Applications of Data Mining and Machine Learning in State Surveillance and the Oppression of Marginalized Populations: A Systematic Literature Review

Introduction

The rapid proliferation of data mining (DM) and machine learning (ML) technologies has fundamentally transformed the landscape of state surveillance. Once limited by the constraints of analog systems and human labor, contemporary surveillance regimes now leverage automated, algorithmic processes to monitor, predict, and influence the behaviors of entire populations. While these advances promise enhanced efficiency and security, they also raise profound ethical, legal, and social concerns—particularly regarding their deployment against marginalized populations. The intersection of DM/ML with state surveillance is not merely a technical phenomenon; it is a deeply political and social process that can entrench existing inequities, amplify discrimination, and undermine fundamental rights.

This systematic literature review critically examines the applications of data mining and machine learning in state surveillance, with a focus on their roles in the oppression of marginalized groups. Drawing on peer-reviewed academic sources, government reports, and analyses from reputable NGOs, the review synthesizes current knowledge, identifies key trends and gaps, and proposes a testable hypothesis for future research. The review is structured according to systematic review standards, encompassing a clear methodology, synthesis of key findings, analysis of trends and gaps, and a forward-looking hypothesis.

Methodology

Systematic Review Standards in Technology and Ethics

This review adheres to established systematic review protocols, notably the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, which emphasize transparency, replicability, and comprehensiveness in literature synthesis. The methodology is informed by best practices in technology and ethics reviews, including the need for multidisciplinary perspectives, explicit inclusion/exclusion criteria, and critical appraisal of study quality.

Search Strategy and Inclusion Criteria

A comprehensive search was conducted across major academic databases (Scopus, Web of Science, PubMed, IEEE Xplore), supplemented by targeted searches of government and NGO reports (e.g., EPIC, Amnesty International, Human Rights Watch). Search terms included combinations of "data mining," "machine learning," "state surveillance," "predictive policing," "facial recognition," "biometric identification," "social media monitoring," "marginalized populations," "algorithmic bias," and "digital authoritarianism." The search was limited to English-language publications from 2015 to early 2026 to capture recent developments.

Studies were included if they:

- Examined the application of DM/ML in state surveillance contexts.
- Addressed impacts on marginalized populations (e.g., racial, ethnic, religious, sexual, political minorities; migrants; refugees).
- Provided empirical data, systematic reviews, or robust theoretical analyses.
- Were published in peer-reviewed journals, reputable government reports, or by established NGOs.

Exclusion criteria comprised:

- Studies focused solely on private sector surveillance without state involvement.
- Technical papers lacking social, ethical, or legal analysis.
- Opinion pieces without substantive evidence.

Data Extraction and Synthesis

Data were extracted on study objectives, methodologies, surveillance technologies examined, populations affected, key findings, and identified limitations. The synthesis employed both narrative and tabular approaches to compare findings across studies, with particular attention to recurring themes, methodological rigor, and reported outcomes.

Key Findings

Scope and Definitions

Data Mining and Machine Learning in Surveillance:
Data mining refers to the automated extraction of patterns and knowledge from large datasets, often employing statistical, computational, and algorithmic techniques. Machine learning, a subset of artificial intelligence, enables systems to learn from data

and improve their performance over time without explicit programming. In surveillance, these technologies are used to process vast streams of data from sensors, cameras, social media, and other sources, enabling real-time monitoring, behavioral prediction, and automated decision-making.

State Surveillance:
State surveillance encompasses the collection, analysis, and use of information by government actors to monitor, control, or influence populations. Modern surveillance systems are increasingly automated, networked, and capable of integrating multiple data modalities (e.g., video, audio, text, biometrics).

Marginalized Populations:
Marginalized populations include groups systematically disadvantaged by social, economic, or political structures—such as racial and ethnic minorities, migrants, refugees, religious and sexual minorities, and political dissidents. These groups are often disproportionately targeted or adversely affected by surveillance practices.

Historical Overview and Evolution

Surveillance technologies have evolved from manual, analog systems to highly automated, AI-driven architectures. Early surveillance focused on physical observation and record-keeping, expanding in the 20th century to include electronic eavesdropping and database management. The post-9/11 era marked a significant acceleration, with the USA PATRIOT Act and similar legislation worldwide enabling mass data collection and analysis for national security purposes. The integration of DM/ML has further transformed surveillance, enabling predictive analytics, real-time monitoring, and the fusion of disparate data sources.

Technical Applications

Facial Recognition, Biometric Identification, and Computer Vision

Facial recognition and biometric identification systems are now central to state surveillance infrastructures. These systems use ML algorithms to match faces, fingerprints, irises, and other physiological features against large databases, enabling rapid identification and tracking of individuals in public and private spaces. Computer vision techniques extend these capabilities to object detection, crowd analysis, and behavioral inference.

Bias and Discrimination:
Numerous studies have documented significant accuracy disparities in facial recognition systems, with error rates up to 40 times higher for darker-skinned women compared to lighter-skinned men. These biases stem from unrepresentative training datasets and can lead to wrongful arrests, exclusion from services, and heightened surveillance of minority communities.

Deepfakes and Generative AI:
The rise of generative AI, particularly deepfakes, poses new challenges for biometric security. Synthetic identities can be used to evade detection or to fabricate evidence, complicating both security and accountability efforts.

Predictive Policing and Crime Forecasting

Predictive policing algorithms analyze historical crime data to forecast where and when crimes are likely to occur, and in some cases, who is likely to commit them. Systems such as PredPol and HunchLab have been widely adopted in the US and elsewhere.

Feedback Loops and Racial Bias:

Predictive policing systems often rely on arrest and incident data that reflect historical patterns of over-policing in minority neighborhoods. This creates self-reinforcing feedback loops: increased police presence leads to more arrests, which in turn justifies further surveillance and intervention in the same communities. Empirical studies have shown that algorithms like PredPol would send police to Black neighborhoods at twice the rate of White neighborhoods, even when controlling for crime rates.

Effectiveness and Accountability:
Despite claims of objectivity and efficiency, there is limited evidence that predictive policing reduces crime. Accuracy rates vary widely, and the systems are often opaque, lacking transparency and avenues for contestation by those affected.

Social Media Monitoring, NLP, and Sentiment Analysis

States increasingly deploy DM/ML tools to monitor social media platforms, employing natural language processing (NLP) and sentiment analysis to detect dissent, extremism, or other behaviors deemed threatening. These systems can process vast quantities of text, images, and video, identifying networks of activists, protestors, or minority groups.

Chilling Effects and Misinterpretation:
Automated social media surveillance can have chilling effects on free expression, particularly among marginalized groups who are aware of being monitored. NLP tools often struggle with context, slang, and multilingual content, leading to misclassification and over-flagging of minority speech.

Targeting of Activists and Minorities:
Reports document the use of social media monitoring to target Black Lives Matter activists, Muslim communities, and immigration advocates in the US and elsewhere. In some cases, visa revocations and deportations have been based on social media posts or associations, raising due process and discrimination concerns.

Network Analysis, Link Prediction, and Data Fusion

Social network analysis (SNA) and link prediction algorithms are used to map relationships among individuals, organizations, and events, often to identify "key nodes" or leaders within activist or criminal networks. Data fusion techniques integrate information from multiple sources (e.g., surveillance cameras, social media, financial records) to create comprehensive behavioral profiles.

Potential for Abuse:
While SNA can aid in targeting interventions, it also risks unjustly labeling individuals as threats based on their associations, particularly in communities already subject to heightened surveillance.

Public Health Surveillance vs. Repressive Surveillance

The COVID-19 pandemic accelerated the adoption of digital surveillance for public health purposes, including contact tracing, outbreak prediction, and quarantine enforcement. While these applications can benefit population health, they also raise concerns about mission creep, data repurposing, and the normalization of surveillance infrastructures.

Equity and Access:
Digital health surveillance systems often fail to account for the needs and barriers faced by marginalized populations, potentially exacerbating health disparities.

Case Studies

Digital Authoritarianism: China, Russia, Iran

China:
China's surveillance apparatus is among the most sophisticated globally, integrating facial recognition, biometric databases, social credit scoring, and real-time analytics across public and private sectors. The system is used not only for crime prevention but also for social control, targeting ethnic minorities (notably Uyghurs), political dissidents, and religious groups. Technologies developed in China are exported to dozens of countries, spreading the model of digital authoritarianism.

Russia and Iran:
Both countries have implemented extensive digital surveillance programs, often justified by national security concerns but used to monitor and suppress political opposition, journalists, and minority groups.

Democratic States: US, UK, Australia

The US has deployed DM/ML in surveillance across law enforcement, immigration, and public health. Predictive policing, facial recognition, and social media monitoring are widespread, with documented impacts on Black, Latino, Muslim, and immigrant

communities. Oversight and transparency remain limited, and federal privacy protections are fragmented.

United Kingdom and Australia:
Both countries have adopted predictive policing and digital health surveillance, with similar concerns about bias, transparency, and the disproportionate impact on marginalized groups.

Impacts on Marginalized Populations

Racial and Ethnic Minorities

Surveillance systems disproportionately target and harm racial and ethnic minorities. Predictive policing, facial recognition, and social media monitoring have been shown to amplify existing patterns of discrimination, leading to higher rates of arrest, misidentification, and exclusion from services.

Migrants, Refugees, and Asylum Seekers

Automated border control, biometric identification, and social media screening are increasingly used in migration management. These systems often exhibit racial and linguistic biases, leading to higher rejection rates for Black, Brown, and Indigenous applicants, and raising concerns about due process and non-refoulement.

Religious, Sexual, and Political Minorities

Surveillance is frequently deployed against religious minorities (e.g., Muslims in China and Western democracies), LGBTQ+ individuals, and political dissidents. Algorithmic systems may misclassify religious expression as extremism, leading to censorship, exclusion, or worse.

Psychological and Social Effects

The pervasive nature of surveillance induces chilling effects, reducing willingness to engage in political activism, express dissent, or participate in public life—especially among those already marginalized. Studies show that algorithmic surveillance, in particular, is associated with lower perceived autonomy, increased resistance, and diminished trust.

Legal and Policy Frameworks

Domestic Laws

Legal protections vary widely. The EU's General Data Protection Regulation (GDPR) and AI Act provide robust frameworks for data protection and algorithmic accountability, though enforcement and scope remain challenges. The US lacks comprehensive federal

privacy legislation, relying on a patchwork of state laws and sector-specific regulations. China and other authoritarian states have enacted data protection laws that primarily serve state interests.

International Human Rights

International human rights instruments (e.g., Universal Declaration of Human Rights, International Covenant on Civil and Political Rights) establish rights to privacy, non-discrimination, and due process. However, enforcement is limited, and many surveillance practices operate in legal gray zones or under national security exemptions.

Oversight and Accountability

Oversight mechanisms are often inadequate. Many surveillance systems are proprietary "black boxes," shielded from public scrutiny by trade secrets or national security justifications. Algorithmic impact assessments, audits, and transparency requirements are emerging but remain inconsistently applied.

Ethical Analyses and Normative Critiques

Privacy and Autonomy:
AI-driven surveillance challenges traditional notions of privacy and informed consent, as data collection is often covert, involuntary, and irreversible. The loss of autonomy is particularly acute under algorithmic surveillance, which is perceived as more dehumanizing and less accountable than human monitoring.

Chilling Effects:
Surveillance induces chilling effects, discouraging dissent, experimentation, and the development of non-mainstream identities—especially among marginalized groups.

Fairness, Bias, and Discrimination:
Algorithmic systems often reproduce and amplify existing social biases, leading to disparate impacts on marginalized populations. Technical fixes (e.g., fairness metrics, bias audits) are necessary but insufficient without broader structural and democratic reforms.

Transparency and Explainability:
The opacity of many DM/ML systems undermines accountability and due process. Calls for explainable AI, open algorithms, and participatory governance are growing but face resistance from both state and corporate actors.

Data Quality, Label Bias, and Feedback Loops

Data quality issues—such as incomplete, inaccurate, or biased data—undermine the reliability and fairness of surveillance systems. Label bias and feedback loops are

particularly pernicious in predictive policing and risk assessment, where historical inequities are encoded and perpetuated by algorithms.

Mitigation Strategies

Technical Safeguards:
Fairness-aware algorithms, bias audits, and representative datasets are essential for mitigating harm, but must be complemented by robust oversight and participatory design.

Legal and Policy Reforms:
Comprehensive privacy laws, algorithmic impact assessments, and independent oversight bodies are critical for ensuring accountability and protecting rights.

Civil Society and Advocacy:
NGOs and advocacy groups play a vital role in exposing abuses, litigating for rights, and mobilizing public awareness.

Public Health vs. Repressive Surveillance:
Clear boundaries and safeguards are needed to prevent the repurposing of public health surveillance for repressive ends.

Table: Summary of Key Studies and Findings

Study/Source	Technology/ Context	Marginalized Group(s)	Key Findings/Impacts
PredPol (US, UK)	Predictive policing	Black, Latino communities	Amplifies racial bias, feedback loops, limited efficacy stateofsurveillance.org + 1
China Social Credit System	Biometric, social scoring	Uyghurs, dissidents	Comprehensive profiling, repression, export model society.sciencearray.com + 1
US Immigration Screening	Social media, biometrics	Migrants, refugees	Bias in facial recognition, visa denials, due process concerns academic.oup.com + 1
France Safe City Project	CCTV, analytics	Ethnic minorities	Misidentification, privacy concerns sdgs.un.org
Australia Suspect Target Plan	Predictive policing	Aboriginal youth	Disproportionate targeting, systemic bias sdgs.un.org
LAPD, NYPD	Predictive policing, SNA	Black, Latino communities	Over-policing, lack of transparency, legal challenges stpp.fordschool.umich.edu + 1
EU iBorderCtrl	Automated border control	Migrants, minorities	Racial bias in lie detection, exclusion academic.oup.com + 1
US Social Media Monitoring	NLP, sentiment analysis	Activists, minorities	Chilling effects, misclassification, visa revocations www.brennancenter.org + 1
Denmark, Serbia Welfare	Algorithmic decision-making	Roma, women, disabled	Exclusion from services, lack of recourse www.amnesty.org

Trends and Gaps

Emerging Trends

- **Expansion of Generative AI and Multimodal Surveillance:**
  The integration of generative AI (e.g., deepfakes) and multimodal analytics

(combining video, audio, text) is expanding the scope and sophistication of surveillance, raising new challenges for detection, accountability, and harm mitigation.

- **Real-Time and Ubiquitous Surveillance:**
  Advances in edge computing, 5G, and IoT enable real-time monitoring of individuals across physical and digital spaces, blurring boundaries between public and private life.
- **Global Diffusion of Surveillance Technologies:**
  Surveillance models developed in China, the US, and Europe are being exported worldwide, often without adequate safeguards or adaptation to local contexts.
- **Algorithmic Governance and Automated Decision-Making:**
  States increasingly rely on automated systems for decisions affecting rights and access to services, with limited transparency or avenues for contestation.

Research Gaps

- **Empirical Evidence and Impact Evaluation:**
  There is a paucity of longitudinal, empirical studies evaluating the real-world impacts of surveillance technologies on marginalized populations. Most evidence is cross-sectional or anecdotal, limiting causal inference.
- **Intersectionality and Contextual Nuance:**
  Existing research often treats marginalized groups as monolithic, neglecting intersectional dynamics (e.g., race, gender, class, disability) and local contexts.
- **Transparency and Access to Data:**
  Proprietary algorithms and restricted access to data impede independent audits and accountability efforts.
- **Mitigation and Redress Mechanisms:**
  There is limited research on effective mitigation strategies, redress mechanisms, and participatory governance models that center the voices of affected communities.
- **Global South Perspectives:**
  Most research is concentrated in North America, Europe, and China, with insufficient attention to the experiences and needs of marginalized populations in the Global South.

Proposed Hypothesis

Hypothesis:

Under-representation of marginalized populations in training datasets used for machine

learning-based surveillance systems leads to systematic disparities in algorithmic performance, resulting in higher rates of false positives and false negatives for these groups, which in turn amplify their exposure to state surveillance and its associated harms.

This hypothesis is testable through empirical studies that compare algorithmic performance metrics (e.g., accuracy, false positive/negative rates) across demographic groups, and that assess downstream impacts on surveillance outcomes (e.g., arrests, service denials, visa rejections).

Conclusion

The integration of data mining and machine learning into state surveillance systems marks a profound shift in the modalities and scale of social control. While these technologies offer potential benefits for public safety and efficiency, they also pose significant risks—particularly for marginalized populations who are disproportionately targeted, misclassified, and harmed by algorithmic systems. The evidence synthesized in this review demonstrates that DM/ML-enabled surveillance often amplifies existing inequities, entrenches feedback loops of discrimination, and undermines fundamental rights to privacy, autonomy, and due process.

Legal and policy frameworks have struggled to keep pace with technological developments, resulting in fragmented protections and inadequate oversight. Technical mitigation strategies, such as fairness-aware algorithms and bias audits, are necessary but insufficient without broader structural reforms, participatory governance, and robust accountability mechanisms. Civil society, NGOs, and affected communities play a crucial role in exposing abuses, advocating for rights, and shaping the future of surveillance governance.

Future research must prioritize empirical evaluation, intersectional analysis, and the development of redress mechanisms that center the experiences and needs of marginalized groups. The proposed hypothesis offers a pathway for rigorous, data-driven inquiry into the mechanisms by which algorithmic surveillance perpetuates harm. Ultimately, the challenge is not merely technical but fundamentally political and ethical: ensuring that the deployment of DM/ML in state surveillance serves the public good without sacrificing the rights and dignity of the most vulnerable.