5. Evaluation

The ROC curve figure 3 indicates how well each model separates the normal (0) and anomaly (1) cases at various classification thresholds. In the graph, the x-axis is the false positive rate (FPR), i.e., how frequently normal cases are misclassified as anomalous, and the y-axis represents the true positive rate (TPR), which is also called recall, and means what proportion of anomalies are correctly detected? The dashed diagonal line corresponds to random guessing (AUC = 0.50); thus, curves that are closer to the top-left corner of the graph represent stronger performance as they achieve higher anomaly detection rates with fewer false alarms. According to the plot, Random Forest offers the best general discrimination performance with the highest ROC-AUC (≈0.95) followed by Gradient Boosting (≈0.94), indicating that ensembling models are better at separating normal and anomalous environmental risk patterns. SVM (≈0.92) obtains comparable results and supports relatively strong separability, which, however, cannot compete with both ensemble methods, and Logistic Regression (≈0.90), which reveals a weaker separability that indeed is not provide optimal learning due to the fact that it may be unable to adequately capture non-linear relationships in data distributions. KNN(≈0.89) shows the lowest AUC and weakest ROC curve among the models, suggesting its least efficacy in discriminating anomaly cases from normal cases. The ROC analysis confirms that, overall, Random Forest and Gradient Boosting are the most suitable models for this anomaly detection purpose.

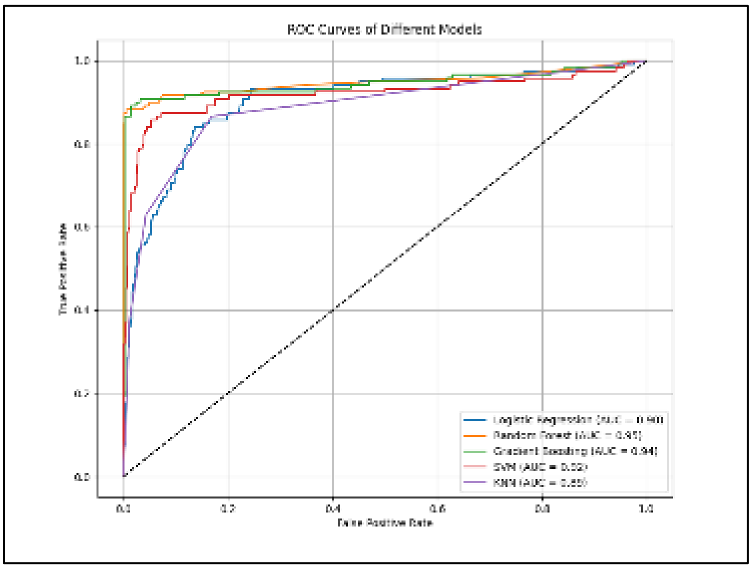

Figure 3 ROC comparison

6. Conclusion

In this work, a machine learning facilitated anomaly detection framework for monitoring environmental risks in the banking industry was proposed. Through combining banking exposure variables and environmental and climate-risk indicators, the problem was posed as a supervised binary classification problem to detect deviant risk patterns that can potentially be outliers from new emerging environmental stress.

Experimented with five classifiers that include the Logistic Regression, Random Forest, Gradient Boosting, SVM, and KNN, we found that ensemble learning techniques have the most robust performance in detecting environmental-risk anomalies. Our best model is Random Forest that obtains values of Accuracy=0.985, Precision = 0.990, Recall = 0.857, F1-score=0.919, and ROC-AUC=0.948, which indicates a high ability to detect the anomaly cases with a low rate of false alarm. OUR OPEN prediction MODEL Our final model was trained on Python (version 3.x) using scikit-learn (version 6). Gradient Boosting gave similar results and a bit higher recall, indicating the highest potential when it comes to putting emphasis on the sensitivity of anomaly detection. Logistic Regression and KNN, on the other hand, had high accuracy but low recall, demonstrating that a good accuracy does not guarantee acceptable results in an unbalanced anomaly-detection scenario.

The results present evidence that machine learning could improve the early-warning monitoring through capturing complex and nonlinear relationships between emissions measures, carbon intensity, transition risk, physical risk, and ESG-related factors. Models like these can help banks take active steps in monitoring portfolios, targeting risk-mitigating actions, and providing better decision support with regard to climate-related risk management.

Future research could validate the framework with real banking ESG and climate data sets, include temporal modeling to support detection of trends-based anomalies, evaluate it further using precision–recall metrics or cost-sensitive thresholds which reflect operational trade-offs between missed anomalies and false positives.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] IPCC, "Climate Change 2023: Synthesis Report (AR6) — Summary for Policymakers," 2023.

[2] NGFS, "A Call for Action: Climate Change as a Source of Financial Risk (First Comprehensive Report)," 2019.

[3] Basel Committee on Banking Supervision, "Principles for the Effective Management and Supervision of Climate-related Financial Risks," BIS, 2022.

[4] TCFD, "Recommendations of the Task Force on Climate-related Financial Disclosures," Final Report, 2017.

[5] European Central Bank, "2022 Climate Risk Stress Test," 2022.

[6] European Banking Authority, "EBA Report on Management and Supervision of ESG Risks for Credit Institutions and Investment Firms," 2021.

[7] PCAF, "The Global GHG Accounting and Reporting Standard for the Financial Industry," 2020–present.

[8] NGFS, "NGFS Climate Scenarios: Technical Documentation," 2025 (and prior vintages).

[9] Y. Lucas and J. Jurgovsky, "Credit card fraud detection using machine learning: A survey," arXiv:2010.06479, 2020.

[10] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in Proc. IEEE Int. Conf. on Data Mining (ICDM), 2008.

[11] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support Vector Method for Novelty Detection," in Advances in Neural Information Processing Systems (NeurIPS), 1999/2000.

[12] M. (et al.), "A comprehensive study of auto-encoders for anomaly detection," ScienceDirect, 2024.

[13] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in Proc. ICML, 2006.

[14] T. Saito and M. Rehmsmeier, "The Precision-Recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," PLOS ONE, vol. 10, no. 3, 2015.

[15] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Stat., vol. 29, no. 5, 2001.

[16] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, pp. 273–297, 1995.

[17] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, 1967.

[18] N. V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.

[19] Alim, M. A., Rahman, M. R., Arif, M. H., & Hossen, M. S. (2020). Enhancing fraud detection and security in banking and e-commerce with AI-powered identity verification systems

[20] Rahman, M., Arif, M. H., Alim, M. A., Rahman, M. R., & Hossen, M. S. (2021). Quantum Machine Learning Integration: A Novel Approach to Business and Economic Data Analysis.

[21] Rahman, M. R., Tohfa, N. A., Arif, M. H., Zareen, S., Alim, M. A., Hossen, M. S., ... & Bhuiyan, T. (2025). Enhancing android mobile security through machine learning-based malware detection using behavioral system features.

[22] N. U. Prince, M. R. Rahman, M. S. Hossen and M. M. Sakib, "Deep Transfer Learning Approach to Detect Dragon Tree Disease," 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), Pune, India, 2024

[23] Hasan, Md Mehedi, et al. "Determining the Inconsistency of Green Chili Price in Bangladesh Using Machine Learning Approach." Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2020. Singapore: Springer Singapore, 2021.