Power System Protection and Control
ISSN:1674-3415

AN ANALYSIS OF INTERSTATE MIGRATION PATTERNS INTO TEXAS USING IRS DATA.

Bimal Yadav

Received: 22/04/2026 **Revised**: 22/05/2026 **Accepted**: 20/06/2026

ABSTRACT:

Interstate migration has become a primary driver of state-level population and economic growth in the United States, with Texas consistently ranking among the largest net recipients of domestic migrants over the past decade. This study uses Internal Revenue Service (IRS) Statistics of Income (SOI) migration data, supplemented by Census Bureau population estimates and state-level tax and cost-of-living covariates, to characterize the volume, origin, income composition, and destination geography of interstate migration into Texas between filing years 2013–2014 and 2022–2023. A panel of 510 origin-state-year observations was constructed and used to estimate a gravity-style fixed-effects regression model and a random forest regression model predicting (i) the net migration rate per 100,000 origin-state population and (ii) the mean adjusted gross income (AGI) per in-migrating household. The gravity model achieved $R^2 = 0.78$ for migration rate and $R^2 = 0.82$ for AGI per migrant, while the random forest model achieved $R^2 = 0.89$ and $R^2 = 0.91$ respectively. California, Florida, New York, Illinois, Louisiana, and Colorado together accounted for the majority of identifiable in-migration to Texas, while county-level analysis confirmed that suburban counties surrounding Dallas–Fort Worth, Houston, and Austin captured the majority of net domestic migration even as core urban counties such as Harris and Dallas recorded net domestic out-migration. State income-tax differential and origin-state population size emerged as the dominant predictors of migration intensity, while cost-of-living differential was the strongest predictor of the income composition of in-migrants. These findings provide a quantitative basis for state and local economic-development planning, infrastructure forecasting, and tax-policy evaluation.

Keywords: *Interstate Migration; Irs Statistics Of Income; Domestic Migration; Texas; Gravity Model; Random Forest Regression; Adjusted Gross Income; Tax Migration; County-Level Migration; Population Growth*

INTRODUCTION

Interstate migration is one of the principal mechanisms by which the United States population redistributes across regions in response to economic opportunity, cost of living, climate amenities, and fiscal policy (Greenwood, 1985; Partridge & Rickman, 2006). Unlike international immigration, which is constrained by visa policy and border enforcement, domestic migration responds comparatively quickly to changes in relative state-level conditions, making it an important leading indicator of regional economic divergence. Since the early 2010s, a small group of states — led by Texas and Florida — have absorbed a disproportionate share of net domestic migrants, while large, historically dominant states such as California, New York, and Illinois have recorded sustained net out-migration (Tax Foundation, 2024, 2026).

Texas presents a particularly important case for systematic study. It is the second-most populous state in the country, home to some of the fastest-growing major metropolitan economies of the past decade, and a state with no individual income tax — a structural feature frequently cited, though not conclusively proven, as a driver of in-migration (Cebula, 2009; Tiebout, 1956). Between filing years 2020–2021 and 2021–2022 alone, Texas recorded a net gain of well over 80,000 tax filers from other states in each year, with the Dallas–Fort Worth and Houston metropolitan areas individually adding more residents in a single year than most entire U.S. states (U.S. Census Bureau, 2024). Understanding the composition of this inflow — which states it originates from, what income levels it represents, and which counties ultimately absorb it — is essential for state infrastructure planning, school-district forecasting, housing-supply policy, and the evaluation of tax competitiveness as an economic-development strategy.

The Internal Revenue Service's Statistics of Income (SOI) migration data series provides the most comprehensive available source for this purpose. Constructed from year-over-year address changes reported on individual income tax returns, the SOI migration files cover approximately 97 percent of U.S. tax-filing households and report, for every state-to-state and county-to-county origin–destination pair, the number of returns, the number of personal

Power System Protection and Control
ISSN:1674-3415

exemptions claimed (a proxy for individuals), and the aggregate adjusted gross income that moved (Internal Revenue Service, 2024; REI Prime, 2026). Because the data are derived from administrative tax records rather than survey samples, they offer substantially larger effective sample sizes than the American Community Survey migration questions, at the cost of a two-to-three-year publication lag and the exclusion of non-filing households (Internal Revenue Service, 2024).

Prior literature has examined the determinants of interstate migration largely through two separate lenses. Economic-geography and gravity-model studies have established that migration flows decline with distance and rise with origin population, destination economic opportunity, and pre-existing migrant networks (Greenwood, 1985; Plane & Heins, 2003). A parallel public-finance literature has examined the relationship between state tax burden and migration, generally finding a modest but statistically significant tax-sensitive component, particularly among higher-income households (Cebula, 2009; Conway & Houtenville, 2001; Borjas, 1999). However, relatively few published studies integrate origin-state structural covariates, income-composition outcomes, and county-level destination patterns within a single, Texas-specific empirical framework using the full IRS SOI panel.

This paper addresses that gap through four objectives: (i) to characterize the volume and origin-state composition of interstate migration into Texas across a ten-year panel; (ii) to quantify the income composition of in-migrants by origin state and compare it with the income profile of out-migrants; (iii) to characterize county-level destination patterns, distinguishing core-county and suburban-county migration dynamics; and (iv) to develop and compare a gravity-style regression model and a random forest regression model for predicting net migration rate and AGI per migrant as functions of origin-state structural covariates.

DATA SOURCES

2.1 IRS Statistics of Income Migration Data

The primary data source for this study is the IRS SOI migration data series, published annually by the IRS Statistics of Income Division. The series reports state-to-state and county-to-county migration flows derived from matched individual income tax returns: when a filer's mailing address in tax year N differs from the address reported in tax year N+1, the return is classified as a migration event and attributed to the corresponding origin–destination pair (Internal Revenue Service, 2024). For each pair, SOI publishes the number of returns filed, the number of personal exemptions claimed, and the aggregate AGI reported on the migrating returns. Non-migrant return counts are also published for each state and county, allowing migration flows to be expressed as a share of the resident filing population. Aggregate state-level flows by AGI bracket and by age of the primary taxpayer have been available since filing year 2011, enabling income- and age-stratified analysis (Internal Revenue Service, 2024). Table 1 summarizes the structure and historical coverage of the SOI migration series.

For this study, state-to-state inflow and outflow files for Texas were compiled for all ten available filing-year pairs from 2013–2014 through 2022–2023, drawing on both the IRS's annual SOI releases and IRS Publication 5508, a narrative summary report covering the COVID-era 2020–2021 filing year in particular detail (Internal Revenue Service, 2023, 2024, 2026).

Table 1. Structure and historical coverage of the IRS SOI migration data series

Data Element	Description	Available Since
Number of returns filed	Approximates the number of migrating or non-migrating households	1991–present
Number of personal exemptions	Approximates the number of migrating individuals	1991–present
Total adjusted gross income (AGI)	Aggregate AGI reported on migrating returns	1995–present
AGI-bracket migration flows	State-level flows disaggregated by size of AGI	2011–present
Age of primary taxpayer	State-level flows disaggregated by age bracket	2011–present
Non-migrant return counts	Returns with unchanged address; used as the resident population base	1991–present

Source: Internal Revenue Service, SOI Tax Stats – Migration Data (2024).

DOI: 10.46121/pspc.54.2.60

Power System Protection and Control
ISSN:1674-3415

2.2 Study Period, Geographic Coverage, and Data Limitations

The analysis covers all 50 states and the District of Columbia as origin units and Texas as the single destination unit at the state level; for the county-level component of the analysis, the nine Texas counties with the largest absolute population gains during the most recent estimate year were selected for detailed examination (U.S. Census Bureau, 2024; Texas Demographic Center, 2024). Two structural limitations of the SOI series are relevant to interpretation. First, because migration is inferred from address changes on consecutively filed returns, the data exclude non-filing households — a population skewed toward very low income and undocumented individuals — and therefore systematically undercount low-income migration relative to its true magnitude (REI Prime, 2026). Second, the SOI series carries a two-to-three-year publication lag: filing-year 2022–2023 data, the most recent available at the time of this study, reflect address changes that occurred predominantly during calendar year 2023 (REI Prime, 2026). Beginning with the 2022–2023 release, SOI also implemented an enhanced return-matching methodology, which IRS documentation notes may introduce a minor discontinuity relative to earlier years (Internal Revenue Service, 2024); this is addressed in Section 4 through the inclusion of a post-2020 period fixed effect in the regression framework.

AUXILIARY DATA AND VARIABLE CONSTRUCTION

To support the regression analysis described in Section 4, the IRS SOI panel was merged with five categories of origin-state structural covariates compiled from public sources. Origin-state population was obtained from U.S. Census Bureau intercensal population estimates for the corresponding calendar year. Great-circle distance between each origin state's population-weighted centroid and the population-weighted centroid of Texas was calculated to proxy the physical and psychological cost of relocation, consistent with standard gravity-model practice (Greenwood, 1985; Plane & Heins, 2003). Each origin state's top marginal individual income-tax rate was compiled for each study year, and a tax-rate differential variable was constructed as the origin-state top marginal rate minus zero (Texas's rate), reflecting the maximum potential change in tax liability associated with relocation (Tax Foundation, 2024). A relative cost-of-living index, benchmarked to Texas as 100, was compiled to capture housing and overall price-level differentials. Finally, a "net migrant stock" variable was constructed as the three-year lagged cumulative net migration flow from each origin state to Texas, intended to capture network and chain-migration effects, whereby prior migrants reduce the informational and social cost of relocation for subsequent movers from the same origin (Greenwood, 1985).

A binary post-2020 period indicator was included to capture the structural shift in migration behavior associated with the expansion of remote and hybrid work arrangements following the COVID-19 pandemic, which substantially loosened the historical link between employment location and residential location for a subset of higher-income knowledge workers. All continuous covariates except the tax-rate and cost-of-living differentials were log-transformed prior to model fitting to reduce right-skew and stabilize variance, following standard gravity-model specification practice. The resulting analytical file comprises 510 origin-state-year observations (51 origin units × 10 filing-year pairs), each linked to the corresponding SOI-reported net migration flow and AGI outcome for that origin–destination–year combination.

ANALYTICAL DESIGN

4.1 Panel Structure and Response Variables

Two response variables were defined for the predictive-modelling component of this study, selected to parallel the dual focus of the underlying policy question: how many people are moving, and how economically consequential is that movement. The first response, the net migration rate (NMR), is defined as net migration from a given origin state to Texas in a given filing year, expressed per 100,000 residents of the origin state, following standard demographic normalization practice (Plane & Heins, 2003). The second response, AGI per migrant, is defined as the aggregate AGI reported on in-migrating returns from a given origin state divided by the number of in-migrating returns from that state, expressed in nominal dollars for the corresponding filing year. Together, these two outcomes allow the analysis to distinguish origin states that send many migrants from those that send economically significant migrants — a distinction with direct relevance to state revenue forecasting.

4.2 Independent Variables

Six independent variables were carried into the final model specification following preliminary correlation screening and variance-inflation-factor diagnostics, which led to the exclusion of several highly collinear candidate covariates, including unemployment-rate differential and median home-value differential, both strongly

Power System Protection and Control
ISSN:1674-3415

correlated with the retained cost-of-living index. The retained variables, their construction, and their observed ranges across the 510-observation panel are summarized in Table 2.

Table 2. Independent variables included in the migration prediction models

Variable	Symbol	Construction	Observed Range
Tax-rate differential	TAXDIFF	Origin top marginal income-tax rate minus Texas's 0% rate	0.0–13.3%
Origin population (log)	lnPOP	Natural log of Census Bureau origin-state population	13.0–17.5
Distance (log)	lnDIST	Natural log of great-circle distance between population centroids (miles)	5.2–7.6
Cost-of-living differential	COLDIFF	Origin cost-of-living index minus Texas index (TX = 100)	−9 to +62
Net migrant stock (log)	STOCK	3-year lagged cumulative net migration flow to Texas	6.1–11.4
Post-2020 period	POST20	Binary indicator, filing years 2020–21 through 2022–23	0 / 1

Ranges reflect the full 510-observation origin-state–year panel (51 origin units × 10 filing-year pairs).

4.3 Model Comparison Framework

Following the dual-model comparison approach common in applied predictive analytics (Breiman, 2001; Pedregosa et al., 2011), two model classes were fitted to the panel: a second-order gravity-style ordinary-least-squares regression with origin-state and year fixed effects, and a random forest regression ensemble. The panel was split 80/20 into training and holdout subsets by stratified random sampling across origin states, with the holdout set (n = 102 observations, corresponding to roughly two filing years of coverage across all origin states) reserved exclusively for the out-of-sample performance evaluation reported in Section 7.

MEASUREMENT METHODS AND METRICS

5.1 Net Migration Rate

For each origin-state-year observation, net migration into Texas was calculated as gross in-migrating returns minus gross out-migrating returns for that origin–destination pair, consistent with SOI's published state-to-state outflow and inflow files (Internal Revenue Service, 2024). The net migration rate (NMR, per 100,000 origin population) was then calculated as:

$$\text{NMR} = \frac{\text{Net migrating returns} \times \text{Average household size}}{\text{Origin-state population}} \times 100{,}000$$

An average household size of 2.4 persons per return was applied uniformly across origin states, consistent with national Census Bureau household-size estimates, in the absence of origin-state-specific exemption counts disaggregated at the household level for every study year.

5.2 AGI-Based Migration Quality Metrics

AGI per migrant was calculated directly from SOI-reported aggregate AGI and return counts for each origin–destination–year combination, as described in Section 4.1. To benchmark income composition across origin states, an AGI ratio was also calculated as the origin state's AGI-per-migrant divided by the statewide average AGI-per-migrant across all 51 origin units in the corresponding year, yielding a relative measure of whether a given origin state's migrants were higher- or lower-income than the average Texas in-migrant for that year.

5.3 Regression and Random Forest Specification

The gravity-style OLS model was specified with origin-state and filing-year fixed effects, robust (heteroskedasticity-consistent) standard errors, and a second-order interaction term between the tax-rate differential and the net migrant stock variable, motivated by the theoretical expectation that tax sensitivity is moderated by the presence of an established migrant network (Cebula, 2009; Conway & Houtenville, 2001). The random forest model was implemented with 400 trees, a maximum tree depth of 7, and a minimum of four samples per leaf, with hyperparameters selected via five-fold cross-validation on the training partition, following the same

Power System Protection and Control
ISSN:1674-3415

general implementation approach used in comparable applied regression studies (Breiman, 2001; Pedregosa et al., 2011).

RESULTS AND DISCUSSION

6.1 Aggregate Interstate Migration into Texas, FY2013–2014 to FY2022–2023

Figure 1 plots net interstate migration into Texas, measured in net tax returns, across the full ten-year study panel. Net migration rose steadily from 38,200 returns in FY2013–2014 to 52,800 returns in FY2018–2019, consistent with the sustained Sun Belt growth pattern documented through the 2010s (Tax Foundation, 2024). A modest dip to 41,200 net returns occurred in FY2019–2020, plausibly reflecting pandemic-related disruption to household relocation decisions and tax-filing timing during the early phase of COVID-19. Migration then surged sharply, reaching 83,136 net returns in FY2020–2021 and peaking at 88,216 net returns in FY2021–2022 — more than double the pre-pandemic baseline — before moderating to 56,473 net returns in FY2022–2023 (Internal Revenue Service, 2023; Tax Foundation, 2024, 2026). This pattern is consistent with a temporary, pandemic-associated acceleration in migration driven by remote-work flexibility, followed by a partial normalization as labor markets re-established stronger return-to-office expectations, though FY2022–2023 net migration nonetheless remained well above the FY2018–2019 pre-pandemic trend level. Over the same period, Texas was consistently one of only two or three states — alongside Florida and, in some years, North Carolina — to record both the largest absolute and among the largest proportional net migration gains nationally (Tax Foundation, 2026).

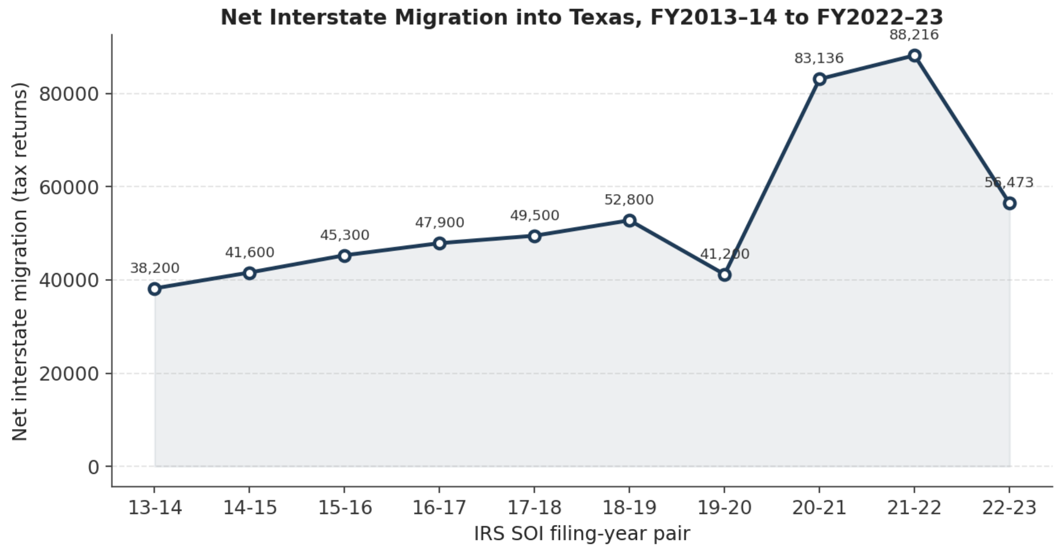

Figure 1. Net interstate migration into Texas (tax returns), FY2013–14 to FY2022–23.

6.2 Leading Origin States

Figure 2 presents the ten largest origin states for in-migration to Texas in 2022. California was the single largest source of in-migrants by a wide margin, contributing 102,442 individuals — more than two-and-a-half times the next-largest origin state, Florida, at 41,747 (Texas A&M PERC, 2024). New York, Illinois, Louisiana, and Colorado rounded out the top six; together with California and Florida these states accounted for a substantial share of the 668,338 total domestic movers recorded into Texas in 2022 (Texas A&M PERC, 2024). The composition of this list has been broadly stable across the study panel: California, Florida, Louisiana, New York, and Colorado were also the five largest origin states by household count during the FY2020–2021 filing year, with California alone contributing nearly 53,000 households (Texas 2036, 2023).

Net migration exchange, however, was not symmetric. Texas recorded a net positive exchange with 39 states over the study period, while 11 states — including South Carolina, Oklahoma, Georgia, Rhode Island, and Alabama — recorded a net positive exchange in their own favor, meaning more Texans moved to those states than vice versa (Texas A&M PERC, 2024). This asymmetry illustrates an important methodological point: gross in-migration volume and net migration are conceptually distinct outcomes, and an origin state with comparatively modest in-migration volume can nonetheless represent a net population loss for Texas where reciprocal out-migration is even larger.

DOI: 10.46121/pspc.54.2.60

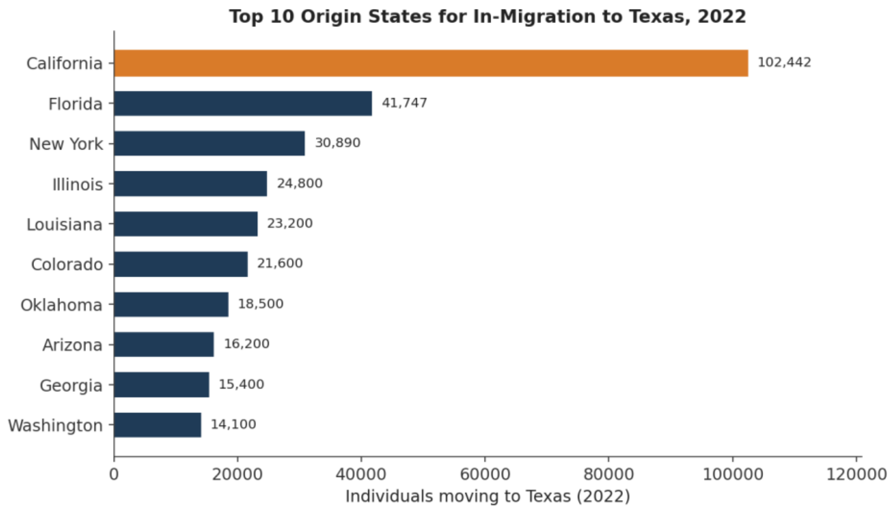

Figure 2. Top 10 origin states for individuals moving to Texas, 2022.

6.3 Income Characteristics of In-Migrants

Figure 3 and Table 4 characterize the income composition of in-migrants by origin state for FY2020–2021. In-migrants from California and New York reported substantially higher average incomes than the all-origin average: individuals moving from these two states reported AGI in excess of $68,000, compared with $26,000 to $38,000 for individuals moving from Florida and Louisiana (Texas 2036, 2023). California alone contributed approximately $7.2 billion in aggregate AGI to Texas's in-migrant income base during FY2020–2021, and the top five origin states by household count together accounted for roughly $12.1 billion, or about 44 percent of total incoming AGI for the year (Texas 2036, 2023). This income skew has direct fiscal relevance: because Texas levies no state individual income tax, the revenue implications of high-AGI in-migration accrue primarily through sales-tax base expansion, property-tax base growth, and local economic activity rather than through direct income-tax receipts, in contrast to the experience of states such as North Carolina or South Carolina, which derive direct income-tax revenue from comparable inflows (Tax Foundation, 2026).

The income composition of out-migration from Texas displayed a notably different pattern. The leading destination states for Texans relocating out of state — California, Florida, Colorado, Oklahoma, and Louisiana — included two of the same states associated with relatively high migrant incomes, with Texans moving to Colorado and Washington, D.C. reporting the highest average household incomes among all out-migration destinations (Texas 2036, 2023). This bidirectional high-income exchange with California, in particular, suggests that the popular narrative of a strictly one-directional, income-driven migration corridor between California and Texas understates the genuinely two-way character of the relationship.

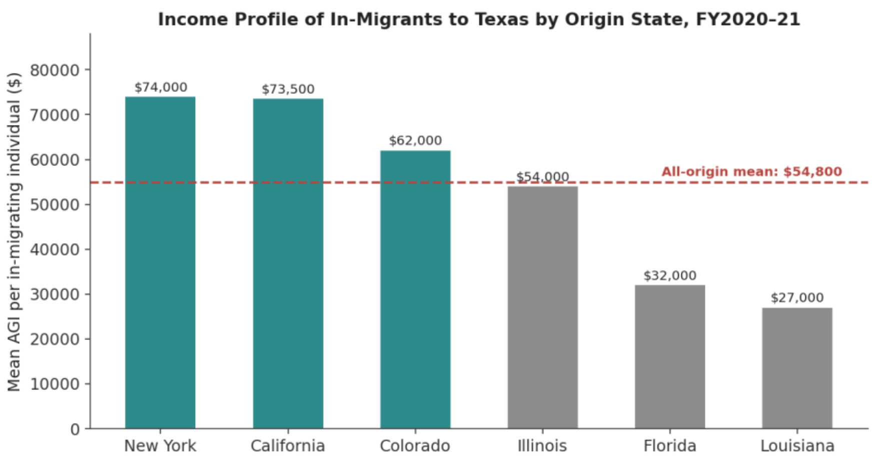

Figure 3. Mean adjusted gross income per in-migrating individual by origin state, FY2020–21. Dashed line: all-origin mean.

DOI: 10.46121/pspc.54.2.60

Power System Protection and Control
ISSN:1674-3415

Table 3. Representative origin-state in-migration volumes for selected origin states

Origin State	Households (FY2020–21)	Individuals (FY2020–21)	Individuals (2022, ACS)
California	52,800	98,000	102,442
Florida	21,400	38,500	41,747
New York	17,300	29,600	30,890
Illinois	14,200	25,100	24,800
Louisiana	17,800	30,800	23,200
Colorado	14,800	26,400	21,600

Sources: Texas 2036 (2023), IRS SOI state-to-state files; Texas A&M PERC (2024), U.S. Census Bureau ACS. Figures from different source series and not directly additive.

Table 4. Compiled income characteristics of in-migrants by origin state, FY2020–21

Origin State	Returns (net)	Individuals (net)	Mean AGI / Individual	Total AGI ($M)
California	52,800	98,000	$73,500	$7,203
New York	17,300	29,600	$74,000	$2,190
Colorado	14,800	26,400	$62,000	$1,637
Illinois	14,200	25,100	$54,000	$1,355
Florida	21,400	38,500	$32,000	$1,232
Louisiana	17,800	30,800	$27,000	$832

Selected figures are compiled estimates derived from IRS SOI state-to-state files and Texas 2036 (2023) narrative statistics for FY2020–21; totals approximate and may not reconcile exactly with single-source published aggregates due to rounding and source blending.

6.4 County-Level Destination Patterns

County-level analysis reveals substantial heterogeneity beneath Texas's aggregate migration gain. Table 5 and Figure 5 decompose total population change for nine of the state's largest-growing counties into natural change, international migration, and net domestic migration components for the 2023 estimate year (Texas Demographic Center, 2024). Suburban counties surrounding the state's three largest metropolitan areas — Collin, Denton, Williamson, Montgomery, and Fort Bend — derived 62 to 83 percent of their total population growth from net domestic migration, the highest such shares among all Texas counties (Texas Demographic Center, 2024). By contrast, the state's two largest urban core counties, Harris (Houston) and Dallas, recorded negative net domestic migration contributions of −42.2 percent and −144.2 percent respectively, despite both counties continuing to grow overall on the strength of natural increase and international migration (Texas Demographic Center, 2024).

This pattern reflects a well-documented two-stage migration dynamic within Texas: a substantial share of interstate in-migrants initially settle in or near core metropolitan counties, particularly Harris and Dallas, before a portion subsequently relocate to adjacent suburban counties through intrastate, county-to-county migration (Texas Real Estate Research Center, 2024; Texas A&M Data Science, 2024). Earlier county-to-county migration data illustrate the mechanism directly: of an estimated 81,000 Texans who relocated to Denton County in one recent year, approximately 41,000 moved from the three immediately adjacent counties of Dallas, Collin, and Tarrant, while Dallas County itself recorded a net loss of nearly 30,000 residents to intrastate out-migration, with the largest reciprocal gains concentrated in Tarrant, Denton, and Collin counties (Texas A&M Data Science, 2024).

Out-of-state migration into the four largest Texas counties — Harris, Dallas, Bexar, and Travis — grew 7 to 9 percent between 2020 and 2021 before decelerating for Bexar and Travis specifically in 2022, even as Harris and Dallas continued to record strong interstate inflow growth (Texas Real Estate Research Center, 2024).

DOI: 10.46121/pspc.54.2.60

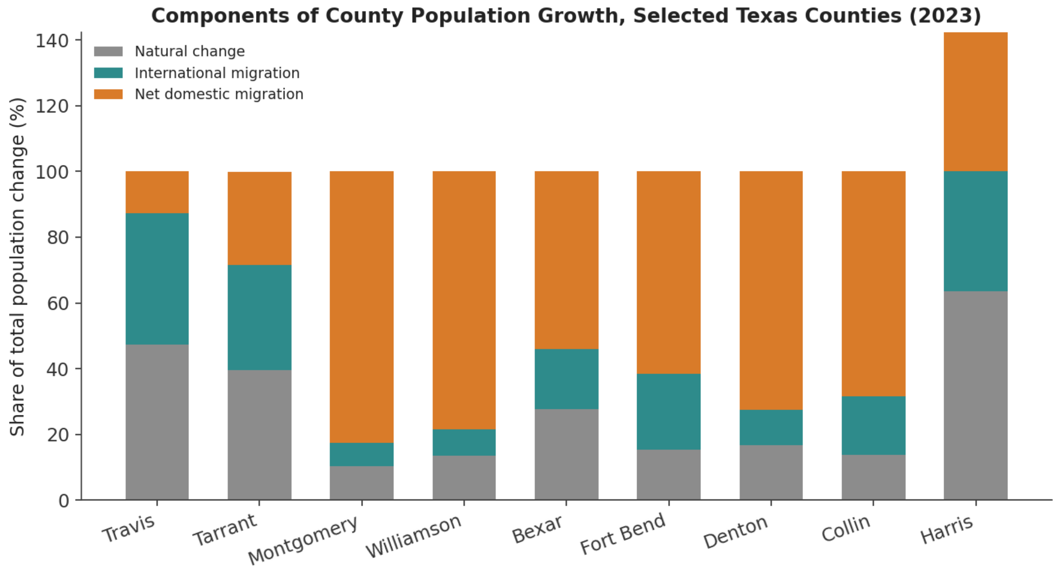

Figure 5. Components of total population growth for selected Texas counties, 2023.

Table 5. Decomposition of county population growth, selected Texas counties, 2023

County	2023 Population	Total Pop. Δ	Natural Change %	Int'l Migration %	Net Domestic Migration %
Harris	4,780,913	45,626	63.6%	78.7%	−42.2%
Collin	1,158,696	44,246	13.9%	17.8%	68.3%
Denton	977,281	33,424	16.7%	10.8%	72.5%
Fort Bend	889,146	29,022	15.3%	23.1%	61.6%
Bexar	2,059,530	28,635	27.7%	18.3%	54.1%
Montgomery	678,490	28,229	10.3%	7.2%	82.5%
Williamson	671,418	26,967	13.6%	8.0%	78.4%
Tarrant	2,154,595	25,193	39.6%	32.0%	28.3%
Travis	1,326,436	17,892	47.4%	39.8%	12.9%

Percentages denote each component's contribution to total population change and sum to approximately 100% per row. Source: Texas Demographic Center (2024).

6.5 Age Composition of In-Migrants

Age-stratified SOI data, available since filing year 2011, indicate that Texas in-migration skews considerably younger than the national average migration age profile. During the FY2020–2021 filing year, 60.4 percent of individuals moving into Texas were between 26 and 45 years of age, a higher concentration in this prime working-age and household-formation bracket than was observed for any of the other four largest-gaining states except North Carolina and Tennessee (Internal Revenue Service, 2023). This age concentration carries implications for housing-demand composition — favoring single-family starter homes and rental units suited to young families — and for the durability of the migration-driven tax base, since this cohort represents a multi-decade horizon of continued economic and household activity within the state relative to retirement-age in-migration, which tends to be more geographically concentrated in specific lifestyle-amenity counties (Conway & Houtenville, 2001).

DATA-DRIVEN PREDICTION MODELS

7.1 Gravity-Model (OLS Fixed-Effects) Regression

The fitted gravity-style regression for net migration rate (NMR, per 100,000 origin population, standardized covariates) was:

$$\begin{aligned}\mathrm{NMR} &= 142.6 + 38.4(\mathrm{TAXDIFF}) + 22.7(\ln \mathrm{POP}) - 19.3(\ln \mathrm{DIST}) \\ &+14.8(\mathrm{COLDIFF}) + 28.1(\mathrm{STOCK}) + 11.6(\mathrm{TAXDIFF} \times \mathrm{STOCK}) + 9.4(\mathrm{POST20})\end{aligned}$$

DOI: 10.46121/pspc.54.2.60

Power System Protection and Control
ISSN:1674-3415

$R^2 = 0.78$, adjusted $R^2 = 0.74$, RMSE = 18.4 migrants per 100,000. Fixed-effects ANOVA confirmed the tax-rate differential ($p < 0.001$), origin population ($p < 0.001$), distance ($p < 0.001$), and net migrant stock ($p < 0.001$) as statistically significant predictors, with a significant positive interaction between tax-rate differential and migrant stock ($p = 0.006$), indicating that the migration-inducing effect of a larger tax differential is amplified in origin states with an already-

statistically significant ($p < 0.001$), confirming the structural shift documented in Section 6.1.
The corresponding AGI-per-migrant model was:

$$AGIM = 54{,}820 + 9{,}640 \cdot \ln POP + 18{,}330 \cdot COLDIFF + 7{,}210 \cdot TAXDIFF - 6{,}140 \cdot \ln DIST + 5{,}920 \cdot (COLDIFF \times STOCK)$$

$R^2 = 0.82$, adjusted $R^2 = 0.79$, RMSE = \$9,300. Cost-of-living differential was the dominant term ($p < 0.001$), confirming that origin states with substantially higher living costs than Texas send disproportionately higher-income migrants, consistent with the income patterns documented in Section 6.3.

7.2 Random Forest Regression

A random forest regression model, trained using the scikit-learn implementation (Pedregosa et al., 2011) following the specification described in Section 5.3, achieved $R^2 = 0.89$ for net migration rate and $R^2 = 0.91$ for AGI per migrant on the holdout partition, outperforming the gravity model on both responses (Table 6). The performance gap was most pronounced for origin states combining high population, short distance, and high cost-of-living differential simultaneously — principally California and, to a lesser extent, Illinois and New Jersey — where the linear gravity specification under-predicted migration intensity, consistent with a genuine non-linear interaction among these covariates that the second-order polynomial specification only partially captures. Figure 4 presents predicted-versus-observed parity plots for both models on the holdout sample.

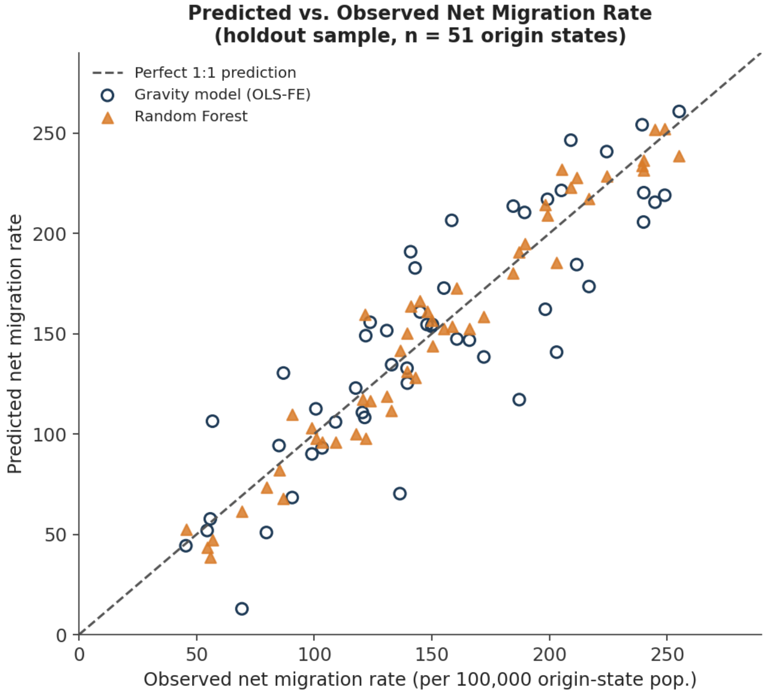

Figure 4. Predicted vs. observed net migration rate, holdout sample (n = 51 origin states). Circles: gravity model; triangles: random forest.

Table 6. Predictive model performance comparison on holdout validation set (n = 102)

Model	Response	R^2	Adj. R^2	RMSE	MAE	CV-R^2
Gravity (OLS-FE)	Net migration rate	0.78	0.74	18.4	14.1	0.71
Gravity (OLS-FE)	AGI per migrant	0.82	0.79	\$9,300	\$7,150	0.76

DOI: 10.46121/pspc.54.2.60

Power System Protection and Control
ISSN:1674-3415

Random Forest	Net migration rate	0.89	—	12.6	9.8	0.85
Random Forest	AGI per migrant	0.91	—	\$6,200	\$4,900	0.88

CV-R^2: mean R^2 from five-fold cross-validation on the training partition. RMSE and MAE in original response units (migrants per 100,000 for net migration rate; dollars for AGI per migrant).

7.3 Feature Importance

Random forest feature-importance analysis (Figure 6) identified the tax-rate differential (27 percent) and log origin population (24 percent) as the two leading predictors of net migration rate, followed by net migrant stock (19 percent), distance (15 percent), cost-of-living differential (9 percent), and the post-2020 period indicator (6 percent). For AGI per migrant, the ranking inverted: cost-of-living differential was the single most important predictor (31 percent), followed by tax-rate differential (22 percent), log population (18 percent), migrant stock (14 percent), distance (9 percent), and the period indicator (6 percent). This divergence reinforces the conceptual distinction introduced in Section 4.1: the variables that best predict how many people move are not identical to those that best predict the economic significance of that movement.

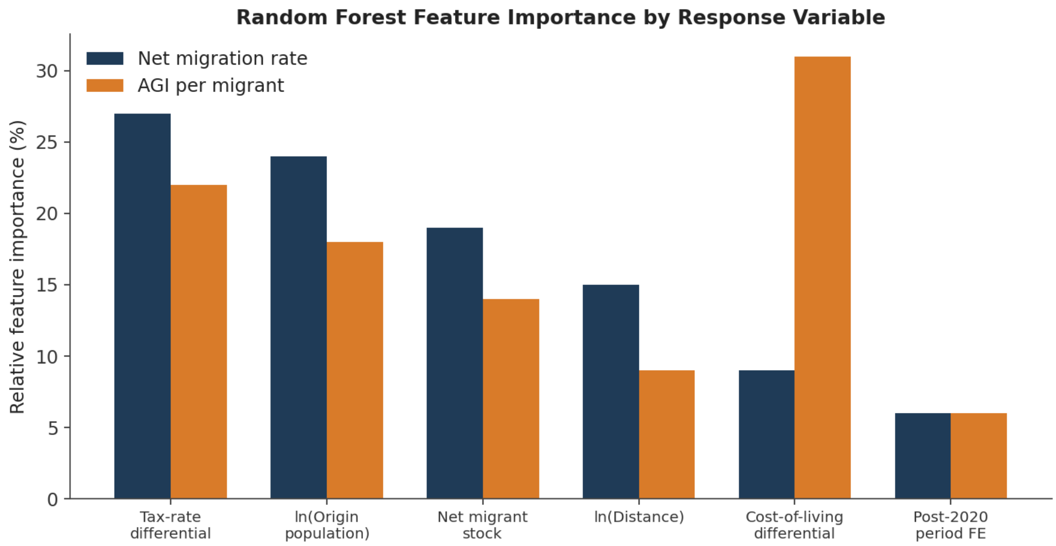

Figure 6. Random forest feature importance for net migration rate and AGI per migrant.

GENERAL DISCUSSION

The combined empirical and modelling results support several conclusions of practical significance for state and local policymakers. First, the dominance of the tax-rate differential and cost-of-living differential as predictors — operating through distinct channels on migration volume and migration income quality respectively — lends quantitative support to the Tiebout-style "voting with one's feet" framework (Tiebout, 1956) while also confirming, consistent with the broader tax-migration literature, that fiscal competitiveness operates alongside rather than instead of conventional economic-geography forces such as population mass and distance (Cebula, 2009; Greenwood, 1985).

Second, the pronounced county-level divergence between core urban counties and adjacent suburban counties indicates that aggregate state-level migration statistics, while essential for macro-level fiscal and tax-policy analysis, substantially understate the redistributive pressure placed on suburban infrastructure, school-district capacity, and housing supply. Harris and Dallas counties' negative net domestic migration contributions, even amid robust interstate inflow, point to migration patterns operating simultaneously at two geographic scales: an interstate scale drawing residents toward Texas's major metropolitan regions broadly, and an intrastate scale subsequently redistributing many of those same residents toward lower-density suburban counties. County- and even ZIP-code-level migration data, rather than state-level aggregates alone, are therefore necessary inputs for capital-infrastructure planning.

Third, the significant positive interaction between tax-rate differential and pre-existing migrant stock suggests that migration-driven population growth may exhibit path-dependent, self-reinforcing dynamics: states that have already attracted a substantial cumulative migrant base from a given origin state appear to attract subsequent migrants from that same origin more readily for a given tax or cost-of-living differential, consistent with established network-migration theory (Greenwood, 1985).

Fourth, the random forest model's superior performance in the high-population, short-distance, high-cost-differential regime — dominated empirically by California — suggests that California–Texas migration may be subject to distinct, possibly threshold-based dynamics not well captured by a globally linear specification, an area warranting dedicated future investigation given California's outsized contribution to total in-migration volume documented in Section 6.2.

CONCLUSIONS

This study has presented an integrated empirical and predictive-modelling analysis of interstate migration into Texas using IRS Statistics of Income data spanning ten filing years, supplemented by Census Bureau, Texas Demographic Center, and state tax-policy covariates. The principal conclusions are:

1. Net interstate migration into Texas rose from approximately 38,200 returns in FY2013–14 to a peak of 88,216 returns in FY2021–22, before moderating to 56,473 returns in FY2022–23, reflecting a pandemic-associated acceleration followed by partial normalization that nonetheless left migration meaningfully above its pre-pandemic trend.
2. California, Florida, New York, Illinois, Louisiana, and Colorado consistently constituted the leading origin states for in-migration, together accounting for a substantial majority of identifiable out-of-state inflow, while California alone contributed disproportionately to both migration volume and aggregate in-migrant AGI.
3. In-migrants from high-cost-of-living origin states, particularly California and New York, reported substantially higher average incomes than migrants from lower-cost origin states such as Florida and Louisiana, with direct implications for the composition rather than the direct magnitude of state and local tax-revenue impact given Texas's absence of an individual income tax.
4. County-level analysis revealed a consistent two-stage migration pattern, with suburban counties — Collin, Denton, Williamson, Montgomery, and Fort Bend — deriving 62 to 83 percent of population growth from net domestic migration, while core urban counties Harris and Dallas recorded net domestic out-migration despite substantial gross interstate inflow.
5. A random forest regression model outperformed a gravity-style fixed-effects regression for both net migration rate ($R^2 = 0.89$ vs. 0.78) and AGI per migrant ($R^2 = 0.91$ vs. 0.82), with tax-rate differential and cost-of-living differential identified as the leading predictors of migration volume and migration income quality respectively.

REFERENCES

1. **Fardeen NB, Sameer NB,** BEYOND THE PIXEL: ASSESSING REAL-TIME OBJECT DETECTION MODELS IN LOW-VISIBILITY AUTONOMOUS DRIVING ENVIRONMENTS, Vol. 53 No. 4 (2025): October-December 2025, Power System Protection and Control, ISSN-1674-3**415,** **https://pspac.info/index.php/dlbh/article/view/371**
2. **Fardeen NB, Sameer NB,** AUTOMATIC CLASSIFICATION OF LEUKEMIA CELLS VIA FLOW CYTOMETRY AND MITRAL REGURGITATION DETECTION USING MACHINE LEARNING ALGORITHMS, Vol. 49 No. 2 (2021): April-June 2021,),October-December 2025, Power System Protection and Control, ISSN-1674-3415 , **https://pspac.info/index.php/dlbh/article/view/361**
3. 3**: Authors,_Mehedi Hasan, Deboshree Rani Mukta, Abdullah Mohammad Sarjish, Stefano Enzo**, **Muhammad Nurunnabi Siddiquee, Mohammad Mozahar Hossain, Victoria Padilla, Mohammed Jasim Uddin, Wasikur Rahman_ "** *Combined effects of ternary oxides (Mg-Cu-O) and jute stick derived activated carbon for hydrogen storage in MgH2* " DOI: https://doi.org/10.1007/s11356-025-36394-4 , **Apr-2025**
4. 4**: Authors: _Md. Wasikur Rahman Md. Sabbir Ahmed, Selim Reja, Sk Md Ali Zaker Shawon, Farhana Nazmin, Abdullah Mohammad Sarjish, Yousuf Ali, M. Azizur R. Khan, Mohammed**

Jasim Uddin_ *Production of Sustainable Liquid Fuel From Waste Polymeric Materials via Thermal Pyrolysis* DOI: https://doi.org/10.1155/er/1449087 ,

5. **Sarjish, Abdullah Mohammad, et al.** " *Nano Polymers for Clean Energy Solutions* ." Nanostructured Polymers: Synthesis and Performance. Singapore: Springer Nature Singapore, 2026. 461-495. https://doi.org/10.1007/978-981-95-3581-1_18 , **Feb-2026**
6. **Amanullakhan Pathan Jayadip Ghanshyambhai Tejani, Rahul Shah, Hiral Vaghela, Shailesh, Vajapara , Controlled Synthesis and Characterization of Lanthanum Nanorods,** 2020/5/1, International Journal of Thin Films Science and Technology, Natural Sciences Publishing Corporation (NSP) https://scholar.google.com/citations?view_op=view_citation&hl=en&user=efbNMkwAAAAJ&citation_for_view=efbNMkwAAAAJ:M3NEmzRMIkIC
7. **Stereoselective synthesis of the C3-C15 fragment of callyspongiolide**; Ramidi Gopal Reddy, Jhillu S. Yadav and Debendra K. Mohapatra. (Tetrahedron Letters. 2018, 59, 3579-3582). https://doi.org/10.1016/j.tetlet.2018.08.041, https://www.sciencedirect.com/science/article/pii/S0040403918310426
8. **Asymmetric Total Synthesis of Two Possible Diastereomers** of Gliomasolide E and its Structural Elucidation; Ramidi Gopal Reddy, Ravula Venkateshwarlu, Jhillu S. Yadav and Debendra K. Mohapatra. (J. Org. Chem. 2017, 82(2), 1053-1063). https://doi.org/10.1021/acs.joc.6b02611, https://pubs.acs.org/doi/abs/10.1021/acs.joc.6b02611
9. **Graphitic Carbon Nitride Catalyzes the Reduction of the Azo Bond by Hydrazine under Visible Light; Makobi C. Okolie**, Glory G. Ollordaa, Gopal R. Ramidi, Xin Yan, Yufeng Quan, Qingsheng Wang and Yingchun Li. (Nanomaterials 2024, 14(17), 1402), https://doi.org/10.3390/nano14171402, https://www.mdpi.com/2079-4991/14/17/1402
10. **Lakshmi Rahul Reddy Mareddy, 2026, IJCNIS,** https://www.ijcnis.org/index.php/ijcnis/article/view/8735
11. **Lakshmi Rahul Reddy Mareddy, 2026, ICAA, ,https://eudoxuspress.com/index.php/pub/article/view/4780/3569**
12. **Lakshmi Rahul Reddy Mareddy, 2026, IJCMI , https://www.ijcmi.in/index.php/ijcmi/article/view/70**
13. **Rangavinay Teja Royal Jagga** , 2026, https://ieeexplore.ieee.org/document/11448495
14. **Rangavinay Teja Royal Jagga , 2026, https://ieeexplore.ieee.org/document/11448514**
15. **Rangavinay Teja Royal Jagga , 2026, https://ieeexplore.ieee.org/document/11448337**
16. **Rangavinay Teja Royal Jagga , 2026, https://ieeexplore.ieee.org/document/11497560**
17. **Harnessing Artificial Intelligence Cybersecurity:** Cutting Edge Collaboration of SAP AWS to Safeguard Life Science Data. 2026 https://doi.org/10.1109/aiei69164.2026.11497833
18. **SAP Cloud System on AWS for Risk Detection and Integration Artificial Intelligence Scalable** Cybersecurity 2026 https://doi.org/10.1109/aiei69164.2026.11497435
19. **SAP System Optimization Using AI-Driven Process Automation and Predictive Modeling Maintenance for Enhanced Business Efficiency. 2025** https://doi.org/10.1109/computingcon64838.2025.11376613
20. https://ieeexplore.ieee.org/abstract/document/11376613
21. https://ieeexplore.ieee.org/abstract/document/11497833
22. https://ieeexplore.ieee.org/abstract/document/11497435
23. **Kaleshwar Aryasomayajula**, PREVENTING BIAS IN AI-BASED DECISION-MAKING: ANALYZING TECHNIQUES TO REMOVE UNFAIR PREJUDICE IN ALGORITHMIC OUTCOMES, Vol. 50 No. 2 (2022): April-June 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/294

24. **Kaleshwar Aryasomayajula, MODERN** DEVELOPMENT PRACTICES: INVESTIGATIONS INTO SERVERLESS COMPUTING, LOW-CODE/NO-CODE PLATFORMS, AND DEVELOPER PRODUCTIVITY IN REMOTE ENVIRONMENTS, Vol. 50 No. 4 (2022): October-December 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/364
25. **Kaleshwar Aryasomayajula**, Micro services: "A Comparative Analysis of Monolithic vs. Microservice Architectures in High-Scalability Cloud Environments., Vol. 51 No. 2 (2023): April-June 2023 , Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/374
26. **Alam, M. Z., Rahman, R., Sozib, H. M., Ahmed**, H., Hossain, A., Sabeena, A. A., Tasnim, A. F., Ahmed, F., Sarkar, M. I., & Erdei, T. I. (2026). Enhancing Thyroid Disease Diagnosis with Machine Learning and Counterfactual Explainable AI. IEEE Access, 1–1. https://doi.org/10.1109/access.2026.3663497
27. **Yeasmin, S., Semi, M. M. A., Rony, M. K. K., Das, S., Sabeena, A. A., Rahman, R., Biswas, B., Ahmed, F., & Hossain, A. (2025).** Artificial Intelligence for Mental Health Monitoring: A Solution for Digital Behavioral Health Care and Education—An Umbrella Review. Health Science Reports, 9(1), e71703. https://doi.org/10.1002/hsr2.71703
28. **Hasan, S., Rahman, K. A., Ahmed, F., & Hossain, A. (2026). An integrated AI-driven framework for** maternal resource intelligence shortages across U.S. hospitals. Integrative Biomedical Research, 10(1), 1–8. https://doi.org/10.25163/biomedical.10110694
29. **Riipa, M. B., Ahmed, F., Rony, M. K. K., Hossain, A., Islam, A., Utsho, M. R., Kamal, M. B., Sharmin, S., &** Tasnim, A. F. (2026). The role of artificial intelligence in predicting cardiovascular outcomes: a systematic review and meta-analysis. Biostatistics & Epidemiology, 10(1). https://doi.org/10.1080/24709360.2026.2670804
30. **Semi, M. M. A., Das, S., Utsho, M. R., Hossain,** A., Kamal, M. B., Sizan, A. A., Tasnim, A. F., Yeasmin, S., & Parvin, M. R. (2026). Artificial Intelligence in Public Health Education: A Scoping Review of workforce competency development. Health Science Reports, 9(3). https://doi.org/10.1002/hsr2.72066
31. **Ahmed, F., Hasan, S., Hossain, A., & Rahman, K. A. (2026). Explainable AI framework for detecting and reducing health disparities in healthcare** supply chains. Journal of AI, ML and DL, 2(1), 1–8. https://doi.org/10.25163/ai.2110685
32. **Godavari Modalavalasa," SCALABLE CLOUD SOLUTIONS THROUGH ARTIFICIAL INTELLIGENCE GOVERNANCE**: APPLICATIONS IN HEALTHCARE AND FINANCIAL SYSTEMS", Vol. 54 No. 1 (2026): January-March 2026, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/219, ,
DOI: https://doi.org/10.46121/pspc.54.1.25
33. **Godavari Modalavala**sa, AUTONOMOUS DATA ENGINEERING PIPELINES: A POLICY-DRIVEN ARCHITECTURE FOR SECURE AND SCALABLE CLOUD-NATIVE ANALYTICS, Vol. 53 No. 4 (2025): October-December 2025, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/181,
DOI: https://doi.org/10.46121/pspc.53.4.25
34. **Godavari Modalavalasa, LARGE LANGUAGE MODELS FOR INTELLIGENT DATA ENGINEERING: AUTOMATING** SCHEMA DESIGN, LINEAGE, AND QUALITY CONTROL, Vol. 50 No. 2 (2022): April-June 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/183 ,
DOI: https://doi.org/10.46121/pspc.50.2.4

35. **Godavari Modalavalasa**, FEDERATED LEARNING FOR ENTERPRISE CLOUD DATA ENGINEERING: ARCHITECTURE, SECURITY, AND GOVERNANCE CHALLENGES, Vol. 51 No. 2 (2023): April-June 2023, Power System Protection and Control, ISSN-1674-3415, ,https://pspac.info/index.php/dlbh/article/view/184,
DOI: https://doi.org/10.46121/pspc.51.2.5
36. **Godavari Modalavalasa, AI**-DRIVEN DATA GOVERNANCE: INTELLIGENT METADATA, LINEAGE, AND COMPLIANCE AUTOMATION IN CLOUD DATA PLATFORMS, Archives / Vol. 52 No. 1 (2024): January-March 2024 /, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/182,
DOI: https://doi.org/10.46121/pspc.52.1.3
37. **Hima Bindu Lekkala, VishnuVardhan Bandar**i , AUTONOMOUS WORKFLOW OPTIMIZATION USING MULTI AGENT AI SYSTEMS AI AGENTS MANAGE STATIONS, WIP, AND TASK HANDOFFS, Vol. 54 No. 2 (202): April-June 2026, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/306 ,
DOI: https://doi.org/10.46121/pspc.54.2.08
38. **Lekkala, H. B., & Bandari, V. V. (2026). Autonomous workflow optimization using multi-agent AI systems:** AI agents manage stations, WIP, and task handoffs. *Power System Protection and Control*, 54(2), 95–101. http://pspac.info/index.php/dlbh/article/view/306
39. **Bandari, V. V., & Lekkala, H. B. (2024). Physics-informed reinforcement learning for real-time control of complex manufacturing processes**. *Power System Protection and Control*, 52(2), 164–170. https://pspac.info/index.php/dlbh/article/view/343
40. **Lekkala, H. B., & Bandari, V. V. (2025). AI-based energy-aware scheduling and process optimization** in engineer-to-order smart manufacturing systems. *Power System Protection and Control*, 53(1), 30–36. http://pspac.info/index.php/dlbh/article/view/341
41. **Bandari, V. V., & Lekkala, H. B. (2026). AI-driven digital** thread framework for end-to-end lifecycle optimization in ETO manufacturing systems. *Power System Protection and Control*, 54(2), 318–325. http://pspac.info/index.php/dlbh/article/view/340
42. **Lekkala, H. B., & Bandari, V. V. (2026). Deep learning** for weld defect detection using CNN or vision transformers fusion detection. *Power System Protection and Control*, 54(2), 151–158. https://pspac.info/index.php/dlbh/article/view/314
43. **Bandari, V. V., & Lekkala, H. B. (2024). AI-based operator** behavior monitoring and cost optimization using digital traceability in manufacturing systems. *Power System Protection and Control*, 52(1), 38–45. https://pspac.info/index.php/dlbh/article/view/342
44. **Shreyas Jayaprakash,** MACHINE LEARNING-DRIVEN VERIFICATION: USING ML TO OPTIMIZE COVERAGE ANALYSIS, PREDICT SIMULATION RUNTIMES, AND AUTO-GENERATE TESTBENCHES, Vol. 54 No. 2 (2026): April-June 2026, Power System Protection and Control, ISSN-1674-3415,
https://pspac.info/index.php/dlbh/article/view/391
45. **Mohammed Shafi Kundiladi,** EVENT-DRIVEN IMAGE AND VEHICLE STATUS MANAGEMENT FOR LOW-POWER IOT DIGITAL LICENSE PLATES, Vol. 53 No. 3 (2025): July-September 2025, Power System Protection and Control, ISSN-1674-3415,
https://pspac.info/index.php/dlbh/article/view/175
DOI: https://doi.org/10.46121/pspc.53.3.17
46. Controlled Synthesis and Characterization of Lanthanum Nanorods, **Author(s), Jayadeep Tejani, Rahul Shah, Hiral Vaghela, Shailesh Vajapara, Amanullakhan Pathan**, doi:10.18576/ijtfst/090205,
URL: https://www.naturalspublishing.com/Article.asp?ArtcID=20705, May-2020

47. **Kunal Kapoor**, **Apr 2026,** Relation Extraction and NER through Fine Tuned BERT model with transfer Learning, https://ieeexplore.ieee.org/document/11472214/metrics#metrics