

AI-Led Governance in Hospital & Healthcare Management in India: A narrative review of applications, frameworks and regulatory considerations

Meghna Mishra¹, Bincy Pothen²

¹Research Scholar, Department of Hospital Administration, Shri Guru Ram Rai University, Dehradun, Uttarakhand, India
ORCID: 0000-0001-6795-283X

²Associate Professor, Department of Hospital Administration, Shri Guru Ram Rai University, Dehradun, Uttarakhand, India

DOI: <https://doi.org/10.5281/zenodo.23119629>

Published Date: 03-October-2026

Abstract: Artificial intelligence (AI) is reshaping healthcare through prediction, classification, and automation, and Indian hospitals are beginning to adopt machine learning (ML), natural language processing (NLP), computer vision (CV), and robotic process automation (RPA) across both clinical and administrative functions. This narrative review synthesises the emerging evidence on AI-led governance and hospital management in India, tracing how these techniques are being used to strengthen diagnostic accuracy, forecast patient demand, and optimise bed and staff allocation in real time. The evidence points to genuine but uneven progress: national instruments such as NITI Aayog's #AIforAll strategy and the Digital Personal Data Protection (DPDP) Act, 2023 establish a policy direction, yet institutional readiness, workforce capacity, and enforcement mechanisms remain inconsistent across public, private, and charitable providers. Persistent concerns include diffuse accountability, uneven data protection, and the black-boxing of clinical algorithms whose reasoning administrators and clinicians cannot easily inspect. To address these gaps, this review proposes a five-layer governance framework, built so that each layer's functions are mutually exclusive and, together, collectively exhaustive of the governance task, spanning data infrastructure, analytical intelligence, operational application, governance and safeguards, and measured outcomes. A situated-ethics approach, one that treats India's geographic disparity and uneven infrastructure as starting conditions rather than afterthoughts, runs through the analysis. Good AI governance, this review argues, treats AI as a monitored and accountable partner in hospital administration rather than a self-sufficient solution.

Keywords: Artificial intelligence; healthcare governance; hospital management; digital health; algorithmic governance; India.

1. INTRODUCTION

Healthcare governance is the web of structures, processes, relationships, and accountability arrangements that regulate how health services are planned, financed, delivered, monitored, and improved. Hospital management is its most visible operational expression, covering patient flow, staffing, procurement, quality assurance, clinical support, financial control, infrastructure upkeep, and communication with patients and regulators. In India, these functions play out against an unusually varied backdrop: public hospitals sit alongside private corporate networks, charitable institutions, teaching hospitals, small nursing homes, and a fast-growing set of digital-health platforms, each with its own governance culture and resource base.

The pressures on this system extend well past patient volume. Geographic inequality, uneven institutional capacity, an unequal distribution of specialists, patchy access to diagnostic equipment, and the survival of paper-based or disconnected information systems all complicate the picture. Even basic planning figures are hard to pin down: estimates of workforce density and hospital-bed numbers vary considerably depending on which denominator, data source, and reference year a study uses, and this measurement inconsistency is itself a governance problem, since administrators cannot plan efficiently around numbers they cannot trust. The Ayushman Bharat Digital Mission aims to improve the portability and linkage of health information, but digital infrastructure by itself is not governance (Mishra et al., 2024) .

AI differs from routine digitisation in one crucial respect: it can detect patterns, generate predictions, classify information, recommend actions, and automate decisions, rather than simply recording them (Beam & Kohane, 2018; Rajkomar et al., 2019). A conventional hospital information system registers admissions and discharges after the fact; an AI-enabled system can use historical and real-time data to forecast emergency department demand, flag patients at risk of clinical instability, project medicine consumption, or surface anomalous billing patterns (Obermeyer & Emanuel, 2016). Used well, these capabilities can shift hospital management from after-the-fact reporting toward anticipatory planning. Prediction is not truth, however, and automation is not the same as responsibility. An algorithm can just as easily reproduce the biases in its training data, entrench existing inequities, or introduce new risks when deployed outside the setting in which it was validated (Amann et al., 2020; London, 2019) .

India sharpens each of these concerns. A model trained on data from a high-resource tertiary hospital may perform quite differently in a district hospital with lower-quality imaging, fewer specialists, less reliable connectivity, or a different patient mix (He et al., 2019). Linguistic diversity, together with the coexistence of paper and electronic records, adds further difficulty for natural language processing and data integration. And because budgets are finite, the cost of acquiring an AI solution has to be weighed against equally urgent claims on staff recruitment, medicines, laboratory capacity, infection control, and basic infrastructure.

This review treats AI governance as a socio-technical undertaking rather than a set of standalone applications, and asks three questions: which AI techniques and application areas matter most for healthcare governance and hospital management in India; what conditions must hold before those applications deliver genuine value; and how Indian hospitals can judge AI on quality, safety, equity, efficiency, accountability, and patient experience. Its scope covers hospital management, clinical decision support, infrastructure, resource management, and digital-health governance and policy. It does not attempt a systematic estimate of diagnostic accuracy for any single disease, nor does it treat any commercial AI product as independently validated simply because it is discussed here.

Several reviews already cover AI in medicine, prediction, clinical imaging, precision medicine, and ethical governance (Char et al., 2018; Floridi et al., 2018; Miotto et al., 2018; Topol, 2019). Fewer connect hospital management, institutional accountability, and India-specific implementation contexts within a single governance lens; a companion review, for instance, maps national and international guidance on privacy and data protection for AI-driven diagnostics in the Indian public health system but does not extend to hospital administration or resource management. This review fills that adjacent gap by bringing clinical and administrative applications into one framework and by separating what the technology promises from what the evidence so far shows about its organisational value (Joshi et al., 2026).

2. METHODOLOGY

This review follows a narrative design, built on a structured literature search and a systematic screening process. A narrative approach was chosen because the literature spans a wide range of study designs, AI technologies, clinical fields, organisational settings, and policy instruments; this heterogeneity makes a pooled effect estimate inappropriate. Reporting follows the Scale for the Assessment of Narrative Review Articles (SANRA) criteria (Baethge et al., 2019); PRISMA was not applied, since the review synthesises conceptual, regulatory, and technical literature rather than pooling quantitative outcomes across comparable study designs. Even so, the review used explicit search concepts, defined eligibility criteria, and a documented screening rationale throughout.

The search covered PubMed/MEDLINE, Scopus, Web of Science, Google Scholar, IEEE Xplore, and ScienceDirect, together with the websites of the World Health Organization, the Government of India, the National Health Authority, NITI Aayog, the Indian Council of Medical Research, the European Union, and the United States Food and Drug Administration. It included literature published between January 2015 and August 2026, while retaining earlier landmark papers where they were foundational to the conceptual development of AI in healthcare.

Studies qualified if they addressed AI applications, governance, hospital operations, digital-health infrastructure, healthcare policy, ethics, regulation, or implementation; were published in English; and gave enough detail on the technology, context, governance challenge, or outcome under discussion. Papers, policy documents, technical standards, and reports from recognised institutions were all eligible. Publications unrelated to health services, purely promotional material without supporting evidence, duplicate studies, and papers offering only a theoretical algorithm with no demonstrated link to healthcare delivery or management were excluded.

3. CONCEPTUALIZING AI-LED GOVERNANCE

Hospital governance has traditionally rested on professional hierarchy, committees, standard operating procedures, audits, and reporting. These mechanisms remain necessary: healthcare carries legal obligations, depends on professional expertise, and answers to public accountability in ways few other sectors do. Their limitation is speed, not legitimacy, they are often too slow to keep pace with rapidly shifting needs. AI-based governance does not replace boards, medical superintendents, clinicians, nurses, managers, or regulators; it changes the kind and timeliness of the data on which those actors already rely, feeding continuous analysis and prediction into structures that would otherwise depend on periodic review (Floridi et al., 2018; Morley et al., 2020). Reddy et al. (2023) reach a related conclusion from a different starting point, proposing a governance model built around explicit accountability, transparency, and clinical-safety functions rather than around a data-processing pipeline. The two views are complementary rather than competing: where their model asks who is answerable for an AI system, the layered model below asks where in the system each safeguard has to sit (Reddy, 2023).

This review proposes a five-layer framework, shown in Figure 1, deliberately built to satisfy the MECE test: each layer covers a function no other layer covers (mutually exclusive), and the five layers together account for every governance-relevant activity identified in the literature (collectively exhaustive). The first layer is data infrastructure: electronic health records, laboratory and imaging systems, the pharmacy information system, billing records, staffing databases, sensor data, and public health data, everything the rest of the framework depends on being available in usable form. The second layer, analytical intelligence, is where that data is turned into inference: statistical learning, deep learning, language models, computer vision, optimisation, and anomaly detection. The third layer, operational application, is where a model's output is embedded into a specific workflow: patient flow and triage, scheduling and bed management, clinical decision support, diagnostics, supply chain, workforce planning, and financial controls. The fourth layer, governance and safeguards, covers the oversight activities that apply across every application regardless of which one is running: informed consent, accountability, privacy and cybersecurity, bias and equity audits, explainability, procurement controls, human oversight, and model monitoring, together with the mechanisms by which a decision can be contested or appealed. The fifth layer, outcomes, is where effects are measured: quality and safety, access and equity, efficiency, patient experience, workforce trust, and public accountability.

Two of these layers are easy to conflate, since procurement and maintenance decisions are made about specific applications. This review places them in the governance layer rather than the application layer because the same procurement standard, the same maintenance-contract review, and the same audit trail apply whether the tool in question is a scheduling algorithm or a diagnostic one. Keeping oversight functions in a single layer, rather than restating them separately for every application, is what keeps the model mutually exclusive; without that separation, a governance requirement would have to be duplicated for each new use case, which is exactly the redundancy a MECE structure is meant to remove.

The five layers are not independent of one another. Poorly structured data cannot be rescued by a sophisticated model; a technically sound model can still harm patients if it is introduced into a workflow without redesigning that workflow and training the staff who use it; and even a well-designed application can fail through inadequate procurement, weak accountability, or neglected maintenance. AI-based governance, on this view, is a recurring cycle of data collection, analysis, translation into action, monitoring, auditing, and improvement, not a one-time deployment (World Health Organization, 2021).

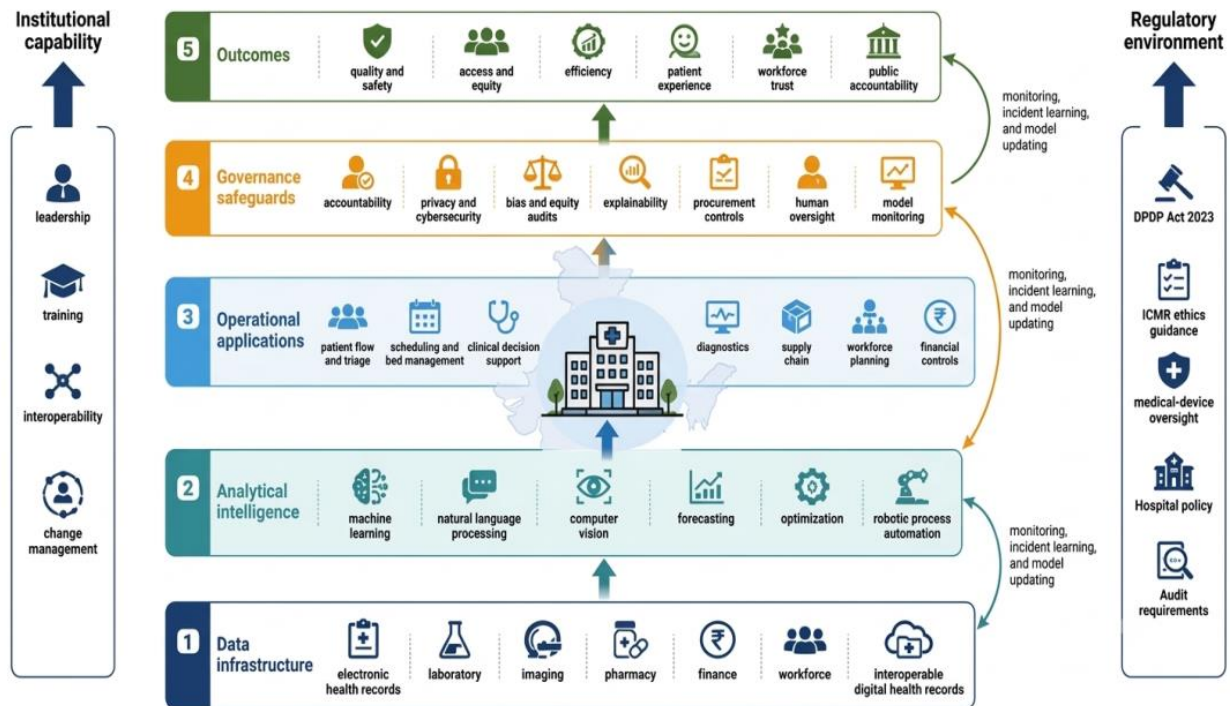


Figure 1: Multi-layered governance framework for AI in Indian healthcare

4. AI MODELS AND TECHNIQUES IN HEALTHCARE GOVERNANCE

4.1 Machine learning and predictive analytics

Machine learning algorithms find patterns in data and use them to classify, estimate, or predict outcomes. In hospital management, this translates into forecasting patient arrivals, estimating length of stay, predicting readmission risk, projecting operating-room demand, and informing staffing decisions (Miotto et al., 2018; Rajkomar et al., 2019). Supervised learning applies where outcome labels already exist; unsupervised methods handle clustering and anomaly detection instead. Deep learning tends to suit images, signals, and text, but generally needs larger datasets, more computing power, and closer supervision than simpler methods (Beam & Kohane, 2018). A predictive model's performance cannot be judged on discrimination or accuracy alone. Calibration, false-positive rate, subgroup performance, clinical applicability, and effect on workflow all matter just as much, arguably more, since a model that correctly predicts high demand is worthless if that prediction never changes a staffing or bed-allocation decision. Local validation is especially important in the Indian context, where case mix, referral patterns, coding practice, and access to care differ sharply across settings (Rajkomar et al., 2019).

4.2 Natural language processing

Natural language processing converts unstructured text into a structured form that can be searched and analysed. In hospital settings, it can extract diagnoses from discharge summaries, summarise records, support procedure coding, flag adverse events, translate patient instructions, and enable multilingual communication (Shickel et al., 2018). Its central difficulty is context-dependence: negation, abbreviation, code-mixing across languages, misspelling, and missing information can all distort output in ways that are not obvious from the text alone. Any system that generates language, rather than merely extracting it, also needs safeguards against fabricated statements, data leakage, and clinically inappropriate advice.

4.3 Computer vision

Computer vision analyses images and video, and has been applied in dermatology, retinal screening, radiology, pathology, surgical video review, and safety monitoring (Esteva et al., 2017; Gulshan et al., 2016). For hospital governance specifically, it can support quality checks, equipment monitoring, infection-control assessment, and the prioritisation of imaging worklists. Performance depends heavily on image quality and on how representative the training population is; a system trained on images from a different country, scanner, or demographic group needs fresh validation before it is trusted on an Indian population (He et al., 2019).

4.4 Robotic process automation and intelligent automation

Robotic process automation (RPA) uses rules-based software to carry out repetitive digital tasks; pairing it with AI adds the ability to classify, predict, or interpret language. In hospital administration, this covers appointment reminders, claims-document validation, invoice matching, report routing, and record reconciliation (**Table 1**). Automation eases clerical workload, but it can also propagate an error through every interconnected system it touches almost instantly. Hospitals therefore need exception handling, access control, audit trails, and a clear process by which staff can override or correct an automated decision (Davenport & Kalakota, 2019).

Table 1: AI techniques, application domains, and governance-relevant outcomes

Technology	Application domains	Governance-relevant outcomes
Machine learning	Demand forecasting, readmission risk, staffing, length of stay	Potentially earlier planning and better risk stratification; requires local calibration
Natural language processing	Clinical-document extraction, coding, summarisation, patient communication	Improved information retrieval and reduced clerical work; risks include ambiguity and hallucination
Computer vision	Radiology, pathology, retinal imaging, safety monitoring	Faster prioritisation and decision support; performance depends on image quality and population
Automation / RPA	Scheduling, claims, billing, reconciliation, reminders	Reduced repetitive workload; requires exception management and auditability
Optimization	Bed allocation, operating rooms, logistics, workforce rosters	More efficient resource allocation; objectives must include fairness and service quality

5. AI IN HOSPITAL ADMINISTRATION

Administrative processes are among the most realistic starting points for AI adoption, because they already generate structured data and involve repetitive, rule-governed decisions (Jiang et al., 2017). Patient-flow systems that link emergency admissions, inpatient bed occupancy, discharge predictions, operating-room schedules, and diagnostic turnaround time can give an administrator a real-time view that surfaces a bottleneck well before it would appear in a monthly report. Realising that benefit, however, depends on timely data processing, integration across departments, and enough managerial authority and resources to act on the alerts the system generates.

Appointment and queue-management tools can forecast waiting times, prioritise emergencies, reduce missed appointments, and balance demand against available capacity. AI-enabled revenue-cycle management can flag incomplete claims, duplicate submissions, coding inconsistencies, and suspicious billing patterns (Reddy et al., 2019). These tools need careful handling: an anomaly is not automatically fraud, and an automated claim rejection can penalise a patient or a provider if the underlying data was simply incomplete rather than dishonest.

Readiness for these systems varies widely across Indian hospitals. Large private chains and tertiary academic centres typically have in-house IT teams, data warehouses, clinical-informatics expertise, and the budget to integrate disparate information systems; smaller hospitals more often struggle with licensing, connectivity, security, data cleaning, and staff training. The public sector's e-hospital initiative and related digital-health policies can support standardisation, but interoperability is not achieved by installing software: it requires agreed identifiers, shared data definitions, exchange standards, governance agreements, and sustained technical support (Mishra et al., 2024).

A claim that a named institution uses a particular AI system needs corroboration, an institutional report, independent evaluation, procurement documentation, or a published case study before it can be treated as evidence. Generic references to well-known Indian hospitals should not stand in for documented implementation and outcomes; a literature review that wants to keep its academic credibility has to hold that line consistently.

Table 2: AI applications in Indian hospital management and clinical delivery

Domain	AI application	Impact on healthcare delivery
Hospital operations	Real-time capacity management	Optimised patient flow and reduced bottlenecks in high-volume settings
Diagnostics	Early cancer and retinopathy screening	Supported earlier intervention in rural and underserved populations
Supply chain	Inventory and logistics optimisation	Supported continuity of medical resources and reduced avoidable wastage
Clinical care	Precision medicine and treatment tailoring	Supported diagnostic accuracy and more individualised therapeutic planning
Security	AI-assisted fraud and anomaly detection	Potentially strengthened transparency and security in administrative processes

6. AI IN HEALTHCARE DELIVERY

AI-supported healthcare delivery spans clinical decision support, imaging interpretation, risk prediction, teleconsultation, patient monitoring, and treatment-planning assistance. The strongest evidence so far comes from imaging classification and prioritisation dermatology and diabetic-retinopathy screening among them (Esteva et al., 2017; Gulshan et al., 2016). Strong performance on a benchmark dataset, though, does not automatically translate into better patient outcomes: referral pathways, clinician response, diagnostic confirmation, treatment choice, and patient adherence all sit between a model's output and the outcome it is meant to improve.

Where radiologists, pathologists, and other specialists are in short supply, AI may help extend the reach of the specialist expertise that does exist. Language processing in particular can support telemedicine consultation and triage. Access, though, depends on more than the algorithm: smartphone penetration, internet connectivity, digital literacy, language, disability, women's access to technology, and the cost of an additional consultation all shape whether a given patient can actually use the service.

Surgical robotics and intelligent monitoring can improve procedural precision and catch clinical deterioration earlier in some cases, but these technologies are expensive and demand ongoing maintenance, clinical governance, and training. Hospitals adopting them should compare interventions directly and report on complications, length of stay, patient experience, and treatment cost, rather than relying on vendor claims alone (Topol, 2019).

Human oversight has to remain real at every step. Clinicians need to understand an algorithm's purpose, limitations, contraindications, and failure modes, and patients should never be led to treat the algorithm as an authority in its own right. Consent and communication processes should make clear whether AI is being used, where a patient can raise a concern, and who holds the final decision (Amann et al., 2020; Char et al., 2018).

7. AI IN INFRASTRUCTURE OPTIMIZATION

Infrastructure optimisation covers beds, operating rooms, diagnostic equipment, energy, water, transport, biomedical devices, and physical space. Predictive analytics can anticipate admission and discharge patterns, speed up bed turnover, and flag bottlenecks in laboratory or radiology workflows. IoT sensors can report equipment temperature, utilisation, vibration, and power consumption, letting maintenance teams anticipate failures rather than react to them. Smart-hospital initiatives are often described as seamless, holistic systems, but integration itself creates new dependencies: sensors can misreport, devices can run on incompatible protocols, network outages can interrupt monitoring, and a cybersecurity gap can expose operational or personal data. Responsible implementation therefore needs clear service-level expectations, manual fallbacks, defined maintenance duties, and an incident-reporting mechanism (Davenport & Kalakota, 2019).

Building the business case for infrastructure AI means costing the whole life cycle, not just the purchase price. Integration, validation, updates, cybersecurity, training, storage, support, and eventual replacement all add to that initial figure, and a fair assessment weighs total cost of ownership against the service outcomes actually delivered, not against a vendor's initial quote.

8. AI IN RESOURCE OPTIMIZATION

Staffing-management systems can forecast personnel needs, flag scheduling conflicts, and assign nurses, technicians, and support staff. Used carefully, they reduce administrative burden, but the optimisation target has to include safe staffing levels, skills mix, fatigue, leave entitlements, professional autonomy, and staff well-being, not headcount alone. On the

supply side, AI can forecast demand for medicines, consumables, reagents, blood products, and protective equipment, flag consumption anomalies, support reorder points, and cut both stockouts and overstock. Those forecasts have to account for seasonality, disease outbreaks, procurement delays, supplier reliability, expiry dates, and the fact that historical consumption often reflects past shortages rather than genuine demand (Obermeyer & Emanuel, 2016).

Financial governance can draw on similar tools for automation, coding support, anomaly detection, and fraud-risk assessment. But an automated financial algorithm can penalise institutions that treat a more complex or more disadvantaged patient population, purely as an artefact of the data it was trained on; any fraud-detection system therefore needs human review, defined thresholds, an appeals process, and regular evaluation for systematic bias (London, 2019).

A claimed saving or efficiency gain is only credible when the underlying study defines its baseline, intervention, comparator, timeframe, and method of financial or statistical analysis. Without that evidence, it is more honest to describe an application as having potential than to attach a specific percentage improvement to it.

9. REGULATORY AND POLICY LANDSCAPE IN INDIA

AI governance in India draws on several overlapping instruments: data-protection legislation, digital-health architecture, professional guidance, public-sector strategy, sectoral regulation, and institutional policy. The Digital Personal Data Protection Act, 2023, governs the collection and processing of digital personal data, though its practical effect still depends on the final rules, their implementation, and their interaction with other legislation (Joshi et al., 2026). Hospitals also carry confidentiality obligations, cybersecurity duties, and, where applicable, medical-device regulation, on top of whatever contractual obligations they owe their technology vendors.

The National Strategy on AI, #AIforAll, singled out socially useful applications and placed healthcare among its priorities (Roy Anna, 2017). The Ayushman Bharat Digital Mission aims at interoperable digital health through a health identifier, shared registries, and consent-driven information exchange (Mishra et al., 2024). More recently, the Ministry of Electronics and Information Technology released the India AI Governance Guidelines in November 2025, setting out a cross-sectoral framework built around seven guiding principles and recommending a coordinating body for AI oversight rather than a single new AI-specific law (Sood Ajay Kumar & Krishnan S., 2025). The guidelines apply to AI broadly and are not written for hospitals specifically, so how they will interact with sector rules such as medical-device regulation and the DPDP Act should become clearer once health-specific implementing guidance follows. None of these instruments, on its own, validates a given algorithm or settles the question of liability. The ICMR's guidance on biomedical and health research using AI addresses ethical review, scientific validity, safety, transparency, and responsible conduct in research specifically, and hospitals need to distinguish that research activity clearly from clinical-tool deployment; a research prototype may need ethics approval and testing, while a clinical tool needs credentialing, safety monitoring, documentation, and its own regulatory pathway (Kundu & Bardhan, 2025).

International models offer useful points of comparison, though not a template to copy wholesale. The EU's AI Act sets out a risk-based regulatory structure with tighter requirements for high-risk uses. In the United States, the FDA has built specific pathways and guidance for software as a medical device, AI/ML-enabled tools included. Reddy (2023) argues that regulation aimed at healthcare AI needs to be precise rather than generic, since a rule written for AI in general tends to under- or over-regulate any specific clinical use; that argument applies with particular force in India, where the health system already mixes public and private providers, uneven resources, linguistic diversity, and ambitious digital-health goals (Reddy, 2023). Importing a foreign framework wholesale, without adapting it to that mix, risks adding compliance burden without a matching gain in patient safety (Chauhan et al., 2025).

Table 3: Key Indian regulatory and policy initiatives for AI in healthcare, with comparative international references

Initiative	Lead agency	Core focus	Key impact / status
#AIforAll strategy	NITI Aayog	National AI roadmap	Established a national vision and exploratory health applications
DPDP Act 2023	Ministry of Law and Justice	Data protection	Created foundational obligations for data fiduciaries and personal-data processing
AI ethics guidance	ICMR	Clinical research ethics	Provided sector-specific principles for safer and more equitable AI adoption

Ayushman Bharat Digital Mission	National Health Authority	Digital health infrastructure	Established ABHA identifiers, registries, and consent-based health-data exchange
India AI Governance Guidelines	MeitY	Cross-sectoral AI governance	Set out seven guiding principles and a coordinating-body model; healthcare-specific guidance still pending
EU AI Act	European Union	Risk-based AI regulation	Introduced tiered obligations for high-risk AI systems, informing comparative analysis
AI/ML-enabled device guidance	US FDA	Medical-device regulation	Provided a comparative regulatory pathway for software as a medical device

10. CHALLENGES AND ETHICAL CONSIDERATIONS

Data quality is one of the most basic constraints. Hospital datasets carry missing values, inconsistent coding, duplicate records, free-text notes, delayed entry, terminology that differs across departments, and the general fragmentation of a mixed public-private system. Interoperability, in this setting, is not a technical connection problem so much as a semantic one: it depends on shared meaning, governance, identity, consent, security, and agreement on acceptable use (Shickel et al., 2018).

Algorithmic bias can arise from underrepresentation, measurement disparity, structural inequity, selection into treatment, and the use of surrogate variables as stand-ins for what a model actually needs to predict. A model trained on patients who reached a tertiary hospital may perform poorly for patients who never got there in the first place. The clearest illustration of how a surrogate variable distorts outcomes comes from outside India: Obermeyer et al. (2019) found that a widely used US commercial algorithm, built to predict health-care cost rather than illness itself, under-identified Black patients for extra care because less money had historically been spent on their treatment for the same level of need; correcting the proxy variable, not the algorithm's accuracy, removed most of the disparity (Obermeyer et al., 2019). The lesson generalises directly to an Indian setting that also spends unevenly across region, provider type, and socioeconomic group: any Indian model built on billing or utilisation data as a proxy for health need carries the same structural risk, and bias auditing has to examine subgroup performance, missing-data patterns, calibration, and the balance of false positives against false negatives, not overall accuracy alone (Amann et al., 2020; Morley et al., 2020). Explainability matters because clinicians, administrators, patients, and regulators each need to understand why a recommendation was made, and in language suited to that audience; an explanation that implies more certainty than the model actually has is worse than no explanation at all. In some cases, a simpler, more interpretable model is the right choice even when a more accurate but opaque alternative is available (London, 2019).

Interconnected systems raise the stakes on privacy and cybersecurity, adding data theft, ransomware, data leakage, model tampering, and secondary misuse to the list of risks a hospital has to manage. The practical response includes data minimisation, role-based access control, encryption, logging, incident response, vendor due diligence, retention controls, and clear information to patients about how their data is used, wherever law and clinical need allow it (Chauhan et al., 2025). Accountability has to be assigned before deployment, not after something goes wrong. Contracts should specify who verifies a system, who maintains it, who conducts periodic updates, who investigates incidents, and who communicates with patients. Human oversight only counts as oversight if it is substantive: if clinicians feel pressured to defer to an algorithm regardless of their own clinical judgement, the presence of a human in the loop satisfies the letter of the requirement without meeting its purpose (Char et al., 2018; Floridi et al., 2018).

11. DISCUSSION

Across every domain examined here, the same pattern recurs: AI adds value when it strengthens an existing hospital-management or care-delivery process, not when it stands in as a demonstration of the technology itself. Predictive analytics helps only if hospitals act on its predictions in staffing, bed allocation, procurement, or referral; NLP eases documentation burden only if the underlying records are consistent enough to parse; computer vision aids diagnosis or monitoring only if image acquisition, verification, and follow-up are themselves reliable; automation reduces repetitive work only if its exceptions and errors stay visible to someone who can act on them (Davenport & Kalakota, 2019; Topol, 2019). The evidence base reflects this unevenly: technical feasibility for image analysis and prediction is comparatively well established, while evidence on governance outcomes at the level of the whole hospital remains thin. Most published work reports technical performance on retrospective datasets; far fewer studies track implementation over time, cost, equity, workforce impact, patient experience, or the negative outcomes that a purely technical evaluation would miss. That gap matters for administrators specifically, because a model can be technically accurate and still fail in practice, through alert fatigue, workflow mismatch, weak adoption, or an inability to act on what the model outputs.

India shares much of this predicament with other large, heterogeneous lower- and middle-income health systems: patchy infrastructure, limited specialist access, and a strong appetite for digital solutions. AI could help these systems clear some administrative hurdles, but it could equally widen inequality if high-quality AI ends up concentrated in urban, private hospitals alone. Judging public AI deployment on innovation grounds only is therefore insufficient; scope, cost, sustainability, and distributive justice belong in the same assessment. Four priorities follow from this. First, hospitals need a reliable data foundation before attempting complex models. Second, every AI product needs a defined purpose, a named owner, a validation procedure, monitoring indicators, and a plan for retiring it if it stops earning its place. Third, procurement should weigh interoperability, security, transparency, maintenance, and total cost ahead of a vendor's sales pitch. Fourth, implementation should involve clinicians, nurses, administrators, patients, legal counsel, information-security staff, and representatives of disadvantaged groups from the outset, not as a late-stage consultation (**Fig. 2**).

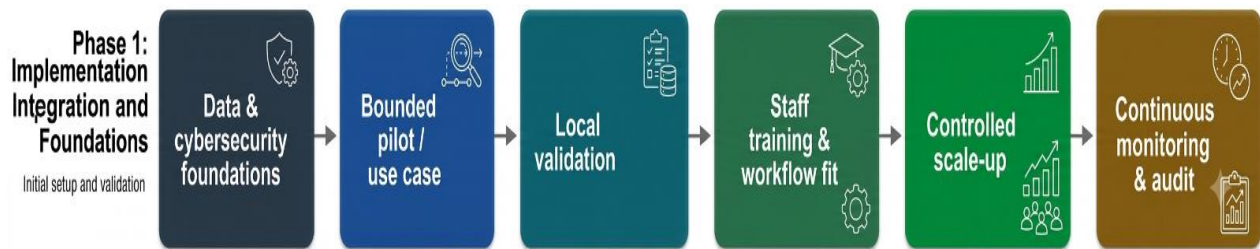


Figure 2: Incremental path of AI implementation in hospitals

Consistent with the layered framework above, judging AI-assisted governance well requires more than an efficiency metric (**Fig. 3**). Efficiency indicators include waiting time, bed turnover, length of stay, stock-outs, and processing time. Safety indicators include adverse events, missed warnings, and inappropriate recommendations. Equity indicators include subgroup calibration, access by location and socioeconomic status, and differential referral or denial rates across population groups. Organisational indicators include workload, training coverage, staff trust, and the practical ability to override the system when it is wrong.

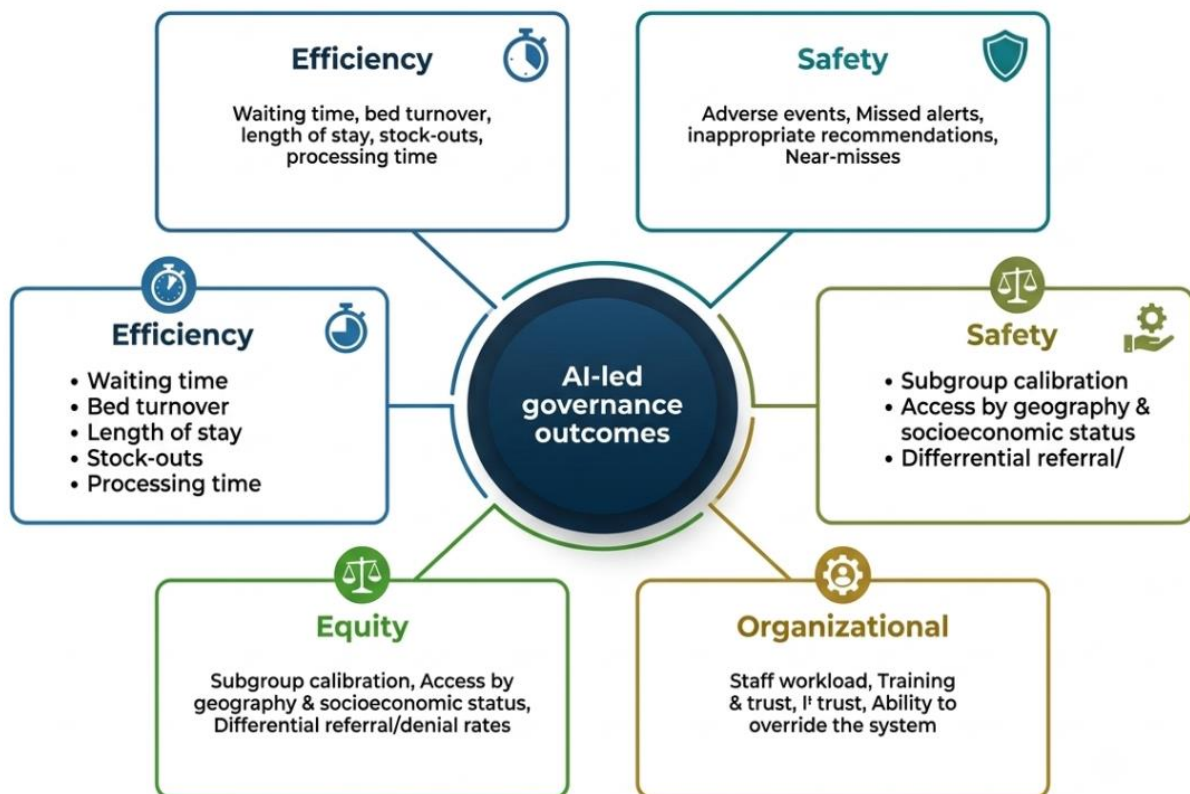


Figure 3: Multidimensional approach to the evaluation of AI-led governance outcomes

Table 4: Exemplary outcome measures mapped to the five-layer governance framework

Outcome dimension	Illustrative indicators	Primary governance layer
Efficiency	Waiting time, bed turnover, length of stay, stock-outs, processing time	Operational application layer
Safety	Adverse events, missed alerts, inappropriate recommendations, near-misses	Governance and safeguards layer
Equity	Subgroup calibration, access by geography and socioeconomic status, differential referral/denial rates	Outcome layer
Organisational	Staff workload, training coverage, trust in the system, safe override rate	Governance and safeguards layer

12. LIMITATIONS

This review carries several limitations. It is a narrative synthesis, not a registered systematic review or meta-analysis, so evidence selection and interpretation inevitably reflect the authors' judgement. The search was restricted to English-language sources, which likely underrepresents relevant evidence published in Indian languages. The literature itself is heterogeneous, technical articles, implementation reports, policy documents, ethical analyses, and conceptual papers sit alongside each other and given how quickly new AI tools appear, not all of it has been rigorously evaluated; implementation detail is often simply not made public. Finally, the working manuscript underlying this revision did not retain a search log or screening file, so record counts are not reported here rather than estimated. Going forward, the authors need to keep and archive the search strategy, screening decisions, an extraction table, and a reference-verification file as the manuscript moves toward submission.

13. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

AI-led governance could meaningfully change how Indian hospitals operate, through earlier prediction, better-coordinated services, more efficient resource use, and wider access to certain forms of clinical expertise. What matters more than any single algorithm, though, is the institution's capacity to use these tools well. AI is best understood as a socio-technical instrument for professional practice, governed by hospital policy, public regulation, and patients' rights, not as a self-directing solution. The most defensible path to implementation is incremental: start with low-risk applications, build data and cybersecurity foundations first, validate performance locally, train users, and monitor results continuously. Higher-risk clinical applications should face a correspondingly higher bar, more evidence, independent assessment, thorough documentation, and clearly assigned liability for error. Public hospitals, in particular, need support toward open standards, technical assistance, and procurement practices that avoid locking them into a single proprietary vendor.

Future research on AI in Indian hospitals would benefit most from pragmatic, multicentre studies that report baseline context, comparator processes, implementation cost, user adoption, patient satisfaction, and outcomes over an adequate follow-up period, broken down by sex, age, language, geography, socioeconomic status, comorbidity, and hospital type. Particular gaps remain in multilingual clinical NLP, low-resource imaging, interoperability between public and private systems, and AI designed specifically for district hospitals. Governance research should examine how responsibility is divided between vendors and hospitals, how patients can contest an algorithmic decision in practice, which explanation approaches actually help end users, and what procurement should require in terms of auditability, notification of model change, and post-market monitoring. Health-economics research should weigh AI investment against competing claims on limited hospital budgets, while implementation science can identify the organisational and human factors that make AI's value sustainable rather than a one-off pilot result.

Responsible AI governance in India cannot rest on technical capability alone. It needs trustworthy data, organisational readiness, ethical guidance, regulatory clarity, measurable impact, and genuine human oversight working together. The point of AI governance is not to make hospitals more algorithmic for its own sake, but to make them safer, more responsive, more equitable, and more accountable than they were before.

Declarations

Author Contributions

Meghna Mishra: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing - original draft.

Bincy Pothen: Supervision, Validation, Writing - review & editing.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare no competing interests.

Acknowledgements

The authors express sincere gratitude to Shri Guru Ram Rai University, Dehradun, for institutional support.

REFERENCES

- [1] Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20(1), 310. <https://doi.org/10.1186/s12911-020-01332-6>
- [2] Baethge, C., Goldbeck-Wood, S., & Mertens, S. (2019). SANRA—a scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review*, 4(1), 5. <https://doi.org/10.1186/s41073-019-0064-8>
- [3] Beam, A. L., & Kohane, I. S. (2018). Big Data and Machine Learning in Health Care. *JAMA*, 319(13), 1317. <https://doi.org/10.1001/jama.2017.18391>
- [4] Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *New England Journal of Medicine*, 378(11), 981–983. <https://doi.org/10.1056/NEJMp1714229>
- [5] Chauhan, S. B., Gaur, R., Akram, A., & Singh, I. (2025). Artificial Intelligence Driven insights for Regulatory Intelligence in Medical Devices: Evaluating EMA, FDA and CDSCO Frameworks. *Global Clinical Engineering Journal*, 7(2), 11–24. <https://doi.org/10.31354/globalce.v7i2.210>
- [6] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94–98. <https://doi.org/10.7861/futurehosp.6-2-94>
- [7] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- [8] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- [9] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, 316(22), 2402. <https://doi.org/10.1001/jama.2016.17216>
- [10] He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, 25(1), 30–36. <https://doi.org/10.1038/s41591-018-0307-0>
- [11] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243. <https://doi.org/10.1136/svn-2017-000101>
- [12] Joshi, U., Baisil, S., Kini B, S., Mehta, M., & Datta, S. (2026). AI ethics in Indian healthcare: a scoping review of national and international guidelines on privacy, data protection, and security. *BMC Medical Ethics*, 27(1), 86. <https://doi.org/10.1186/s12910-026-01435-1>
- [13] Kundu, T., & Bardhan, M. (2025). Artificial intelligence in neurology, ethics, recent guideline, and law-an Indian perspective. *Frontiers in Neurology*, 16. <https://doi.org/10.3389/fneur.2025.1515041>
- [14] London, A. J. (2019). Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Center Report*, 49(1), 15–21. <https://doi.org/10.1002/hast.973>

- [15] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6), 1236–1246. <https://doi.org/10.1093/bib/bbx044>
- [16] Mishra, U. S., Yadav, S., & Joe, W. (2024). The Ayushman Bharat Digital Mission of India: An Assessment. *Health Systems & Reform*, 10(2). <https://doi.org/10.1080/23288604.2024.2392290>
- [17] Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine*, 260, 113172. <https://doi.org/10.1016/j.socscimed.2020.113172>
- [18] Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *New England Journal of Medicine*, 375(13), 1216–1219. <https://doi.org/10.1056/NEJMp1606181>
- [19] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- [20] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMra1814259>
- [21] Reddy, S. (2023). Navigating the AI Revolution: The Case for Precise Regulation in Health Care. *Journal of Medical Internet Research*, 25, e49989. <https://doi.org/10.2196/49989>
- [22] Reddy, S., Fox, J., & Purohit, M. P. (2019). Artificial intelligence-enabled healthcare delivery. *Journal of the Royal Society of Medicine*, 112(1), 22–28. <https://doi.org/10.1177/0141076818815510>
- [23] Roy Anna. (2017). *National-Strategy-for-Artificial-Intelligence*.
- [24] Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589–1604. <https://doi.org/10.1109/JBHI.2017.2767063>
- [25] Sood Ajay Kumar, & Krishnan S. (2025). *India AI Governance Guidelines Enabling Safe and Trusted AI Innovation*.
- [26] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- [27] World Health Organization (Ed.). (2021). *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. World Health Organization.