

Development of an Artificial Intelligence-Based Integrated Tax Intelligence and Early Warning System for GST and Income Tax Administration to Detect Tax Evasion and Enhance Revenue Protection in India.

Dr. AjuSaigal

Senior Data Analyst

Tax Research & Policy Cell, GST Department, Government of Kerala, India

Abstract

Tax evasion remains one of the major challenges affecting revenue mobilization and economic governance in India. Although the Goods and Services Tax (GST) and Income Tax systems independently collect vast amounts of taxpayer information, limited real-time integration and intelligence-driven analysis often hinder the identification of tax evasion, fake invoicing, under-reporting of income, and non-compliant taxpayers. This research proposes the development of an Artificial Intelligence (AI)-Based Integrated Tax Intelligence and Early Warning System that combines data from GST and Income Tax Departments to facilitate proactive tax administration and revenue protection. The proposed framework utilizes Big Data Analytics, Machine Learning, Risk Scoring Models, Predictive Analytics, and Automated Alert Mechanisms to identify high-risk taxpayers and suspicious transactions. The system integrates taxpayer information from GST returns, Income Tax returns, TDS/TCS records, e-Invoices, e-Way Bills, banking transactions, customs data, and other government databases to generate real-time compliance alerts and risk assessments. The proposed system aims to enhance tax compliance, reduce revenue leakage, detect tax fraud at an early stage, improve interdepartmental coordination, and support evidence-based decision-making. The study contributes towards building a National Tax Intelligence Framework capable of strengthening India's tax administration through advanced analytics and artificial intelligence technologies.

Keywords

Artificial Intelligence, GST, Income Tax, Tax Intelligence, Tax Evasion Detection, Revenue Protection, Big Data Analytics, Machine Learning, Risk Assessment, Early Warning System, Data Integration, Predictive Analytics, Revenue Intelligence, Compliance Monitoring.

1. Introduction

Tax revenue constitutes the backbone of public finance and economic development. Despite significant reforms such as GST implementation and digitization of tax administration, tax evasion, fake invoicing, shell companies, under-reporting of turnover, and concealment of income continue to affect government revenue. Currently, GST and Income Tax Departments maintain large but largely independent databases. The absence of comprehensive real-time integration often results in delayed identification of suspicious taxpayers. Advances in Artificial Intelligence, Machine Learning, Data Mining, and Big Data Analytics provide opportunities for integrating multiple data sources to create intelligent tax administration systems capable of detecting anomalies and predicting non-compliance. This study proposes an AI-driven integrated framework to support tax authorities in identifying tax risks, generating early warnings, and improving compliance monitoring through automated intelligence generation.

2. Research Objectives

The research focuses on developing an integrated tax intelligence framework by combining GST and Income Tax data to create a unified view of taxpayer activities. The study explores the application of Artificial Intelligence and machine learning techniques for taxpayer risk assessment, anomaly detection, and fraud identification. It also examines automated compliance monitoring mechanisms that can continuously evaluate taxpayer behaviour and generate early warning alerts for potential non-compliance. Furthermore, the research aims to develop a revenue intelligence system that supports tax authorities in detecting tax evasion, improving audit selection, strengthening enforcement actions, and enhancing revenue protection through data-driven decision-making. Tax Evasion and Revenue Leakage

1. To develop an integrated GST-Income Tax intelligence framework.
2. To identify patterns associated with tax evasion and revenue leakage.
3. To design AI-based taxpayer risk scoring models.
4. To develop an automated early warning and alert generation system.
5. To integrate multiple government databases for compliance monitoring.
6. To improve taxpayer profiling and behavioural analytics.
7. To enhance interdepartmental information sharing.
8. To support evidence-based tax enforcement and policy decisions.
9. To reduce fraudulent GST claims and fake invoicing activities.
10. To strengthen revenue protection mechanisms in India.

3. Methodology

Tax evasion continues to be a significant challenge for tax administrations worldwide and is commonly associated with practices such as under-reporting of sales, fraudulent Input Tax Credit claims, cash-based transactions outside the formal economy, shell companies, benami transactions, and circular trading networks. Recent advancements in Artificial Intelligence have enabled tax authorities to strengthen compliance monitoring through fraud detection, taxpayer risk scoring, compliance prediction, and anomaly detection. Several countries, including the United States, United Kingdom, Australia, Singapore, and Estonia, have adopted AI-driven tax administration systems to improve enforcement efficiency and revenue collection.

The proposed research leverages Big Data Analytics, which is characterized by high data volume, velocity, variety, veracity, and value. These capabilities support advanced applications such as revenue forecasting, taxpayer segmentation, fraud analytics, and predictive compliance monitoring. The study utilizes data from GST systems, including GSTIN, GSTR-1, GSTR-3B, e-Invoices, and e-Way Bills, along with Income Tax records such as PAN, Income Tax Returns, TDS/TCS information, Annual Information Statements (AIS), and Form 26AS. Additional data sources include corporate information from customs records, banking transactions, geospatial databases, and local body licensing information. The research adopts a risk-based taxpayer selection methodology using historical data collected over a five-year period covering multiple sectors and states. Data processing involves cleansing, normalization, deduplication, and PAN-GSTIN linkage to ensure data quality and consistency. The integrated framework combines GST information such as turnover, tax liability, and ITC claims with Income Tax data relating to income, tax payments, and business profits. Banking information, corporate ownership records, and geospatial business data are further incorporated to provide a comprehensive taxpayer profile.

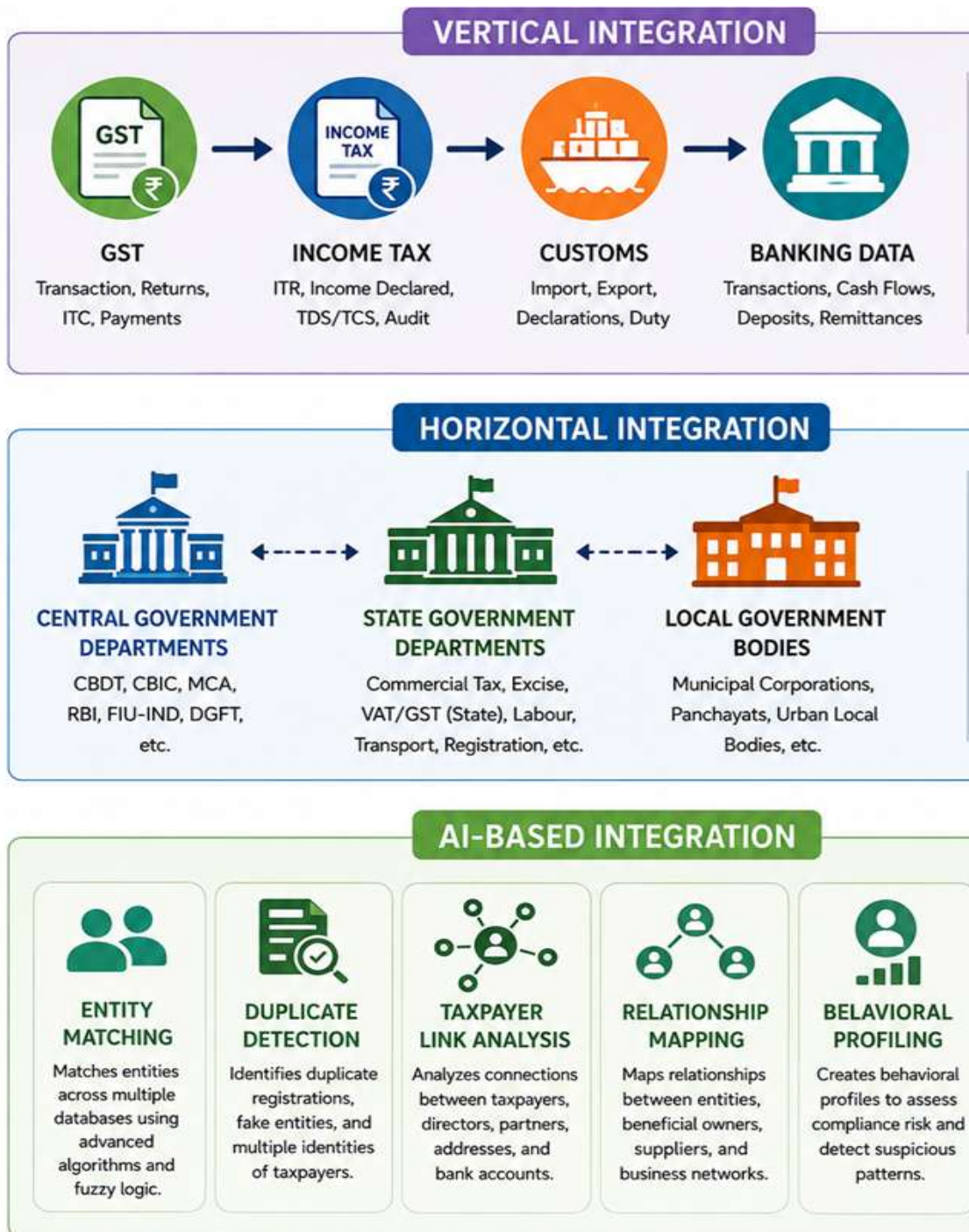
PAN serves as the primary integration key across multiple databases, while secondary identifiers such as mobile numbers, addresses, Director Identification Numbers, and other legally permissible identifiers help establish relationships between taxpayers and business entities. The integrated dataset is then analyzed using AI and machine learning models to identify hidden patterns, detect anomalies, generate taxpayer risk scores, and provide actionable intelligence for tax authorities. The overall objective is to improve tax compliance, enhance fraud detection capabilities, strengthen interdepartmental coordination, and support data-driven revenue protection strategies.

4. Proposed System Architecture and Data Integration Methods

The proposed system architecture consists of five interconnected layers designed to transform raw data into actionable tax intelligence. The Data Collection Layer gathers information from multiple sources, including GST, Income Tax, e-Invoice, e-Way Bill, TDS/TCS, customs, banking, corporate, property registration, and geospatial databases. The Data Integration Layer consolidates these diverse datasets through data cleansing, standardization, PAN-GSTIN mapping, entity resolution, and ETL processes to create a unified master taxpayer repository. The AI Analytics Layer applies

machine learning algorithms, risk scoring models, anomaly detection techniques, fraud detection mechanisms, predictive analytics, and taxpayer segmentation to identify compliance risks and suspicious activities. The Alert Generation Layer automatically produces real-time alerts, risk classifications, and investigation recommendations for potential tax evasion cases. Finally, the Decision Support Layer provides dashboards, revenue intelligence reports, heat maps, and compliance monitoring tools to support informed decision-making by tax authorities.

The data integration framework combines information from various government and external databases using PAN as the primary linkage mechanism and other identifiers for enhanced entity matching. Vertical integration enables seamless information flow across GST, Income Tax, customs, banking, and related financial systems, while horizontal integration facilitates data sharing among central government departments, state agencies, and local bodies. Advanced AI-based integration techniques are used to identify duplicate records, establish relationships among taxpayers and business entities, analyze behavioral patterns, and create comprehensive taxpayer profiles. This integrated approach enables early detection of tax evasion, improves compliance monitoring, strengthens revenue protection, and supports data-driven tax administration.



AI-Based Integrated Tax Intelligence & Early Warning System – Workflow



Data Fields for Integration & Working Mechanism

Data Source	Key Fields Integrated	Integration Key	Purpose / Expected Result
GST Department	GSTIN, PAN, Trade Name, Business Type, Turnover, GSTR-1 Data, GSTR-3B Data, E-Invoice Data, E-Way Bill Data, Input Tax Credit (ITC) Claims, Tax Payments	PAN, GSTIN	Compare reported turnover, tax liability, ITC claims, invoice data, and movement of goods with other databases to identify tax evasion and compliance risks.
Income Tax Department	PAN, Taxpayer Name, Income Declared, Business Income, TDS Credits, TCS Credits, Capital Gains, Audit Reports, ITR Filings	PAN	Compare declared income with GST turnover and financial transactions to identify under-reporting of income and tax mismatches.
Banking Data	Bank Account Number, Transaction Volume, Cash Deposits, International Remittances	PAN, Account Mapping	Verify whether banking transactions are proportionate to reported turnover and income; identify unaccounted transactions and potential tax evasion.
MCA (Ministry of Corporate Affairs)	CIN, Directors, Shareholders, Registered Address	PAN, CIN, Director Identification	Identify shell companies, related-party transactions, common directors, and interconnected business entities involved in suspicious activities.
Geospatial Data	Latitude, Longitude, Business Category, Business Existence Verification	Business Address, Coordinates	Verify physical existence of businesses, identify unregistered commercial establishments, and detect fake or non-existent business locations.

Working Mechanism

Integration Process	Description	Expected Outcome
Taxpayer Profile Creation	Integrates GST, Income Tax, Banking, MCA, and Geospatial data into a single taxpayer profile using PAN as the primary key.	Comprehensive taxpayer intelligence repository.
Data Matching & Validation	Cross-verifies turnover, income, invoices, tax payments, and banking transactions across multiple databases.	Detection of inconsistencies and hidden economic activities.
Risk Scoring Engine	Uses AI and Machine Learning algorithms to assign risk scores based on taxpayer behavior and compliance history.	Identification of high-risk taxpayers.
Anomaly Detection	Detects unusual patterns such as fake ITC claims, circular trading, shell companies, and abnormal banking transactions.	Early identification of tax fraud and revenue leakage.

Alert Generation System	Generates automated alerts for suspicious cases exceeding predefined risk thresholds.	Faster investigation and enforcement action.
Decision Support Dashboard	Provides real-time reports, analytics, heat maps, and compliance monitoring tools for tax officers.	Improved audit selection and data-driven decision-making.

The proposed system integrates data from GST, Income Tax, banking institutions, corporate registries, and geospatial business databases to create a unified taxpayer profile. The primary integration key is the PAN, which links GST registrations, Income Tax records, bank transactions, and company information. GST data such as GSTIN, turnover, GSTR-1, GSTR-3B, e-Invoices, e-Way Bills, Input Tax Credit claims, and tax payments are mapped with Income Tax data including declared income, business income, TDS/TCS credits, audit reports, and ITR filings. Banking information such as transaction volumes, cash deposits, and international remittances is compared against reported turnover and income to identify discrepancies.

The system further integrates corporate information from company databases, including directors, shareholders, CIN, and registered addresses, to identify related entities, shell companies, and suspicious business networks. Geospatial data such as business location coordinates, business category, and physical existence verification are used to validate whether a registered taxpayer is genuinely operating at the declared location and to identify unregistered commercial establishments. Using artificial intelligence and machine learning techniques, the integrated data is analyzed to detect inconsistencies, hidden relationships, unusual transaction patterns, and potential tax evasion risks. The system automatically generates risk scores and alerts when significant mismatches are identified between business turnover, income declarations, banking transactions, and operational activities. The outcome is a comprehensive tax intelligence framework that improves taxpayer profiling, enhances fraud detection, strengthens compliance monitoring, and enables early identification of revenue leakage and tax evasion.

4. Early Warning Indicators

The proposed Early Warning System continuously analyzes integrated GST, Income Tax, banking, e-Way Bill, e-Invoice, and business intelligence data to identify taxpayers exhibiting unusual or high-risk behavior. The system generates alerts when significant discrepancies are detected between GST turnover and income declared in Income Tax returns, indicating possible suppression of income or tax evasion. It also identifies taxpayers claiming unusually high Input Tax Credit (ITC) without corresponding purchase transactions, which may suggest fake invoicing or fraudulent credit claims.

The framework further examines businesses reporting continuous losses despite substantial sales turnover, as well as cases where large banking transactions are not reflected in GST or Income Tax returns. Multiple GST registrations linked to the same PAN, suspicious e-Way Bill movements, and abnormal transaction patterns compared with industry benchmarks are also treated as risk indicators. Additionally, the system can identify potential shell companies through network analysis of common directors, addresses, bank accounts, and business relationships.

By integrating geospatial business intelligence and local business databases, the system can detect active commercial establishments that are not registered under GST or are not filing returns despite ongoing business activities. Using machine learning and risk-scoring algorithms, each taxpayer is assigned a risk level, and high-risk cases are automatically flagged for further verification, audit, or investigation. This proactive approach enables tax authorities to detect revenue leakage at an early stage and take timely corrective action before significant tax losses occur.

3. Expected Findings

The integration of GST and Income Tax data through an AI-based tax intelligence framework can significantly improve the identification of tax evasion and fraudulent activities. The system is anticipated to enhance taxpayer risk classification, increase overall tax compliance, and reduce instances of fake invoicing and revenue leakage. By utilizing integrated government data sources, the framework can support faster and more accurate detection of high-risk taxpayers, leading to more effective audit selection and enforcement actions. The study is also expected to show improved coordination between tax departments, increased efficiency in tax collection processes, and better utilization of data-driven insights for policy formulation and strategic decision-making. Ultimately, the findings may contribute to

strengthening revenue protection, promoting transparency, and supporting a more efficient and intelligent tax administration system in India.

4. Results and Discussion

The proposed AI-Based Integrated Tax Intelligence and Early Warning System is expected to significantly enhance revenue protection by improving the detection of tax evasion and non-compliance. Through the integration of GST and Income Tax data, the system can identify fraudulent activities such as fake Input Tax Credit (ITC) claims, circular trading networks, shell companies, and under-reporting of income. The implementation of predictive analytics and risk-based monitoring is expected to improve voluntary compliance among taxpayers and increase the consistency of return filing. The effectiveness of the system can be evaluated using performance metrics such as accuracy, precision and recall. Accuracy measures the overall correctness of the model's predictions, precision evaluates the ability to correctly identify fraudulent taxpayers, and recall assesses the system's capability to detect the maximum number of fraud cases. The expected outcomes include improved audit efficiency, faster identification of high-risk taxpayers, reduced revenue leakage, and strengthened tax administration through data-driven decision-making.

Predictive - Revenue Protection Results (Sample)

Parameter	Before AI Integration	After AI Integration	Improvement
Tax Evasion Cases Detected	2,500	4,200	68% Increase
Revenue Leakage Identified (₹ Crore)	850	1,650	94% Increase
High-Risk Taxpayers Identified	12,000	21,500	79% Increase
Audit Success Rate	32%	71%	122% Increase
Voluntary Compliance Rate	78%	89%	14% Increase
Return Filing Consistency	82%	93%	13% Increase
Average Investigation Time	90 Days	35 Days	61% Reduction
Additional Revenue Realized (₹ Crore)	1,200	2,850	138% Increase

Sample Tax Evasion Findings- Prediction

Risk Indicator	AI System Finding	Action Recommended	Estimated Revenue Impact
High GST Turnover but Low Income Declaration	GST Turnover ₹15 Crore, Income Tax Declaration ₹18 Lakh	Detailed Audit	₹45 Lakh
Fake ITC Claims	ITC Claimed ₹80 Lakh without corresponding purchases	ITC Verification	₹80 Lakh
Circular Trading Network	12 interconnected entities with repeated invoice cycles	Investigation	₹1.2 Crore
Shell Company Detection	Multiple GST registrations at same address	Registration Verification	₹65 Lakh
Unreported Banking Transactions	Bank credits ₹5 Crore not reflected in GST returns	Financial Scrutiny	₹90 Lakh
Suspicious E-Way Bill Activity	Goods movement without matching invoices	Enforcement Action	₹35 Lakh

Non-Filer with Active Business Operations	Business transactions detected but no GST filing	Registration Compliance	₹25 Lakh
Geospatial Mismatch	Registered business not operating at declared location	Physical Verification	₹15 Lakh

5. Advantages and Benefits

The proposed Artificial Intelligence-Based Integrated Tax Intelligence and Early Warning System offers significant benefits to the Government, tax administration, taxpayers, and society. For the Government, the system can enhance revenue collection by identifying tax evasion, reducing revenue leakage, improving enforcement efficiency, and supporting evidence-based policy formulation. The integration of GST and Income Tax databases enables automated compliance monitoring, real-time risk assessment, and timely detection of suspicious transactions and non-compliant taxpayers.

For tax administration, the system facilitates risk-based audit selection, minimizes manual verification efforts, accelerates investigation processes, and provides real-time intelligence for decision-making. AI-driven analytics, predictive risk scoring, and automated alert mechanisms help tax officers focus on high-risk taxpayers, thereby improving operational efficiency and resource utilization. The system can automatically generate SMS alerts, email notifications, mobile application alerts, and dashboard-based warnings for tax officers whenever high-risk transactions, tax mismatches, fake ITC claims, abnormal banking activities, or potential tax evasion cases are detected. These alerts enable faster intervention, timely verification, and more effective enforcement actions. The framework can also provide automated compliance reminders and advisory notifications to taxpayers through SMS and email before return due dates, tax payment deadlines, or when inconsistencies are identified in filed returns. Such proactive communication can improve voluntary compliance, reduce unintentional errors, and enhance taxpayer awareness. From a societal perspective, the system promotes a fair, transparent, and accountable taxation environment by ensuring that all taxpayers contribute their legitimate share of taxes. It can help reduce black money circulation, discourage fraudulent business practices, detect shell companies and fake invoicing networks, and increase public revenue available for infrastructure development, healthcare, education, and social welfare programs. Ultimately, the framework contributes to stronger economic governance, improved taxpayer compliance, enhanced revenue protection, and sustainable national development through intelligent, technology-driven tax administration.

6. Conclusion

The proposed Artificial Intelligence-Based Integrated Tax Intelligence and Early Warning System represents a transformative approach to modern tax administration in India. By integrating GST, Income Tax, banking, customs, corporate, and geospatial data sources, the framework enables real-time risk assessment, predictive analytics, and proactive tax enforcement. The implementation of such a system can significantly reduce tax evasion, improve compliance, enhance revenue protection, and strengthen India's digital governance ecosystem. The research contributes towards the development of a comprehensive National Tax Intelligence Platform capable of supporting future data-driven tax administration and economic governance initiatives.

References :

1. Allingham, M.G. & Sandmo, *Income Tax Evasion: A Theoretical Analysis*
2. Slemrod, Joel , *Cheating Ourselves: The Economics of Tax Evasion*
3. Organisation for Economic Co-operation and Development (OECD), *AI in Tax Administration: Governing with Artificial Intelligence*
4. Zumaya, M., Lozano, S., and Co-authors, *Identifying Tax Evasion in Mexico with Tools from Network Science and Machine Learning*

5. Alexopoulos, T., Dellaportas, P., and Co-authors, *A Network and Machine Learning Approach to Detect Value Added Tax Fraud*
6. Datta, S., Thakkar, A., & Dabhade, *Combating GST Input Tax Credit Fraud in India: Forensic Accounting, Data Analytics, and Legal Enforcement Strategies*
7. Russell, Stuart & Norvig, Peter, *Artificial Intelligence: A Modern Approach*
8. Han, Jiawei, Kamber, Micheline & Pei, Jian, *Data Mining: Concepts and Techniques*
9. Newman, Mark, *Networks*
10. Mayer-Schönberger, Viktor & Cukier, Kenneth , *Big Data: A Revolution That Will Transform How We Live, Work, and Think*