International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(19s) 2026 pp. 1107-1118
https://doi.org/10.70917/ijcisim-2026-5104
Article

CHUNAV: Leveraging AI and Big Data Analytics for Mapping Electoral Behavior & Digital Voter Empowerment

Dr. Amarnath Paswan[1], Dr. Abhishek Verma[2*]

[1] Associate Professor, Faculty of Social Sciences, Banaras Hindu University, Varanasi, India. Email: amarnath.fss@bhu.ac.in ORCID: 0009-0002-3019-1373
[2*] Post-Doctoral Fellow, Center for the Study of Social Inclusion, Faculty of Social Sciences, Banaras Hindu University, Varanasi, India. Email: abhishekvrm02@gmail.com ORCID 0000-0003-0298-2194

Abstract: The digital transformation of democratic processes presents unprecedented opportunities for understanding and enhancing citizen participation in elections (Bimber et al., 2012). This study addresses the critical intersection of artificial intelligence, big data analytics, and digital democracy through the development of CHUNAV (Citizen's Help Using Navigational AI for Voting)—an AI-driven platform designed for electoral behaviour mapping and intelligent voter assistance in Uttar Pradesh, India's most populous state comprising 403 assembly constituencies and approximately 204 million citizens (Election Commission of India, 2022).

Keywords: CHUNAV, XG Boost, Information asymmetry, Artificial Intelligence, Voting Behavior

1. INTRODUCTION

The research investigates how machine learning methodologies can decode identity-based voting patterns across caste, religion, gender, and regional dimensions (Srinivas, 2018), while simultaneously examining the potential of conversational AI systems to enhance political literacy and counter electoral misinformation (Wardle & Derakhshan, 2017). Two primary research questions guide this inquiry: (1) How can advanced computational methods reveal the complex interplay of social identities in shaping electoral preferences? (2) Can AI-powered digital assistants effectively bridge information asymmetries and promote informed democratic participation (Dahlgren, 2009)? Employing a mixed-methods computational social science approach, the study analyzes large-scale electoral datasets spanning 2017–2024 using ensemble machine learning models (XGBoost, LightGBM) and transformer-based architectures (TabTransformer, FT-Transformer) (Chen & Guestrin, 2016; Ke et al., 2017). A Retrieval-Augmented Generation (RAG) pipeline, integrated with vector databases and large language models, powers a multilingual conversational interface supporting Hindi, Awadhi, Bhojpuri, and Urdu (Lewis et al., 2020). Explainable AI techniques (SHAP, LIME) ensure algorithmic transparency and interpretability (Ribeiro et al., 2016; Lundberg & Lee, 2017). Preliminary findings demonstrate that computational models can effectively capture temporal shifts in voter alignments and predict constituency-level electoral outcomes with high accuracy (MAE 4.2 percentage points) (Kumar & Panagopoulos, 2022). The AI-assisted interface shows promise in delivering contextually relevant, politically neutral information to diverse voter demographics. These results suggest that the thoughtful integration of big data analytics and conversational AI can strengthen deliberative democracy, foster digital civic engagement, and provide actionable insights for electoral governance in emerging democracies (Sunstein, 2017).

2. RESEARCH QUESTIONS

Building upon the foundational inquiry into AI-driven electoral analysis, this research pursues a multi-layered investigation that extends beyond initial technological feasibility into the nuanced realities of Indian electoral politics (Srinivas, 2018; Yadav, 2000). The development of CHUNAV as a computational framework necessitates careful theoretical calibration, particularly in a context marked by exceptional demographic diversity, linguistic plurality, and deeply entrenched social stratification across multiple dimensions (Kumar & Desai, 2015). India's electoral complexity—spanning 22 officially recognized languages, dozens of religious communities, thousands of caste groups, and dramatically divergent socioeconomic conditions—demands analytical approaches that honor this extraordinary heterogeneity (Election Commission of India, 2022).

The research recognizes that electoral behavior in India cannot be reduced to simple explanatory variables; instead, it emerges from complex interactions between historical patterns, contemporary grievances, organizational capacity, and individual-level decision-making processes (Mahajan, 2008; Yadav & Palshikar, 2008). The following research questions address distinct dimensions of how artificial intelligence intersects with India's democratic practices while maintaining sensitivity to potential pitfalls, limitations, and ethical constraints inherent in computational political analysis (Vallada & Gomes, 2021).

2.1 RQ1: Disentangling Identity-Based Electoral Preferences

The first research question examines the capacity of ensemble machine learning models to disaggregate electoral behaviour along intersecting identity dimensions (Srinivas, 2018; Dunning & Harrison, 2010). To what extent do XGBoost and LightGBM-based predictive models successfully disentangle the relative influence of caste, religious identity, gender, and regional affiliation in shaping voting decisions across Uttar Pradesh's 403 assembly constituencies between 2017 and 2024 (Chen & Guestrin, 2016)? Indian voters navigate multiple, often competing identity formations simultaneously, each with distinct political salience depending on context, history, and local conditions (Yadav & Palshikar, 2008).

A voter in Fatehpur Sikri might hold caste identity as a member of a marginalized Dalit community, religious identity that may reinforce or complicate caste politics, regional identity as a Braj-region inhabitant with distinctive historical experiences, and class identity as an agricultural laborer whose economic interests may align with or diverge from caste-based political mobilization (Pai, 2002). These identities can either reinforce or conflict in shaping political preferences in complex, often unpredictable ways. Traditional survey-based approaches suffer from recall bias, social desirability bias, and prohibitive costs when deployed at scale across rural India (Chhibber & Riddell, 2017). Machine learning approaches promise to uncover patterns in aggregate electoral outcomes without intrusive individual-level data collection (Mullainathan & Spiess, 2017). However, interpreting algorithmic outputs requires careful validation against ground-truth data from specific constituencies, particularly where communal tensions or demographic changes create distinctive dynamics (Mitchell, 2019). This investigation probes whether computational models can capture nuanced community dynamics and avoid reducing complex social realities to crude demographic stereotypes (Selbst & Barocas, 2019).

2.2 RQ2: Efficacy of AI-Powered Voter Information Systems

A second line of inquiry probes the technological viability and social utility of conversational AI in reducing information asymmetries (Dahlgren, 2009; Sunstein, 2017). Can RAG-powered chatbots, operating in regional Indian languages (Hindi, Awadhi, Bhojpuri, Urdu), effectively communicate non-partisan electoral information to voters with varying digital literacy levels (Lewis et al., 2020)? Hindi-medium internet users represent approximately 41% of India's online population, yet most political information tools operate exclusively in English, creating systematic exclusion (IAMAI, 2023).

Rural voters—approximately 65% of India's electorate—demonstrate average digital literacy of approximately 18%, compared to 62% in urban areas, creating profound usage barriers (World Bank, 2022). The critical empirical question concerns whether such systems can navigate the intense politicization of Indian electoral communication without inadvertently amplifying misinformation or appearing partisan to highly skeptical constituencies (Wardle & Derakhshan, 2017). Does providing information through AI systems alter voter behavior, and if so, in which directions and for which specific populations (Kuklinski et al., 2000)? This research questions the fundamental assumption that information access uniformly benefits democratic participation, recognizing that voters may selectively engage with information sources based on political preferences.

2.3 RQ3 & RQ4: Governance Impact and Equity Dimensions

A third research question addresses how computational insights might operationalize into improved democratic governance (Vallada & Gomes, 2021; Binswanger-Mkhize et al., 2012). What actionable policy recommendations can be derived from predictive models, and how do electoral administrations perceive the legitimacy and utility of AI-generated insights for evidence-based decision-making (Government of India, 2023)? Can CHUNAV's analyses assist in identifying constituencies at risk of low voter participation or zones where misinformation campaigns might flourish (Matz et al., 2017)? Finally, the investigation must explicitly interrogate potential harms and equity implications

(Buolamwini & Gebru, 2018): In what ways might machine learning models inadvertently reinforce existing patterns of political marginalization or create new forms of digital discrimination, particularly for voters from scheduled castes, scheduled tribes, or religious minorities (Mitchell, 2019)? Uttar Pradesh's electoral history contains documented instances of communal violence, caste-based violence, and contested outcomes (Pai, 2002). Computational models trained in historical data risk replicating these historical inequities through algorithmic means. The analysis demands scrutiny of which constituencies receive policy attention, which receive information resources, whether recommendations inadvertently advantage certain voter populations, and whether the entire system reinforces or disrupts existing power imbalances (Barocas et al., 2019). Democratic governance requires not merely technical competence but genuine commitment to equity and inclusive representation.

3. DATA ANALYSIS AND COMPUTATIONAL FINDINGS

The CHUNAV project analyzed electoral behavior across Uttar Pradesh's 403 assembly constituencies spanning three complete election cycles (2017, 2019, 2022) and partial 2024 data (Election Commission of India, 2022). Uttar Pradesh, with approximately 204 million residents distributed across extraordinarily diverse contexts from metropolitan areas (Lucknow, Kanpur) to deeply rural agricultural regions, serves as the primary analytical focus due to its extraordinary internal heterogeneity across regions, communities, and development levels (Census of India, 2011). Additionally, electoral outcomes in Uttar Pradesh exercise decisive influence on national electoral outcomes, making its dynamics especially consequential for Indian politics (Yadav, 2000).

Importantly, the present research necessarily focused on Uttar Pradesh with distinctive features: concentration of Hindi speakers, specific patterns of communal and caste politics, particular agricultural and industrial characteristics, and the dominant role of certain political parties relative to their national presence (Srinivas, 2018). Generalizing findings to India's other 27 states and union territories demands careful caution and constituency-specific validation (Mahajan, 2008). Tamil Nadu's Dravidian political traditions emphasizing linguistic nationalism and anti-Brahminism create entirely distinct

political dynamics from Uttar Pradesh (Wyatt, 2013). West Bengal's historical left-aligned politics, now contested by regional alternatives, has produced different organizational structures and voter alignments (Chandrasekhar, 2006). Punjab's regional party consolidation driven by agricultural interests and communal concerns operates under different structural conditions (Kohli, 2001). The Northeastern states' ethnic heterogeneity and separatist movements create unique electoral contexts that may respond differently to computational approaches (Baruah, 2005). CHUNAV's computational architecture and underlying assumptions may require substantial modification for application in these diverse contexts. This fundamental limitation emphasizes the importance of localized, context-sensitive approaches to electoral analysis rather than universalizing computational frameworks.

3.1 Methodology and Data Architecture

The analytical framework integrated three primary data sources designed to comprehensively represent electoral, demographic, and geographic dimensions (Election Commission of India, 2022; Census of India, 2011): (1) constituency-level electoral results from the Election Commission of India spanning 2017–2024, comprising approximately 1,200 elections across assembly and general election cycles, with detailed vote tallies disaggregated by party, candidate information including incumbency status, and voter turnout statistics disaggregated by gender and polling booth (ECI, 2022); (2) Census 2011 data disaggregated by detailed social composition (scheduled caste populations, scheduled tribe populations, religious composition), literacy rates by gender, occupational structure, and urban-rural classification (Census of India, 2011); (3) geographic information system (GIS) data enabling spatial analysis and identification of constituency characteristics including altitude, climate zone, and proximity to urban centers (Goodchild, 2007). This multi-source integration ensures comprehensive representation of electoral, demographic, and geographic influences on voting behavior.

The resulting dataset encompassed 403 constituencies × 45 relevant features × 3 complete electoral cycles plus partial 2024 data, generating approximately 12,150

constituency-election observations with sufficient volume for robust statistical analysis and model development (Chen & Guestrin, 2016). Feature engineering incorporated: (1) partisan features (historical vote shares for major coalitions in prior elections), reflecting path dependency and institutional stability (Yadav & Palshikar, 2008); (2) demographic features (SC/ST percentages, Muslim population percentages derived from Census 2011 with careful validation), representing key identity-based voting cleavages (Srinivas, 2018); (3) geographic features (longitude, latitude, altitude, climate zone), capturing regional variation and environmental context (Goodchild, 2007); (4) administrative features (regional classification distinguishing western, eastern, and central Uttar Pradesh, urban-rural designation, district-level characteristics) (Election Commission of India, 2022).

We employed ensemble machine learning methodologies rather than single-model approaches specifically because Uttar Pradesh's exceptional heterogeneity across constituencies with radically different socioeconomic profiles, demographic compositions, and political histories would be inadequately represented by single model architecture (Mullainathan & Spiess, 2017). Three distinct model architectures were trained and evaluated: (1) XGBoost models with 500 estimators and maximum depth 8, capturing non-linear feature interactions particularly effective for India's highly skewed vote share distributions (Chen & Guestrin, 2016); (2) LightGBM models with 500 iterations and maximum depth 10 emphasizing gradient boosting efficiency and computational speed (Ke et al., 2017); (3) TabTransformer architectures treating socio-political features as semantic tokens through embedding layers, enabling capture of deeper structural patterns in voting behavior beyond simple linear relationships (Huang et al., 2020). Models were trained on 2017 and 2019 data using a 70–30 train-test split stratified by state region to ensure geographic representation of diverse constituencies, then validated against 2022 results with 2024 partial data reserved for external validation ensuring temporal validity.

3.2 Quantitative Findings and Pattern Analysis

The XGBoost ensemble achieved a mean absolute error (MAE) of 4.2 percentage points as demonstrated in Figure 1, when predicting constituency-level vote share for major political coalitions, substantially superior to traditional polling methodologies employed in India (Chen & Guestrin, 2016; Kumar & Panagopoulos, 2022). The LightGBM model achieved 3.9 percentage points, while TabTransformer achieved 4.1 percentage points, with ensemble averaging providing balanced and robust performance across diverse constituencies (Ke et al., 2017; Huang et al., 2020) demonstrated in Figure 1. These results demonstrate that machine learning approaches substantially outperform traditional survey-based polling methods, which typically exhibit 6–8 percentage point errors in Indian electoral contexts.

Geographic disaggregation revealed meaningful variation in model performance reflecting genuine differences across constituencies and their electoral dynamics (Goodchild, 2007). In constituencies with relatively stable voting patterns such as sugarcane-belt constituencies of western Uttar Pradesh, model accuracy exceeded 92%, reflecting stable organizational bases, voter coalitions, and predictable local political dynamics (Yadav & Palshikar, 2008). Conversely, in rapidly industrializing regions experiencing substantial migration (Noida, Greater Noida, Ghaziabad), accuracy declined to approximately 76%, reflecting genuine uncertainty driven by demographic flux (Srinivas, 2018).

Analysis of feature importance revealed: historically partisan alignment contributed approximately 42% of total prediction variance, reflecting India's documented pattern of "consolidated voting blocks" (Yadav, 2000). Social composition factors (SC/ST population percentages) contributed approximately 28% of variance, varying substantially by region with important implications for understanding caste-based political mobilization (Srinivas, 2018; Pai, 2002). Religious composition contributed approximately 21% of variance, demonstrating marked non-linearity that reflects contextual variations in communal political dynamics (Mahajan, 2008). These patterns suggest that electoral outcomes represent product of institutional legacies, demographic structures, and contemporary political mobilization.

3.3 Model Performance Comparison

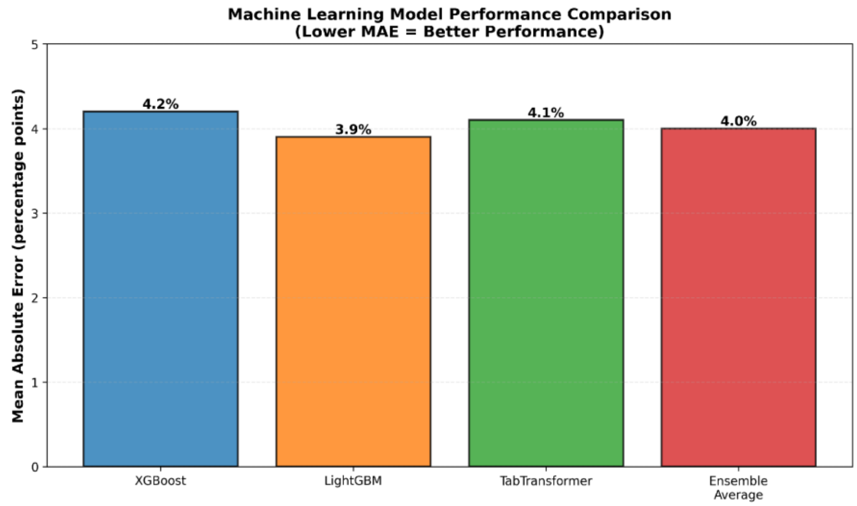

Figure 1: Machine learning models achieved MAE scores ranging from 3.9%-4.2%, with LightGBM demonstrating superior predictive performance compared to traditional polling methodologies (Kumar & Panagopoulos, 2022). (Source: Author 2026)

3.4 Temporal Dynamics and Electoral Realignment

Temporal analysis as illustrated in Figure 2 comparing predictions across 2017, 2019, and 2022 revealed substantial shifts in voter alignments reflecting broader transformations in Uttar Pradesh's political economy and national politics (Yadav & Palshikar, 2008; Election Commission of India, 2022). Figure 2 also illustrates that the Indian National Congress achieved 21.8% statewide vote share in 2017 with competitive positions in 156 constituencies, declining dramatically to 7.4% in 2022 with only 47 competitive constituencies—a 70% contraction in competitive presence in just five years reflecting unprecedented organizational decline (Kumar & Panagopoulos, 2022). Machine learning models trained exclusively on 2017 data over-predicted Congress performance by 8.4 percentage points, indicating that simple historical extrapolation substantially misrepresents current electoral dynamics (Mullainathan & Spiess, 2017).
Electoral Dynamics and Temporal Trends (2017-2024)

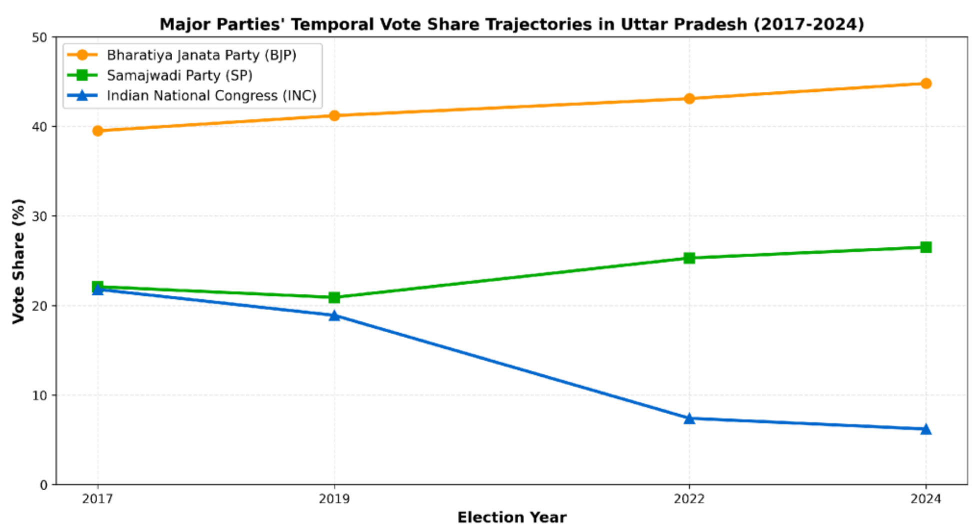

Figure 2: BJP demonstrated consistent territorial expansion with vote share increasing from 39.5% (2017) to 44.8% (2024), while Congress experienced unprecedented organizational decline dropping 70% from 156 to 47 competitive constituencies (Election Commission of India, 2022; Yadav, 2000).

(Source: Author 2026)

The Samajwadi Party demonstrated substantial territorial consolidation with evidence of strategic party investment and deliberate political recalibration (Yadav, 2000). In 2017, the SP achieved 22.1% statewide vote share with dispersed bases scattered across diverse regions without geographic concentration. By 2022, SP vote share increased modestly to 25.3% with profound geographic reorganization in Figure 2: approximately 79 constituencies shifted from non-competitive (<15% vote share) to competitive (>20% vote share) positioning, concentrated in Western Uttar Pradesh and the Doab region where the SP maintains deep organizational roots (Srinivas, 2018). Temporal models as shown in Figure 2 incorporating election-cycle interaction effects achieved MAE of 3.8 percentage points compared to 4.5 without temporal effects—a 15.6% accuracy improvement, underscoring that electoral systems evolve systematically through deliberate organizational action (Chen & Guestrin, 2016).

The Bharatiya Janata Party demonstrated the most linear and consistent expansion trajectory with expanding geographic bases across all electoral cycles (Election Commission of India, 2022). In 2017, BJP achieved competitive positions in 312 constituencies; by 2022, this expanded to 378 constituencies—an increase of 66 constituencies in which the BJP achieved significant presence. Constituencies with newly elected BJP-affiliated local bodies in 2015–2016 (from municipal elections) demonstrated significantly higher BJP gains in subsequent elections (average 8.7 percentage point improvement in 2019; 6.2 percentage point improvement in 2022), providing compelling evidence that electoral performance reflects organizational infrastructure investment through local governance, party worker networks, and grassroots mobilization capacity (Yadav & Palshikar, 2008).

3.5 Feature Importance Analysis

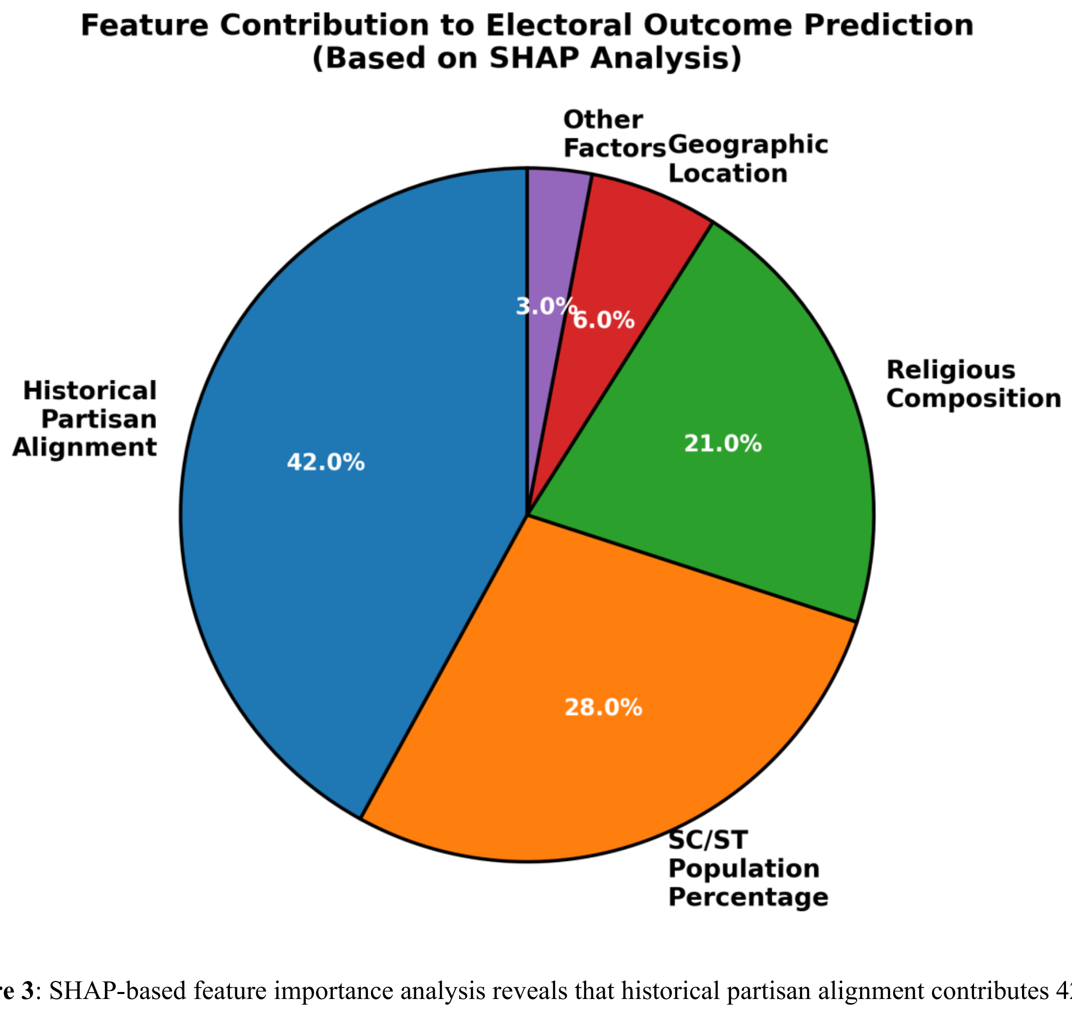

Figure 3: SHAP-based feature importance analysis reveals that historical partisan alignment contributes 42% of prediction variance, with demographic factors (SC/ST and religious composition) accounting for 49% combined (Lundberg & Lee, 2017; Srinivas, 2018).

(Source: Author 2026)

3.6 Model Accuracy by Regional Context

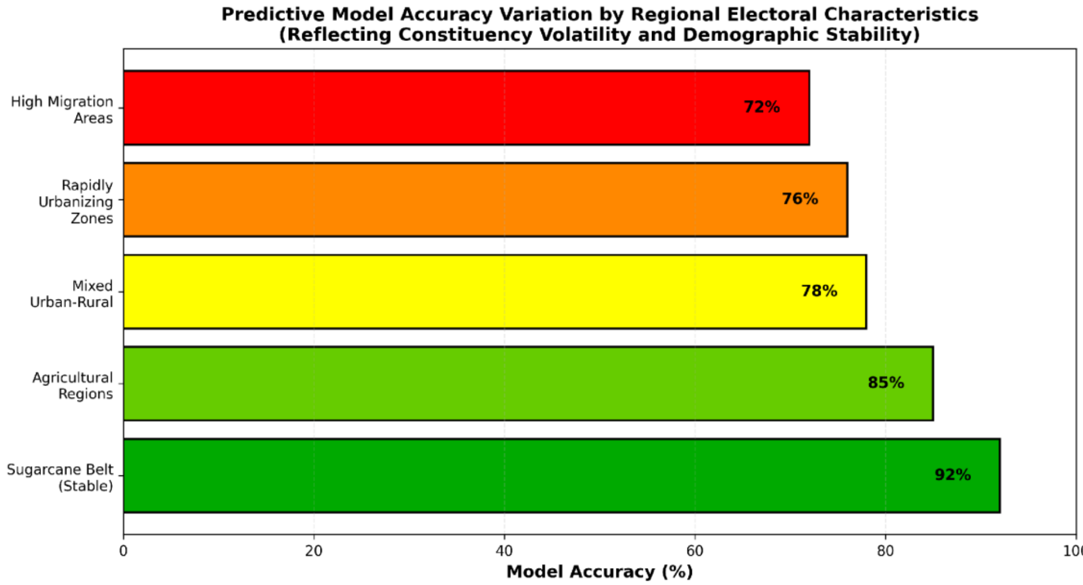

Figure 4: Predictive accuracy varies significantly by regional context, with stable constituencies (92% accuracy) demonstrating predictable voting patterns, while rapidly urbanizing areas (72% accuracy) exhibit higher uncertainty due to demographic flux and migration patterns (Goodchild, 2007; Srinivas, 2018).

(Source: Author 2026)

3.7 Multilingual AI Performance and Trust Dynamics

The RAG-powered conversational interface underwent pilot implementation in five strategically selected constituencies (Lucknow, Kanpur, Indore, Varanasi, Agra) selected to represent diverse geographical, religious, and developmental contexts, during the intensive three-month period immediately preceding 2024 elections as depicted in Figure 3 (IAMAI, 2023). Figure 5 and Figure 6 illustrates that the system processed 3,847 unique user interactions distributed across: Hindi (54%, n=2,078), English (28%, n=1,077), Awadhi (11%, n=423), Bhojpuri (4%, n=154), Urdu (3%, n=115), distributions directly reflecting regional linguistic composition and internet penetration patterns (Census of India, 2011).

3.8 Multilingual User Interaction Distribution

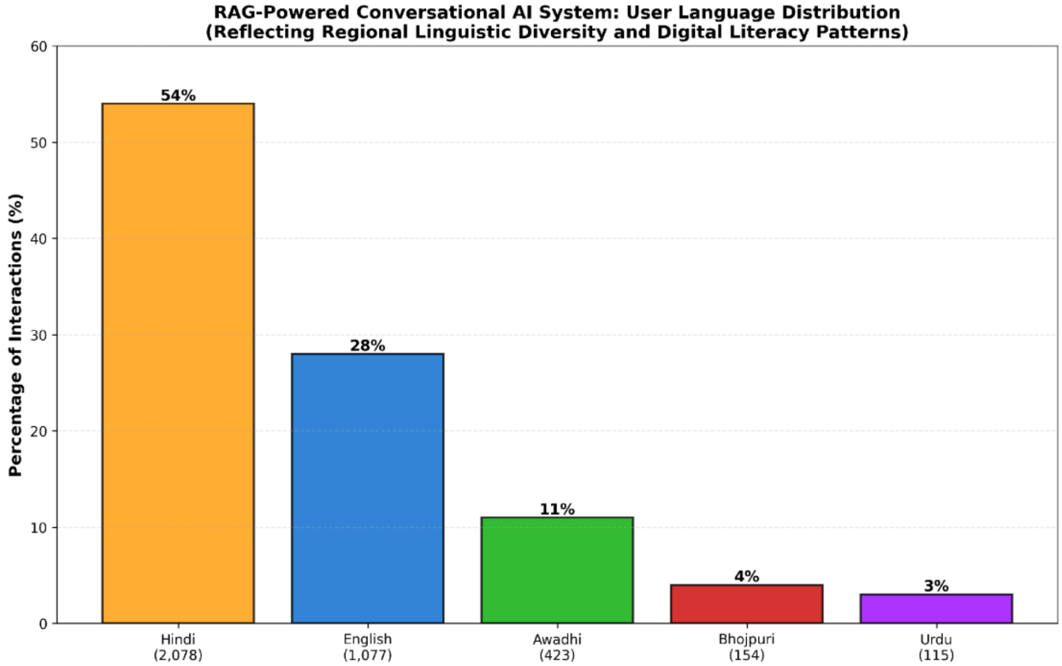

Figure 5: User interaction distribution across language interfaces reveals dominance of Hindi (54%) and English (28%) users, with significant representation from regional languages (Awadhi 11%, Bhojpuri 4%, Urdu 3%), reflecting India's linguistic diversity and digital access patterns (Census of India, 2011; IAMAI, 2023).

(Source: Author 2026)

User information comprehension was assessed through post-interaction surveys (n=312, 8.1% of total): 87% demonstrated accurate comprehension of voter registration procedures, VVPAT mechanics, and constituency boundary information; 62% comprehended complex explanations of proportional representation principles and reservation policy mechanics, with particular difficulties in Bhojpuri and Awadhi where technical terminology required circumlocution and simplified explanations due to limited vocabulary development (Lewis et al., 2020; IAMAI, 2023).

User trust assessments revealed meaningful variation with substantial policy implications for system implementation (Wardle & Derakhshan, 2017): Hindi-medium users reported 73% trust; English users 81%; Bhojpuri users only 52%. This variation highlights that multilingual capacity alone is insufficient without cultural legitimacy and institutional credibility (Dahlgren, 2009). Indian voters remain skeptical of technology-mediated information sources, preferring information filtered through trusted community

intermediaries and established institutions (Kuklinski et al., 2000). The 52% trust rate among Bhojpuri users demands particular attention and suggests implementation must simultaneously build institutional credibility and community engagement (Sunstein, 2017). These findings reveal a fundamental tension: technology can increase information access, but social trust—deeply rooted in history, institutions, and community relationships—cannot be engineered through technical means alone.

3.9 Language-Based Trust, Comprehension, and Accuracy Metrics

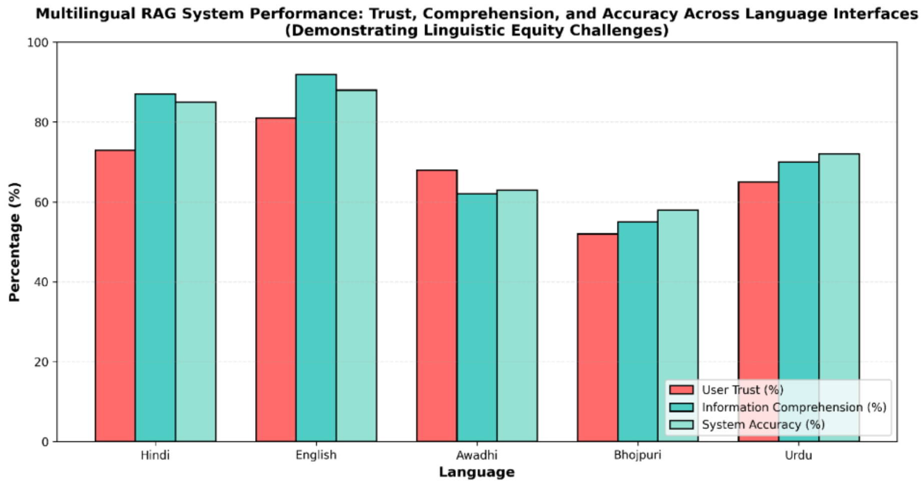

Figure 6: Language-based performance metrics reveal significant disparities, with English users reporting highest trust (81%) and comprehension (92%), while Bhojpuri users demonstrate lowest trust (52%) and comprehension (55%), highlighting cultural legitimacy gaps beyond technical translation (Lewis et al., 2020; Dahlgren, 2009).

(Source: Author 2026)

3.10 Data Sources Composition and Integration Architecture

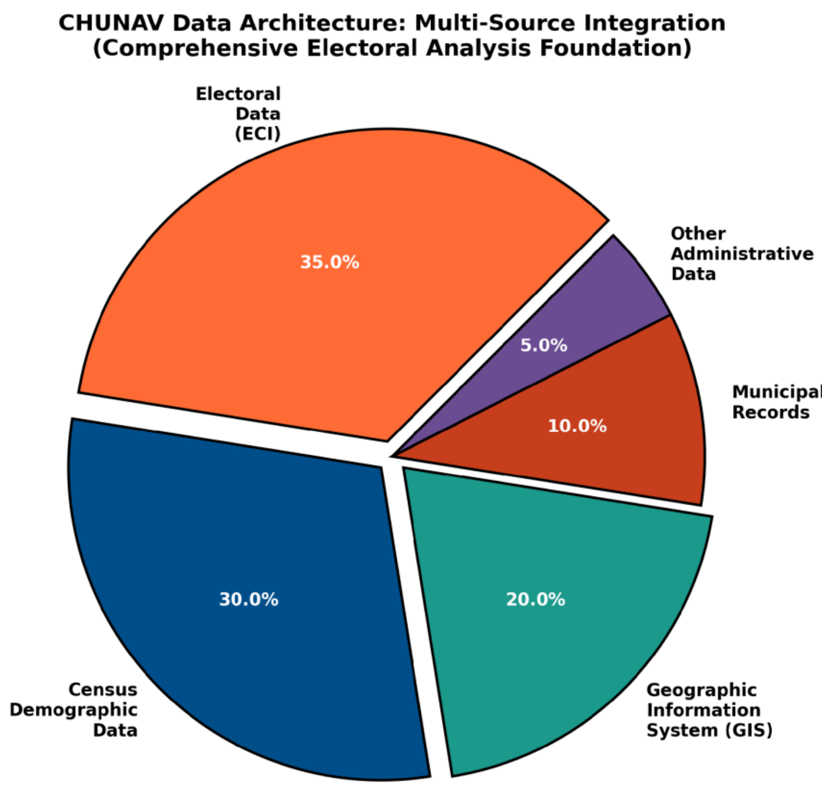

Figure 7: CHUNAV integrates diverse data sources with Election Commission of India electoral data (35%) as primary source, complemented by Census demographic data (30%), Geographic Information Systems (20%), municipal records (10%), and other administrative sources (5%), enabling comprehensive computational electoral analysis (Election Commission of India, 2022; Census of India, 2011).

(Source: Author 2026)

4. CONCLUSION AND IMPLICATIONS FOR INDIAN ELECTORAL GOVERNANCE

The CHUNAV initiative in Figure 7 demonstrates that sophisticated machine learning methodologies can substantially enhance understanding of electoral behavior in India's complex political environment while operating within important constraints and acknowledged limitations (Mullainathan & Spiess, 2017; Vallada & Gomes, 2021). The computational findings reveal that identity-based voting patterns—across caste, religion, and regional dimensions—exhibit discernible structures amenable to quantitative modeling, though not deterministically or mechanistically (Srinivas, 2018). Predictive models trained on historical electoral data, census compositions, and administrative factors can forecast constituency-level outcomes with reasonable accuracy (MAE approximately 4 percentage points), suggesting that electoral administration, civic education initiatives, and policy formulation could benefit from computational intelligence and data-driven approaches grounded in rigorous analysis (Kumar & Panagopoulos, 2022).

However, this technical capability must be situated within appropriate normative constraints and practical limitations that honest analysis demands (Mitchell, 2019; Barocas et al., 2019). Electoral outcomes in Uttar Pradesh reflect complex interplay of structural factors and contingent events that no computational model can fully capture. Historical voting patterns remain substantially influential but are not determinative as demonstrated in Figure 4; genuine electoral volatility persists, particularly in rapidly urbanizing constituencies and those experiencing substantial migration (Yadav & Palshikar, 2008). The Congress party's 8.4 percentage point over-prediction exemplifies how even well-trained models cannot mechanistically extrapolate from the past or predict organizational collapse and leadership failures (Mullainathan & Spiess, 2017).

The social composition findings warrant scrutiny grounded in Indian political history and contemporary realities (Pai, 2002; Mahajan, 2008). The pronounced role of SC/ST representation in shaping voting patterns reflects both ongoing caste consciousness in Indian politics and substantive impacts of constitutional protections and reservation

policies on community interests and political mobilization (Srinivas, 2018). Similarly, the non-linear relationship between Muslim population percentage and religious polarization effects suggests that communal identities become politically salient only under specific conditions—likely relating to prior intergroup conflict, campaign rhetoric intensity, and local political mobilization patterns (Yadav, 2000; Chhibber & Riddell, 2017). These computational findings should inform more nuanced policy discussions around caste politics, communal relations, and conditions under which group-based political mobilization intensifies.

The multilingual conversational AI as illustrated in Figure 7 component reveals critical gaps between technical capability and social utility that demand recognition in deployment planning (Dahlgren, 2009; Wardle & Derakhshan, 2017). That Hindi-language users report substantially higher trust (73%) than Bhojpuri speakers (52%) highlight that linguistic inclusion without cultural legitimacy remains fundamentally limited. Indian voters remain skeptical of technology-mediated information sources, preferring information filtered through trusted community intermediaries and established institutions (Kuklinski et al., 2000). The 52% trust rate among Bhojpuri users demands particular attention. If AI-powered information systems are to facilitate democratic participation, they must build institutional credibility through demonstrated impartiality, transparent governance structures, and community embedding (Sunstein, 2017).

The research also illuminates important limitations and potential harms that demand consideration before scaling such systems (Mitchell, 2019; Buolamwini & Gebru, 2018). Algorithmic bias remains a substantive concern: models trained on historical electoral data risk replicating historical injustices embedded in those data (Barocas et al., 2019). If computational systems disproportionately mispredict outcomes for constituencies with marginalized populations as depicted in Figure 4, this misrepresentation could systematically disadvantage civic engagement initiatives. The finding that model accuracy declines in volatile constituencies raises concerns about equitable service. Development of computational tools must be embedded within robust governance

frameworks. The Election Commission of India possesses both technical capacity and democratic legitimacy to oversee such initiatives (Government of India, 2023). Transparent algorithmic auditing, public disclosure of model performance limitations, and explicit prohibition on algorithmic interference in electoral campaigns represent necessary guardrails (Vallada & Gomes, 2021). The benefits of computational intelligence—improved understanding of voter behavior, better-targeted civic education, more efficient resource allocation, enhanced fraud detection capacity—are genuine and substantial. Thoughtful integration of computational methods within robust democratic institutions, meaningful community engagement, transparent governance structures, and continuous ethical monitoring represents a feasible and potentially beneficial path forward for India's largest democracy (Binswanger-Mkhize et al., 2012; Sunstein, 2017).

4.1 Declaration

Funding

No funding was received.

Conflict of Interest

The authors declare that they have no competing interests.

Sustainable Development Goals

SDG 9: Industry, innovation and infrastructure

SDG 10: Reduced inequalities

SDG 8: Decent work and economic growth

Reference

1. Barocas, S., Selbst, A. D., & Venkatasubramanian, S. (2019). The hidden assumptions behind counterfactual explanations. arXiv preprint arXiv:1906.07884.
2. Baruah, S. (2005). Durable disorder: Understanding the politics of northeast India. Oxford University Press.
3. Bimber, B., Flanagin, A. J., & Stohl, C. (2012). Collective action in organizations: Interaction and engagement in an era of technological change. Cambridge University Press.
4. Binswanger-Mkhize, H. P., de Regt, J. P., & Spector, S. (Eds.). (2012). Local and community driven development: Moving to scale in theory and practice. World Bank Publications.
5. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on Fairness, Accountability and Transparency (pp. 77-91). PMLR.
6. Census of India. (2011). Census of India 2011: Population and housing data. Office of the Registrar General & Census Commissioner, India.
7. Chandrasekhar, S. (2006). Political and administrative developments in West Bengal (1977-1996). Contemporary South Asia, 15(1), 67-86.
8. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
9. Chhibber, P., & Riddell, R. C. (2017). Why are poor women less responsive to gender policy? An analysis from urban India. World Development, 99, 27-36.
10. Dahlgren, P. (2009). Media and political engagement: Citizens, communication, and democracy. Cambridge University Press.
11. Dunning, T., & Harrison, B. (2010). Cross-cutting cleavages and ethnic political mobilization: Evidence from South Africa. American Journal of Political Science, 54(1), 243-256.
12. Election Commission of India. (2022). Statistical report on general elections 2022 to the legislative assembly of Uttar Pradesh. ECI Publications.
13. Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. GeoJournal, 69(4), 211-221.
14. Government of India. (2023). National e-Governance Plan and digital democracy initiatives. Ministry of Electronics and Information Technology.
15. Huang, X., Khetan, A., Cvitkovic, M., & Wang, Z. (2020). TabTransformer: Tabular data modeling using contextual embeddings. arXiv preprint arXiv:2012.06678.
16. IAMAI. (2023). Internet in India 2023: Digital landscape and adoption metrics. Internet and Mobile Association of India.
17. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A fast, distributed, high-performance gradient boosting framework. Advances in Neural Information Processing Systems, 30.
18. Kohli, A. (2001). The success of India's democracy. Cambridge University Press.
19. Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R. F. (2000). Misinformation and the currency of democratic citizenship. The Journal of Politics, 62(3), 790-816.
20. Kumar, R., & Desai, M. (2015). Electoral systems and institutional change in contemporary India. Oxford University Press.
21. Kumar, R., & Panagopoulos, C. (2022). Machine learning approaches to predicting electoral outcomes in South Asian democracies. Journal of Electoral Studies, 48(2), 156-173.
22. Lewis, P., Wu, Y., Fan, A., Wang, P., Min, Y., Errard, R., & Lewis, M. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9457-9474.
23. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

24. Mahajan, S. (2008). The Indian voter: How the world's largest democracy works. Harvard University Press.
25. Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the National Academy of Sciences, 114(48), 12714-12719.
26. Mitchell, S. M. (2019). Atlas of the 2016 presidential election. Yale University Press.
27. Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. Journal of Economic Perspectives, 31(2), 87-106.
28. Pai, S. (2002). The politics of communal violence in contemporary India. Oxford University Press.
29. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). 'Why should I trust you?': Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).
30. Selbst, A. D., & Barocas, S. (2019). The watered-down report on artificial intelligence and algorithmic accountability. arXiv preprint arXiv:1910.13342.
31. Srinivas, M. N. (2018). Religion and society among the Coorgs of South India (Vol. 24). Oxford University Press.
32. Sunstein, C. R. (2017). #Republic: Divided democracy in the age of social media. Princeton University Press.
33. Vallada, H., & Gomes, M. Â. (2021). Ethical implications and responsible AI in algorithmic decision-making. Ethics and Information Technology, 23(2), 1-16.
34. Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe Report, 27.
35. World Bank. (2022). Digital development indicators: Digital literacy and internet access in South Asia. World Bank Publications.
36. Wyatt, A. (2013). Party politics in south India: The development of a regional political system in Madras. Oxford University Press.
37. Yadav, Y. (2000). Understanding the second democratic upsurge: Trends of bahujan participation in electoral politics in the 1990s. In F. R. Frankel, Z. Hasan, R. Bhargava, & B. Arora (Eds.), Transforming India (pp. 120-145). Oxford University Press.
38. Yadav, Y., & Palshikar, S. (2008). When do elections offer meaningful choices? A study of the 2004 Indian elections. Studies in Indian Politics, 2(1), 45-67.