



AN INTEGRATED AI AND GIS FRAMEWORK FOR TRACKING GOVERNMENT FUNDS FROM BUDGET ALLOCATION TO GROUND-LEVEL PROJECT COMPLETION

Pavan S. Badgujar¹, Assoc. Prof. Dinesh D. Puri²

Research Scholar, Department of Computer Applications, Shramsadhana Collage of Engineering & Technology,

Jalgaon, Maharashtra, India¹

Associate Professor, Department of Computer Applications, Shramsadhana Collage of Engineering & Technology,

Jalgaon, Maharashtra, India²

Abstract: Government development projects generate financial, procurement, physical-progress, and geospatial records at different stages of implementation. Because these records are usually maintained in separate systems, it is difficult to judge whether expenditure reported for a project is reasonably consistent with the work reported on the ground. This paper presents and evaluates an Artificial Intelligence (AI) and Geographic Information System (GIS) framework that brings these records together through a common project record. A controlled synthetic benchmark of 3,000 project records was prepared because confidential government data were not available. Fifteen percent of the records contained four deliberately injected anomaly types: financial-physical mismatch, overspending, payment delay, and procurement irregularity. Isolation Forest, Local Outlier Factor (LOF), and a rule-based baseline were compared over five random seeds using precision, recall, F1-score, ROC-AUC, PR-AUC, and Precision@K. LOF gave the highest scores in the benchmark (F1 = 0.985; ROC-AUC = 0.9999), followed by the rule-based method (F1 = 0.874; ROC-AUC = 0.982) and Isolation Forest (F1 = 0.573; ROC-AUC = 0.890). For Isolation Forest, Precision@50 was 0.680 against a 0.150 no-skill baseline. A separate 30-seed Wilcoxon signed-rank test showed a consistent difference between the rule-based method and Isolation Forest ($p = 1.86 \times 10^{-9}$). SHAP analysis highlighted tender variance and the financial-physical gap among the strongest contributors to Isolation Forest scores. A Moran's I test on randomly assigned coordinates found no significant spatial clustering ($I = -0.016$, $p = 0.067$). These results show that the proposed analysis pipeline can be tested in a controlled setting, while the near-perfect LOF result should be treated as a best-case outcome because the simulated anomalies were relatively easy to separate. The framework is therefore intended as an early-warning and decision-support tool, not as an automated means of establishing corruption.

Keywords: Artificial Intelligence, Geographic Information System, Government Funds, E-Governance, Anomaly Detection, Explainable AI, Public Procurement, Project Monitoring, Design Science Research

INTRODUCTION

1.1 Background

Government development projects are funded and implemented to provide infrastructure, public services, and other benefits to citizens. In India, a project normally passes through budget allocation, administrative approval, procurement, fund release, payments, execution, inspection, and completion. Earlier work on digital financial reforms shows that better information flow can strengthen public-fund accountability, while electronic procurement can improve some aspects of infrastructure delivery [1], [2].

Government systems also produce information at different stages of this process. The Public Financial Management System (PFMS) provides facilities for fund-flow, transaction, payment, and expenditure monitoring. GeoMGNREGA, meanwhile, shows how GIS, mobile technology, and geotagged evidence can support monitoring of public assets [3], [4]. These sources are useful, but they are heterogeneous and are not organized around a common project record.



1.2 Research Problem

The main problem considered in this research is the difficulty of reconciling government expenditure with the physical progress and ground-level evidence of an individual development project. A project may show substantial expenditure while the reported physical progress remains comparatively low. Such a difference does not, by itself, mean fraud or misuse; there may be legitimate operational reasons. The practical requirement is therefore to identify records that deserve closer examination rather than to make an automatic allegation.

Recent studies show that machine-learning methods can support early warning in public procurement and can identify unusual expenditure patterns [5]-[7]. Other work has used computer vision, BIM, deep learning, and satellite data to measure or detect physical project changes [8]-[10]. These studies are useful, but they usually address one part of the project lifecycle. A more complete monitoring view would bring together allocation, fund release, expenditure, procurement, contractor information, physical progress, location, and supporting evidence for the same project.

1.3 Research Gap

The literature already provides strong evidence for several individual capabilities: digital financial management for accountability [1], e-procurement for infrastructure delivery [2], financial monitoring through PFMS [3], geospatial monitoring through GeoMGNREGA [4], procurement-risk analytics through machine learning [5]-[7], and automated physical-progress monitoring using AI and geospatial methods [8]-[10]. The unresolved issue is how these information layers can be reconciled at the level of an individual government development project.

The gap addressed here is therefore an integration gap rather than an algorithmic novelty gap. The proposed contribution is a project-level architecture that connects financial, procurement, physical-progress, GIS/evidence, and explainable AI risk layers and uses the combined information for early identification of possible financial-physical mismatches. The paper does not claim a new anomaly-detection algorithm or a first use of AI in government monitoring.

1.4 Scope and Objectives

The study covers budget allocation, fund release, procurement, contractor information, payments and expenditure, reported physical progress, GIS-based evidence, and relevant citizen-generated evidence. AI-based anomaly detection is used to rank unusual project records, while an explainable layer indicates the factors associated with a flag. The empirical evaluation is limited to synthetic data and does not use confidential government records.

The objectives are:

- To design an integrated AI and GIS framework for tracking government funds from budget allocation to ground-level project completion.
- To connect financial, procurement, contractor, physical-progress, geospatial, and relevant evidence records through a common project-level identifier.
- To evaluate anomaly-detection methods on simulated financial, physical, payment, and procurement irregularities.
- To examine whether explainable risk information can support human prioritization of projects for verification or audit.
- To demonstrate how GIS can provide spatial context and evidence linkage without treating synthetic locations as real-world evidence.

1.5 Research Question

Can an integrated AI- and GIS-enabled framework support transparent project monitoring and early identification of potential mismatches between government expenditure and reported ground-level project progress?

The paper makes four main contributions:

- A project-level integration architecture linking financial, procurement, physical-progress, GIS, and evidence records.
- A reproducible synthetic benchmark for evaluating anomaly-detection approaches without requiring confidential government data.
- An explainability layer for examining the factors associated with anomalous project records.
- A bounded novelty claim focused on cross-layer project integration rather than a new AI algorithm.

The remainder of the paper is organized as follows. Section 2 reviews related work and develops the research gap. Section 3 explains the methodology, synthetic benchmark, anomaly-detection methods, GIS analysis, explainability, and statistical testing. Section 4 describes the proposed system design. Section 5 presents and discusses the experimental results. Section 6 concludes the paper and outlines future work.

**LITERATURE REVIEW / RELATED WORK**

The literature was examined across five closely related areas: digital financial management, electronic procurement, AI-based procurement and expenditure risk detection, physical project-progress monitoring, and GIS/geospatial evidence. The review gives preference to peer-reviewed studies and official government sources and, for each source, notes both its contribution and the part of the problem it does not address.

[1] Banerjee et al. (2020). Banerjee et al. evaluated an electronic financial-management reform in India through a large field experiment. The reform reduced leakage-related indicators and program expenditure, although payment delays increased. The results support the use of digital financial controls for accountability, but the study did not connect financial records with the physical progress of individual projects.

[2] Lewis-Faupel et al. (2016). Lewis-Faupel et al. examined e-procurement in public works in India and Indonesia. They reported better road quality in India and shorter project delays in Indonesia, without finding evidence of lower prices. The study establishes the value of digital procurement, but procurement outcomes were not reconciled with later expenditure and ground-level evidence.

[3] PFMS (Government of India). PFMS provides facilities for fund-flow tracking, transactions, payments, expenditure reporting, and accounting. It therefore supplies an important financial-control layer for government monitoring. The financial focus, however, does not provide a unified view of physical progress and geospatial evidence for an individual project.

[4] GeoMGNREGA (Government of Maharashtra). GeoMGNREGA uses GIS, mobile technology, and geotagging to monitor MGNREGS assets and provide spatial evidence during implementation. It demonstrates the usefulness of location-based monitoring, but it does not itself reconcile financial, procurement, contractor, and physical-progress records in one project-level view.

[5] Gallego, Rivero and Martínez (2021). Gallego et al. developed an early-warning model using more than two million Colombian public-procurement contracts. The model was used to predict contracts associated with investigations, breaches, and implementation problems. The study shows that predictive procurement screening is feasible, but the analysis remains focused on procurement records.

[6] García Rodríguez et al. (2022). García Rodríguez et al. compared eleven machine-learning algorithms for collusion detection using procurement datasets from several countries. Their results show that performance changes with the data and the selected features. The work supports comparative model evaluation, but it does not include physical project evidence.

[7] Dutra et al. (2026). Dutra et al. applied statistical and machine-learning methods to large city-expenditure time series and ranked unusual spending using both anomaly frequency and monetary impact. The study is closely related to the expenditure-anomaly part of this work, but its analysis is not connected to procurement, physical progress, or GIS evidence.

[8] Pal et al. (2024). Pal et al. combined site images, BIM, point clouds, and semantic segmentation to estimate activity-level construction progress. Their results show that physical progress can be measured computationally. The work is mainly concerned with construction monitoring rather than reconciliation of public expenditure with project execution.

[9] Shamsollahi, Moselhi and Khorasani (2024). Shamsollahi et al. combined deep learning with real-time locating data for automated construction-progress monitoring. The study demonstrates how several physical-monitoring sources can be integrated, but it does not connect them to government finance or procurement information.

[10] Suh, Zhu and Zhao (2024). Suh et al. used dense satellite time-series data with a U-Net-based method to detect construction changes. The study shows the value of scalable remote-sensing evidence, while financial and procurement reconciliation remains outside its scope.

[11] Nai et al. (2023). Nai et al. used procurement and administrative-justice data with machine learning and recommender methods to identify characteristics associated with complaints and disputes. This adds useful legal and complaint context, but the work remains focused on procurement cases rather than full project execution.



[12] Decarolis and Giorgiantonio (2022). Decarolis and Giorgiantonio tested corruption red flags in Italian tender data using statistical learning models and found that some obvious indicators had limited predictive value. This result is important for the present work because it shows why individual red flags should not be assumed to be reliable without validation.

[13] Medina-Hernández, Kertész and Fazekas (2026). Medina-Hernández et al. combined positive-unlabeled learning, procurement red flags, and network features using procurement and supplier-sanction data from Mexico. The findings support the use of network-derived information in procurement-risk analysis, but the scope remains centered on suppliers and procurement.

[14]-[16] Methodological foundations. Hevner et al. and Peffers et al. provide the design-science and DSRM foundations used to frame and evaluate the proposed information-system artifact [14], [15]. Lundberg and Lee provide the SHAP method used to examine feature contributions to model outputs [16].

[17]-[19] Algorithm foundations. Liu et al. introduced Isolation Forest [17], Breunig et al. introduced Local Outlier Factor [18], and Pedregosa et al. described scikit-learn as a standard Python machine-learning library [19]. These references provide the basis for the methods used in the experiment.

2.1 Literature Synthesis and Research Gap

Taken together, the literature can be viewed as three related streams. The first covers financial governance and procurement, where digital systems have improved traceability, accountability, and implementation outcomes [1]-[4]. The second covers risk analytics, where machine learning, red-flag indicators, and network features are used to identify potentially risky procurement or expenditure activity [5]-[7], [11]-[13]. The third covers physical monitoring, where computer vision, BIM, deep learning, and satellite data are used to quantify or detect project changes [8]-[10].

The main limitation across these streams is not a lack of technology, but the separation of the information itself. The research gap addressed in this paper is the project-level reconciliation of financial expenditure, procurement information, physical progress, GIS/evidence, and explainable risk information under a common project identifier.

2.2 Comparison with Existing Integrated Systems

There are also existing systems that combine more than one information source. India's PM Gati Shakti National Master Plan provides a GIS-enabled platform for integrating infrastructure-related information across ministries and states for planning and monitoring [20]. The European Commission's Arachne+ is a data-mining and risk-scoring tool used to identify higher-risk EU-funded projects, beneficiaries, contracts, and contractors [21]. These examples show that cross-source integration and risk scoring are practical at institutional scale. The proposed framework is narrower: it is a project-level research design that brings financial-physical reconciliation together with GIS/evidence and explainable anomaly signals for human review. It is not presented as a replacement for these existing systems.

Research stream	Representative evidence	Remaining limitation / relevance
Financial & procurement	[1]-[4]	Strong digital controls and monitoring, but financial and physical layers remain separated.
Procurement & expenditure risk	[5]-[7], [11]-[13]	Risk detection is effective within specific data layers; cross-layer project reconciliation is limited.
Physical & geospatial monitoring	[8]-[10], [20]	Progress and location can be measured or visualized, but financial reconciliation is outside scope.
Risk-scoring systems	[13], [21]	Institutional risk tools exist, but the proposed framework focuses on project-level financial-physical consistency and evidence linkage.

Table 1: Synthesis of related work and the remaining integration gap



RESEARCH METHODOLOGY

3.1 Research Design

A design-science approach was used because the research is concerned with a practical monitoring problem, a proposed information-system artifact, and the evaluation of selected analytical components [14], [15]. This keeps the work beyond a purely conceptual discussion while avoiding any claim that a production government system has been deployed.

The DSRM process was used as the organizing logic of the study: problem identification and motivation are covered in Sections 1.2-1.3; objectives are set out in Section 1.4; system design is presented in Section 4; the synthetic benchmark provides the demonstration in Sections 3.2-3.7; Section 5 contains the evaluation; and this paper provides the final communication of the work [15].

3.2 Data Sources and Synthetic Dataset

Two types of information were used. The first is secondary documentary evidence, including peer-reviewed research, official PFMS information, GeoMGNREGA documentation, and official descriptions of integrated public infrastructure and risk-monitoring systems [1]-[4], [20], [21]. The second is a synthetic project-level dataset created for the experiment. These synthetic records are not government data and are not intended to represent a particular department, state, or public scheme.

The dataset contains 3,000 project records: 2,550 normal records (85%) and 450 anomalous records (15%). The anomalous cases were distributed across four injected patterns: (i) high expenditure with low physical progress, (ii) overspending beyond the sanctioned budget ratio, (iii) long payment delays combined with low fund release, and (iv) procurement irregularity represented by high tender variance and an unusual contractor-count pattern. A fixed random seed was used during data generation so that the benchmark could be reproduced.

Feature	Description
Allocation amount	Project-wise sanctioned amount, generated in the range ₹0.5–5.0 million.
Release ratio	Released funds relative to allocation.
Expenditure ratio	Expenditure relative to allocation; values above 1.0 represent overspending.
Physical progress	Reported physical completion ratio.
Financial-physical gap	Expenditure ratio minus physical progress.
Payment delay	Payment delay in days.
Tender variance	Standardized procurement-price variation indicator.
Contractor count	Number of contractors/implementing agencies represented in the record.
GIS evidence count	Synthetic count representing available geospatial evidence metadata.
Citizen evidence count	Synthetic count representing citizen-generated evidence metadata.

Table 2: Synthetic project-level feature set

3.3 Anomaly Detection Models

Three methods were compared. Isolation Forest was selected as the main unsupervised anomaly-ranking method because unusual observations can be isolated from the bulk of the data without requiring labelled fraud cases [17]. Local Outlier Factor (LOF) was used as a density-based alternative that considers how isolated an observation is within its local neighbourhood [18]. A simple rule-based baseline was added as a transparent reference point so that the experiment could compare machine-learning methods with explicit domain rules.

For reproducibility, the rule-based method flagged a record when at least one of these conditions was met: financial-physical gap > 0.40; expenditure ratio > 1.10; payment delay > 90 days together with release ratio < 0.50; or tender variance above the normal-training mean plus two standard deviations. Isolation Forest used 200 trees with contamination = 0.15. LOF used 20 neighbours with contamination = 0.15 and novelty mode. Numeric variables were standardized with a StandardScaler fitted only on normal training records.



3.4 Training, Testing and Thresholding

For each of the five main random seeds, the data were split 70:30 using stratification. The unsupervised models were fitted only on normal records from the training portion, while the test set contained both normal and anomalous records. For the controlled benchmark, the top 15% of test observations by anomaly score were treated as positive predictions, matching the known simulated anomaly prevalence. This is why precision, recall, and F1 are numerically identical within each model. Since an audit team would not know the true anomaly prevalence in practice, Precision@K was also used as a more operational measure.

3.5 Explainable AI

SHAP-based feature attribution was used to examine which variables contributed most to the Isolation Forest anomaly scores [16]. These values are interpreted as indicators of model behaviour, not as causal effects. They also do not establish that an underlying government record is correct or that a flagged contractor has committed misconduct.

3.6 GIS Methodology

Each project identifier was associated with synthetic coordinates and evidence metadata. Because real project coordinates were not available, latitude and longitude values were generated within approximate Maharashtra bounds. A Global Moran's I test using a five-nearest-neighbour spatial-weights matrix was then applied to the Isolation Forest risk flags. The purpose was to check the spatial-processing pipeline; the result cannot be interpreted as a real pattern of geographic risk because the coordinates were assigned randomly.

3.7 Validation and Statistical Test

Five random seeds (11, 22, 33, 44, and 55) were used for the main comparison. Precision, recall, F1-score, ROC-AUC, and PR-AUC were calculated, and Precision@50, Precision@100, and Precision@200 were used to examine audit-prioritization performance. A separate 30-seed experiment compared Isolation Forest with the rule-based baseline using ROC-AUC and a Wilcoxon signed-rank test. The same significance test was not run for LOF, so no pairwise statistical-superiority claim is made for LOF.

3.8 Reproducibility, Ethics and Data Privacy

The complete notebook for dataset generation, model training, evaluation, SHAP analysis, spatial testing, and the 30-seed statistical test was executed in Google Colab. All project records, coordinates, and evidence counts used in the experiment are synthetic. No real government, personal, contractor, citizen, or confidential records were used. In a real deployment, a high risk score would be treated only as an early-warning signal and would require human verification, together with appropriate privacy, security, interoperability, and data-governance controls.

SYSTEM DESIGN AND DIAGRAMS

The proposed framework is described using standard system-modelling diagrams. Together, the diagrams show how information moves through the framework, how users interact with it, how the main data entities are related, how the monitoring process unfolds over time, how a project changes state, and how the components could be organized technically. They describe the proposed design only; they do not represent a live government production system.

4.1 DFD Level 1

The Data Flow Diagram (DFD) Level 1 breaks the framework into budget and fund management, procurement and contractor management, payment and expenditure tracking, physical progress and GIS evidence, AI-based anomaly detection, and risk scoring with human verification. It also shows the main project-financial and progress-evidence data stores.

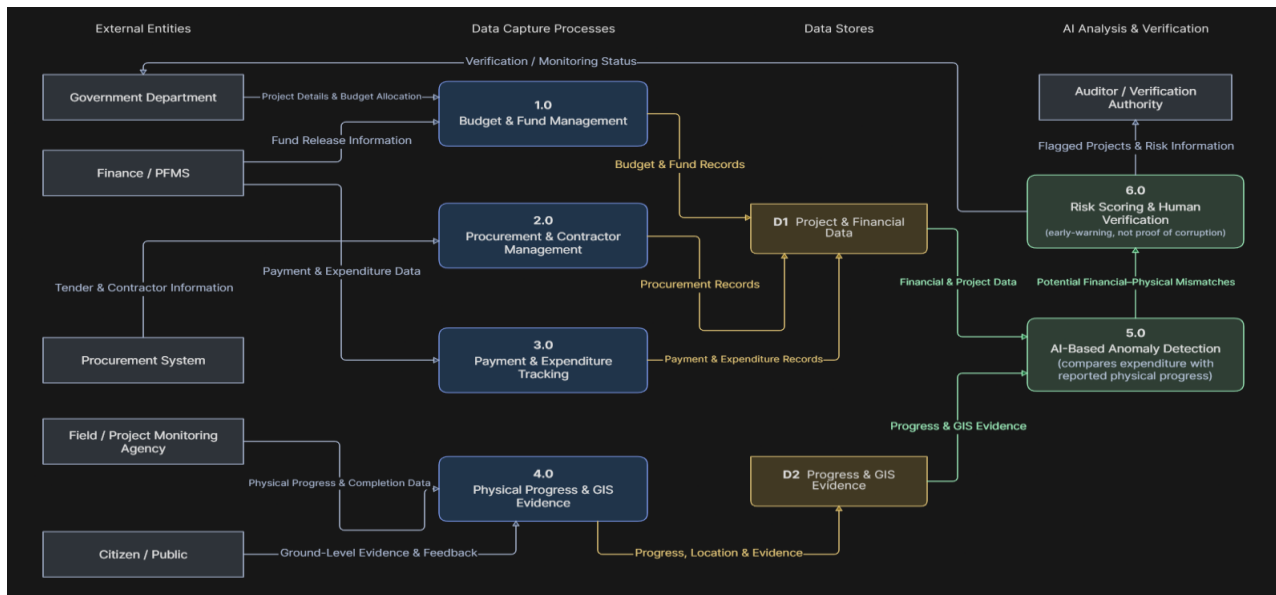


Figure 1: DFD Level 1 - Proposed AI and GIS Government Fund Tracking Framework

4.2 Use Case Diagram

The Use Case Diagram shows the main actors and the functions they perform within the proposed framework. Government departments, finance/PFMS, procurement systems, field monitoring agencies, citizens/public, and verification authorities interact with functions such as fund tracking, procurement monitoring, physical-progress monitoring, evidence viewing, anomaly detection, risk review, verification, and reporting.

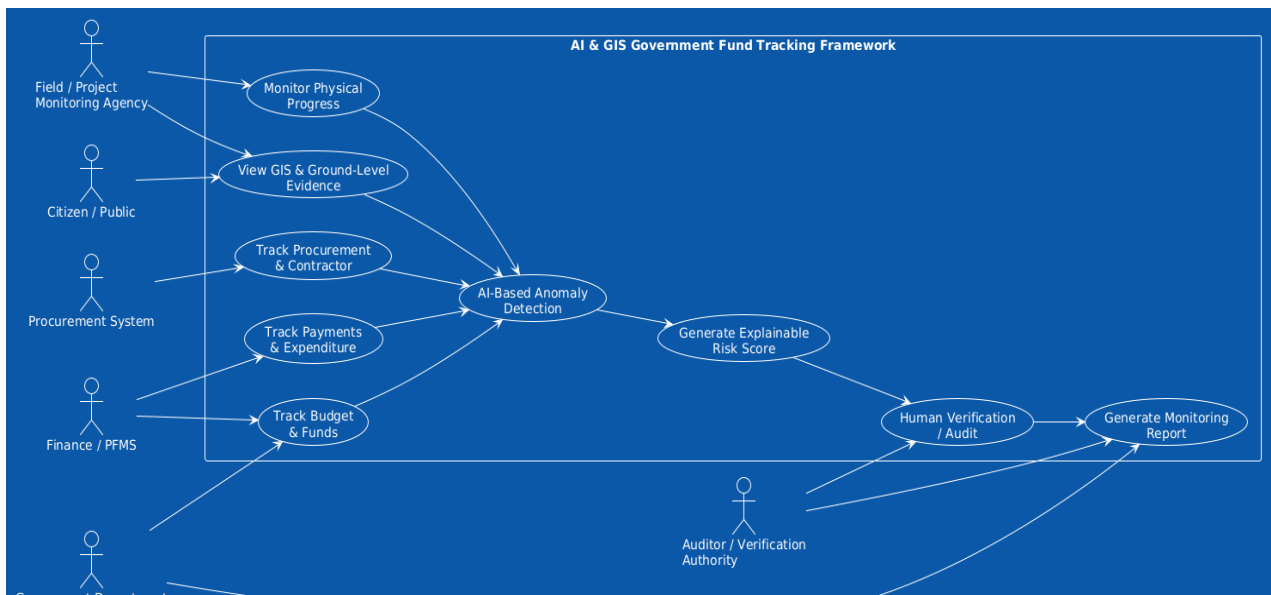


Figure 2: Use Case Diagram - Proposed AI and GIS Government Fund Tracking Framework

4.3 Class Diagram

The Class Diagram defines the principal project-level entities and their relationships: project, budget, fund release, procurement, contractor, expenditure, physical progress, GIS evidence, citizen evidence, AI anomaly, risk score, and verification. These entities form the conceptual data structure used in the proposed framework.

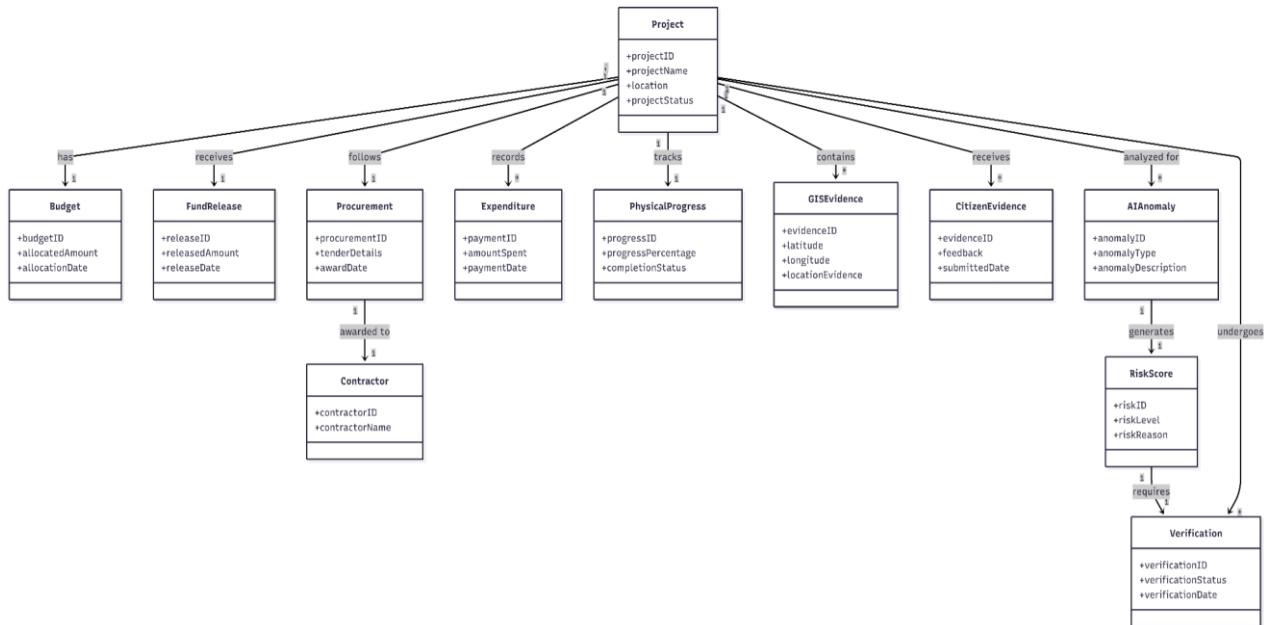


Figure 3: Class Diagram - Proposed AI and GIS Government Fund Tracking Framework

4.4 Sequence Diagram

The Sequence Diagram follows the main exchange of information among the government department, financial system, procurement system, field and GIS evidence sources, AI anomaly-detection module, risk-scoring layer, and verification authority. It presents the proposed monitoring workflow from data collection to human review.

Sequence Diagram — Proposed AI and GIS Government Fund Tracking Framework

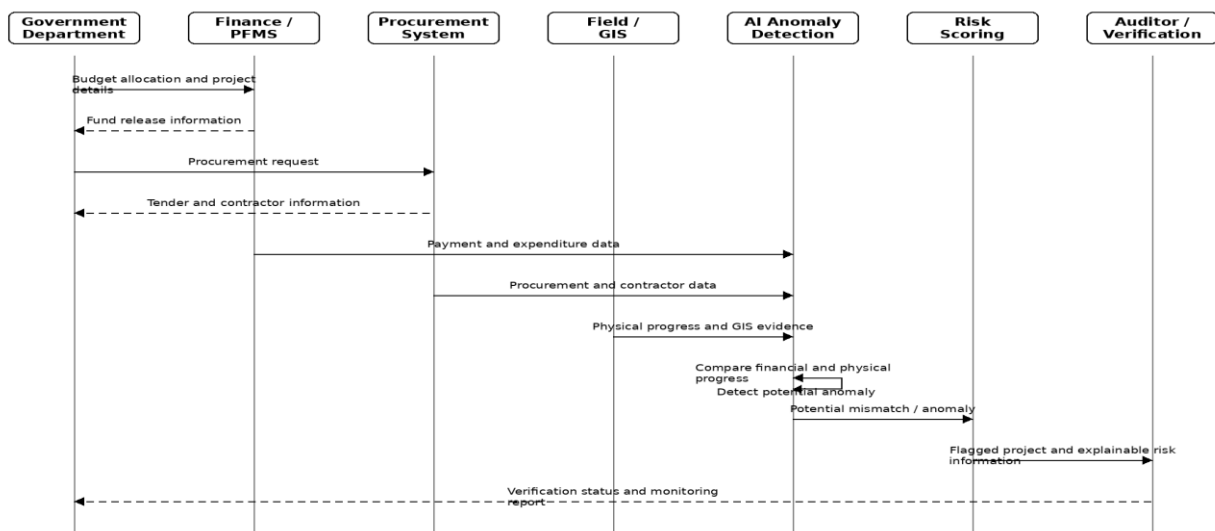


Figure 4: Sequence Diagram - Proposed AI and GIS Government Fund Tracking Framework

4.5 State Chart Diagram

The State Chart follows the lifecycle of a government development project from creation and budget allocation through procurement, execution, physical-progress and evidence updates, AI analysis, risk assessment, verification, and completion. When a potential anomaly is identified, the project moves to human verification or audit before completion

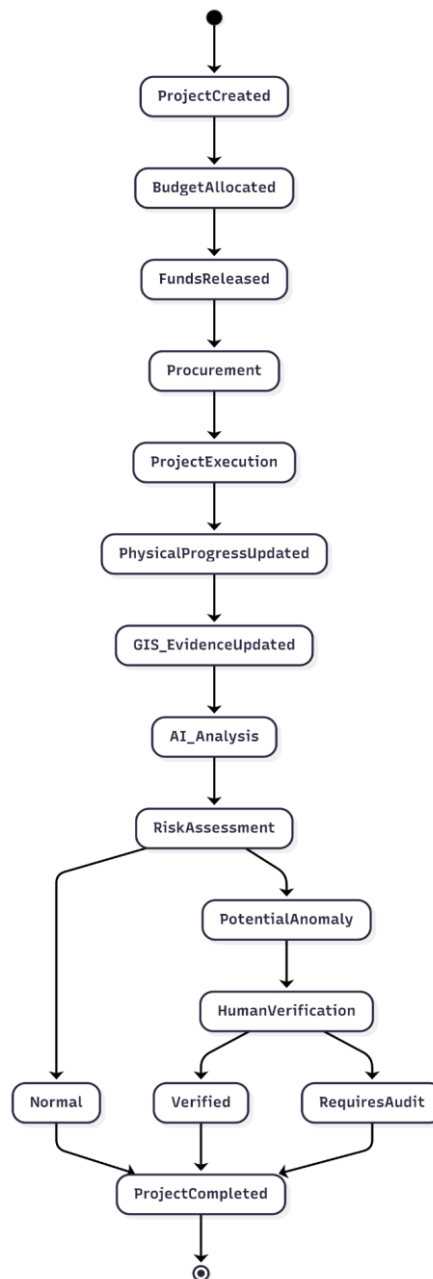


Figure 5: State Chart Diagram - Proposed AI and GIS Government Fund Tracking Framework

4.6 Deployment Diagram

The Deployment Diagram gives the proposed technical organization of the framework. It separates users and external government systems from application services, the AI anomaly-detection and explainable-risk module, and the project/GIS data stores. The diagram is architectural and conceptual rather than a description of an existing production deployment.

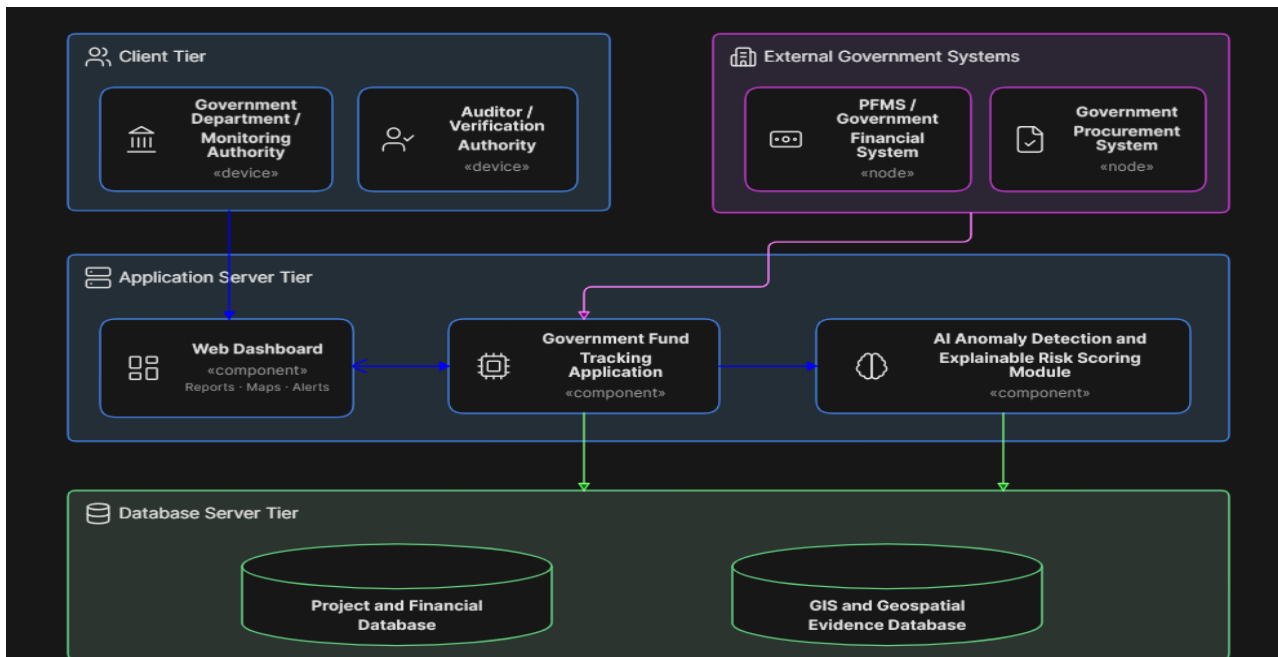


Figure 6: Deployment Diagram - Proposed AI and GIS Government Fund Tracking Framework

RESULTS AND DISCUSSION

The validation dataset contained 3,000 synthetic government-project records, including 2,550 normal records and 450 deliberately injected anomalous records (15%).

The results in this section come from the synthetic benchmark described in the methodology and from the executed research notebook. They show how the proposed analytical pipeline behaves under controlled, simulated conditions. They should not be read as measurements of real government fraud, corruption, or project performance.

5.1 Synthetic Benchmark

The benchmark contains 3,000 project records: 2,550 normal records and 450 injected anomalies. The four anomaly patterns were intentionally created with feature ranges that are largely separated from the normal records. This makes the benchmark suitable for checking whether the analysis pipeline works, but it also makes the task easier than a real government setting.

5.2 Model Performance

Model	Precision	Recall	F1	ROC-AUC	PR-AUC
Isolation Forest	0.573 ± 0.041	0.573 ± 0.041	0.573 ± 0.041	0.890 ± 0.017	0.579 ± 0.057
LOF	0.985 ± 0.005	0.985 ± 0.005	0.985 ± 0.005	0.9999 ± 0.00002	0.999 ± 0.0001
Rule-based	0.874 ± 0.028	0.874 ± 0.028	0.874 ± 0.028	0.982 ± 0.006	0.876 ± 0.019

Table 3: Mean ± standard deviation of anomaly-detection performance across five seeds

LOF produced the strongest scores in the main benchmark, followed by the rule-based baseline and Isolation Forest. The LOF ROC-AUC of 0.9999 is particularly high and should be interpreted cautiously. It is consistent with the way the synthetic anomalies were generated, where several problem types occupy feature ranges that are relatively easy to distinguish from normal records.



Precision, recall, and F1 are identical within each row because the decision rule was set to flag exactly 15% of the test records, matching the known simulated anomaly prevalence. This controlled threshold makes the comparison useful for ranking performance, but it would not be available in a real deployment because the true anomaly rate would not be known in advance. Thresholds would therefore need to be calibrated separately, for example using an audit capacity such as Precision@K.

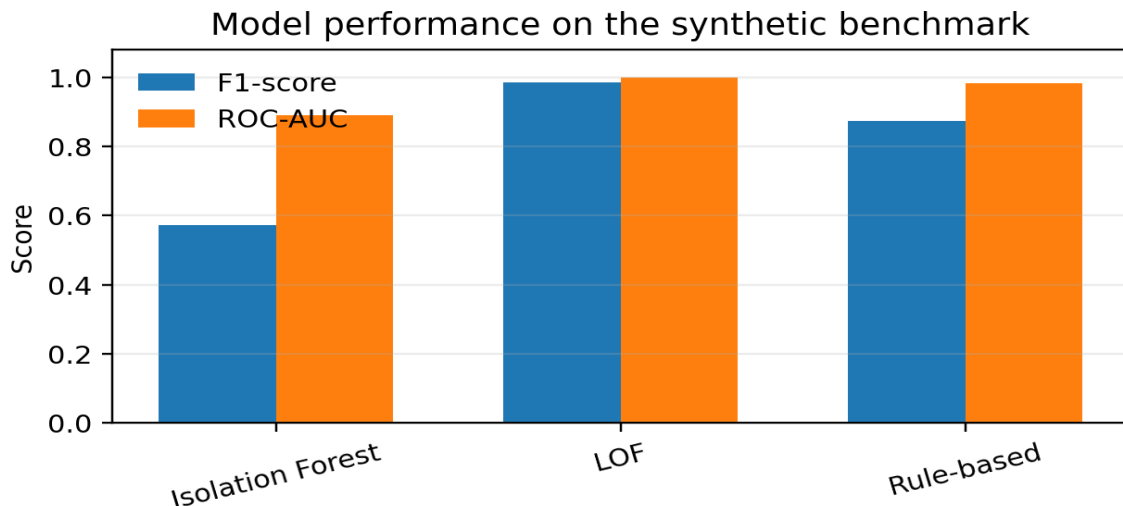


Figure 7: Comparison of F1-score and ROC-AUC across the three detection methods

5.3 Audit Prioritization with Precision@K

Audit budget	Precision@K	No-skill reference
Top 50	0.680	0.150
Top 100	0.620	0.150
Top 200	0.465	0.150

Table 4: Precision@K for Isolation Forest risk ranking (seed = 11)

Precision@50 of 0.680 means that 68% of the top 50 ranked records were known injected anomalies in this controlled benchmark, compared with 15% expected from random selection. The result suggests that risk ranking can concentrate unusual cases when audit capacity is limited. It should not be interpreted as a real-world audit hit rate.

5.4 Explainability Findings

For the Isolation Forest model, the largest mean absolute SHAP contributions were associated with tender variance (0.256), financial-physical gap (0.244), GIS evidence count (0.202), physical progress (0.201), citizen evidence count (0.193), payment delay (0.189), release ratio (0.187), allocation amount (0.186), expenditure ratio (0.182), and contractor count (0.041). In this benchmark, the model therefore combined procurement, financial, progress, and evidence-related signals rather than relying on a single variable. These values describe the synthetic experiment and would need to be recalculated with real project data before operational use.

5.5 GIS Spatial Test

The Global Moran's I test on the Isolation Forest risk flags produced $I = -0.0162$ with $p = 0.0670$ using a five-nearest-neighbour weights matrix. No statistically significant spatial clustering was observed at the 0.05 level. This is consistent with the random assignment of coordinates. The result is therefore best treated as a check of the spatial-processing pipeline, not as evidence about geographic concentration of project risk.



5.6 Statistical Comparison

Across 30 random seeds, Isolation Forest obtained a mean ROC-AUC of 0.8936 ± 0.0165 , compared with 0.9829 ± 0.0042 for the rule-based baseline. The Wilcoxon signed-rank test yielded a statistic of 0.0 and $p = 1.86 \times 10^{-9}$, with a paired Cohen's d of 5.82, indicating a very large standardized effect. Showing a consistent difference between the two methods in this synthetic benchmark. The same test was not performed for LOF, so the paper does not make a statistical-significance claim comparing LOF with the other methods.

5.7 Discussion

One clear result from the experiment is that model choice depends on how the anomaly patterns are formed. LOF was the strongest method in this benchmark because the injected anomalies were relatively distinct in feature space. At the same time, the rule-based method consistently outperformed Isolation Forest. This is useful for the proposed application because transparent rules can remain competitive when the conditions being monitored are well understood [12], [17], [18].

The main contribution of the paper is still the project-level architecture rather than a new algorithm. Earlier studies provide separate capabilities for financial control, procurement risk, expenditure anomalies, physical-progress monitoring, and geospatial evidence [1]-[13]. The proposed framework brings these layers together under a common project identifier and treats the financial-physical relationship as a central monitoring link. In practice, such a system should narrow the set of projects that human reviewers need to examine rather than replace audit judgment.

5.8 Limitations

- The benchmark is synthetic and cannot establish real-world detection accuracy, generalization, or fraud prevalence.
- LOF's near-perfect score is a best-case benchmark outcome caused by relatively easy-to-separate injected anomalies.
- GIS coordinates, GIS evidence counts, and citizen evidence counts are randomly generated and do not represent real projects, photographs, satellite observations, or citizen complaints.
- The spatial test is a pipeline check rather than a substantive study of geographic risk concentration.
- The Wilcoxon significance test compares Isolation Forest and the rule-based baseline only.
- Live integration with PFMS, procurement platforms, or state GIS systems was not performed.
- SHAP values describe model behaviour on synthetic data and do not validate the underlying data or establish causality.
- The study makes no claim that the proposed framework itself reduces leakage or corruption; such a claim would require longitudinal field evaluation.

CONCLUSION AND FUTURE WORK

This paper presented and evaluated an integrated AI and GIS framework for tracking government funds from budget allocation to ground-level project completion. The literature shows that financial monitoring, procurement analytics, physical-progress estimation, and geospatial evidence are already supported by different tools and studies. The remaining opportunity is to connect these layers for an individual project and use the combined information to identify cases that deserve further scrutiny.

The controlled experiment shows that anomaly detection can be evaluated without confidential government data. LOF achieved the highest benchmark scores, while the rule-based method consistently outperformed Isolation Forest in the 30-seed comparison. Precision@50 showed that risk ranking can concentrate known simulated anomalies within a limited review budget. SHAP analysis provided feature-level explanations, and the spatial test demonstrated how GIS risk flags can be processed even when the locations are synthetic.

These findings need to be read in the context of the benchmark. They do not establish that the framework can detect real fraud or corruption, and the synthetic scores should not be assumed to transfer directly to government data. The main value of the work is the framework and the validation procedure: different project records can be connected under a common identifier, unusual cases can be ranked, and the factors behind a flag can be shown to a human reviewer.

6.1 Future Work

- The next step is to evaluate the framework with public or institutionally approved government project datasets and to calibrate the thresholds with domain experts.
- The LOF, rule-based, and Isolation Forest methods should then be compared again under more difficult anomaly settings in which normal and abnormal records overlap more closely.



- Verified geotagged photographs, satellite imagery, and other remote-sensing sources can be added to strengthen the physical-progress and GIS components.
- Another useful direction is to study the human-AI audit workflow, including whether risk-ranked project lists actually reduce review time or help auditors find important cases that would otherwise be missed.
- Before any operational use, the framework would also need testing for interoperability, privacy, security, fairness, and model drift.

6.2 Reproducibility Statement

The research notebook used for synthetic-data generation, model training, evaluation, SHAP analysis, GIS testing, and statistical analysis was executed in Google Colab. The benchmark and analysis code are available in the project repository maintained by the author. No real government, personal, or confidential data were used in producing the reported results.

REFERENCES

- [1]. A. Banerjee, E. Duflo, C. Imbert, S. Mathew, and R. Pande, "E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in India," *American Economic Journal: Applied Economics*, vol. 12, no. 4, pp. 39-72, 2020, doi: 10.1257/app.20180302.
- [2]. S. Lewis-Faupel, Y. Neggers, B. A. Olken, and R. Pande, "Can electronic procurement improve infrastructure provision? Evidence from public works in India and Indonesia," *American Economic Journal: Economic Policy*, vol. 8, no. 3, pp. 258-283, 2016, doi: 10.1257/pol.20140258.
- [3]. Public Financial Management System, "About Public Financial Management System," Controller General of Accounts, Department of Expenditure, Ministry of Finance, Government of India, 2024. [Online]. Available: <https://pfms.nic.in/SitePages/aboutus.aspx>
- [4]. Maharashtra Employment Guarantee Scheme, "Monitoring Initiatives: GeoMGNREGA," Government of Maharashtra. [Online]. Available: <https://mahaegs.maharashtra.gov.in/en/mgnregs/monitoring-initiatives/>
- [5]. J. Gallego, G. Rivero, and J. Martínez, "Preventing rather than punishing: An early warning model of malfeasance in public procurement," *International Journal of Forecasting*, vol. 37, no. 1, pp. 360-377, 2021, doi: 10.1016/j.ijforecast.2020.06.006.
- [6]. M. J. García Rodríguez, V. Rodríguez-Montequín, P. Ballesteros-Pérez, P. E. D. Love, and R. Signor, "Collusion detection in public procurement auctions with machine learning algorithms," *Automation in Construction*, vol. 133, art. 104047, 2022, doi: 10.1016/j.autcon.2021.104047.
- [7]. M. T. Dutra, L. G. L. Costa, G. P. Oliveira, M. O. Silva, and G. L. Pappa, "Detecting and analyzing anomalies in city expenditure time series," *Journal on Interactive Systems*, vol. 17, no. 1, pp. 456-472, 2026, doi: 10.5753/jis.2026.7072.
- [8]. A. Pal, J. J. Lin, S.-H. Hsieh, and M. Golparvar-Fard, "Activity-level construction progress monitoring through semantic segmentation of 3D-informed orthographic images," *Automation in Construction*, vol. 157, art. 105157, 2024, doi: 10.1016/j.autcon.2023.105157.
- [9]. D. Shamsollahi, O. Moselhi, and K. Khorasani, "Data integration using deep learning and real-time locating system (RTLS) for automated construction progress monitoring and reporting," *Automation in Construction*, vol. 168, art. 105778, 2024, doi: 10.1016/j.autcon.2024.105778.
- [10]. J. W. Suh, Z. Zhu, and Y. Zhao, "Monitoring construction changes using dense satellite time series and deep learning," *Remote Sensing of Environment*, vol. 309, art. 114207, 2024, doi: 10.1016/j.rse.2024.114207.
- [11]. R. Nai, R. Meo, G. Morina, and P. Pasteris, "Public tenders, complaints, machine learning and recommender systems: A case study in public administration," *Computer Law & Security Review*, vol. 51, art. 105887, 2023, doi: 10.1016/j.clsr.2023.105887.
- [12]. F. Decarolis and C. Giorgiantonio, "Corruption red flags in public procurement: New evidence from Italian calls for tenders," *EPJ Data Science*, vol. 11, art. 16, 2022, doi: 10.1140/epjds/s13688-022-00325-x.
- [13]. M. Medina-Hernández, J. Kertész, and M. Fazekas, "Learning from sanctioned government suppliers: A machine learning and network science approach to detecting fraud and corruption in Mexico," *Scientific Reports*, vol. 16, art. 22382, 2026, doi: 10.1038/s41598-026-48873-w.
- [14]. A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75-105, 2004, doi: 10.2307/25148625.
- [15]. K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45-77, 2007, doi: 10.2753/MIS0742-1222240302.
- [16]. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems* 30, pp. 4765-4774, 2017.



- [17]. F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in Proceedings of the 8th IEEE International Conference on Data Mining, pp. 413-422, 2008, doi: 10.1109/ICDM.2008.17.
- [18]. M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 93-104, 2000, doi: 10.1145/342009.335388.
- [19]. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
- [20]. Government of India, Ministry of New and Renewable Energy, "Gati Shakti: National Master Plan for Multi-Modal Connectivity," [Online]. Available: <https://gatishakti.mnre.gov.in/home>
- [21]. European Commission Anti-Fraud Knowledge Centre, "What is ARACHNE?," [Online]. Available: https://antifraud-knowledge-centre.ec.europa.eu/useful-tools/what-arachne_en