

IoT-Enabled Predictive Analytics for Real-Time Decision-Making in Saudi Digital Service Platforms

Ahmed Helmi 

Published October 2026

Abstract

Saudi digital service platforms are now becoming more reliant on connected devices, the streaming of telemetry data, and automated service orchestration, but the shift from simply observing events to being able to anticipate operational outcomes is still very uneven. This review looks at how Internet of Things (IoT)-based predictive analytics can assist with real-time decision-making in the various digital service environments in Saudi Arabia, focusing on the architectural, analytical and governance conditions needed for a reliable implementation. The study draws together evidence from peer-reviewed publications between 2020 and 2025 covering the topics of IoT, edge intelligence, machine learning, digital platforms, decision support, privacy, reliability and smart-city applications. Instead of viewing prediction as a separate algorithmic activity, the review presents it as a complete operational capability that involves sensing, data-quality control, contextual processing, model inference, decision policies, workflow execution and feedback on the outcomes. The evidence shows that the value obtained depends more on a low-latency architecture, trustworthy data, explainable decision thresholds, resilient connectivity and institutional ownership of the results than on the complexity of the models. For Saudi Arabia, these requirements are especially important since the country's digital transformation includes large-scale investments in infrastructure, high levels of mobile connectivity, smart-city projects and sector-specific service platforms. A reference architecture and a decision lifecycle are suggested in order to link IoT telemetry with predictive service operations while still ensuring security, privacy and human oversight. The review concludes that the most defensible approach is a federated, edge-cloud operating model in which predictive intelligence is incorporated into measurable service workflows and is governed by clear assurance controls.

1 Introduction

The concept of the Internet of Things has now developed from one of connectivity into an operational data infrastructure which enables organisations to keep a continuous watch on physical assets, environments and user interactions. By means of sensors, gateways and connected devices, event streams are generated that can show the condition of assets, service demand, process bottlenecks and emerging risks. Yet merely observing does not lead to decision value; the key strategic move is to quickly turn the heterogeneous, high-velocity telemetry into forecasts and recommended actions so as to have an impact on the developing service situation. Recent research into real-time organisational decision-making indicates that the IoT becomes genuinely useful in a practical sense when it is combined with analytics, managerial processes and measurable operational outcomes [1]. At the same time, studies focused on Saudi Arabia describe the IoT as a major facilitator of digital transformation and of Vision 2030, especially in cases where cloud computing, advanced connectivity and decision-support capabilities are incorporated into both public and private services [2].

Predictive analytics takes this approach a step further by providing estimates of future states rather than just giving a report on the present situation. For instance, a digital service platform could

forecast transaction demand, detect likely service degradation, anticipate equipment failure, predict congestion surrounding a connected service, or prioritize an intervention before a threshold is breached. The fundamental technical pattern is becoming more and more widespread: devices generate signals; edge or cloud infrastructure cleans and combines them; machine-learning models assess risk or demand; a decision module compares the predictions with operational policies; and a workflow engine then initiates an appropriate digital or human response. These types of systems can be found in smart cities, transportation, industrial automation, healthcare, energy management and supply chains [3–7]. What they have in common is not a specific algorithm but a time-sensitive decision loop.

The fact that Saudi Arabia is involved makes this subject particularly important. Digital service platforms are being extended into areas such as government, mobility, utilities, health, logistics, commerce and large urban projects. Saudi research into the Internet of Things has shown that reliable connectivity, scalable applications and the ability to select services that are appropriate to specific contexts are important [2,3]. Investigations into connected services for Hajj and Umrah also illustrate how large-scale and time-critical environments rely on sensing, knowledge of location, information about crowds and effective digital coordination [3]. On the other hand, studies of smart cities indicate that IoT ecosystems can only generate value when data integration, interoperability, analytics and governance all reach a mature stage [19,20]. From these observations it follows that predictive decision-making should be regarded as an enterprise capability rather than as a set of separate pilot projects.

A common fault in the existing literature is that it is fragmented. While some papers focus on device connectivity and network reliability, others pay attention to machine-learning performance, and a number of them talk about decision support or sustainability. However, a predictive service platform has to meet all of these requirements at the same time. A model that is highly accurate is of little use if the telemetry data arrives late, if the identifiers cannot be matched across different systems, if the model's confidence is not turned into an operational rule, or if it is unclear who is responsible for taking action. On the other hand, a straightforward predictive model can generate considerable value provided that the data pipeline is reliable, the action is clearly defined, and feedback is collected in order to enable continuous improvement. Thus, reliable studies in the field of IoT include latency, resource allocation, security and communication quality together with analytical capability [18].

The review looks at the issue of integration. It investigates how IoT-enabled predictive analytics can be arranged in such a way that it supports real-time decisions within Saudi digital service platforms and at the same time remains operationally reliable, governable and scalable. The paper does not suggest a new prediction algorithm; rather, it combines recent evidence into a socio-technical architecture which links data capture, edge processing, analytics, decision logic, platform execution and assurance. The contribution consists of three parts. Firstly, it identifies the capabilities that set descriptive IoT monitoring apart from predictive decision operations. Second, it creates a layered reference architecture and a closed-loop decision lifecycle appropriate for multi-service platforms. Third, it turns the evidence into implementation priorities for organisations in Saudi Arabia, covering latency design, model governance, privacy, interoperability and performance measurement.

2 Aim, objectives and review questions

This review is intended to explore the way in which IoT-enabled predictive analytics can be arranged so as to provide a reliable real-time decision-making capability for the Saudi digital service platforms.

The synthesis is guided by four objectives: to identify the architectural components needed in order to turn IoT telemetry into operational predictions; to look at the analytical methods and deployment patterns that enable low-latency decisions; to assess the governance, security, privacy and reliability requirements that affect actual use; and to develop an implementation framework that is suited to the scale and variety of services involved in Saudi digital transformation.

The article looks at four questions. The first of these is which data and platform layers are needed in order to enable real-time predictive decision-making? The second question is which analytical methods are most appropriate when it is necessary to act on predictions within operational time limits? The third is what kind of governance and assurance measures are required in order to stop inaccurate, insecure or poorly contextualized predictions from being passed on to services? The fourth question is how Saudi organisations can arrange the adoption of predictive analytics so that it leads to tangible service outcomes rather than merely consisting of separate technical demonstrations? These questions intentionally link technical design with organisational implementation since the evidence examined shows that the value of IoT is achieved through decision-making processes, not through connectivity alone [1,18].

3 Review methodology

A structured integrative review approach was adopted since the subject area involves information systems, IoT engineering, machine learning, smart-city platforms, industrial analytics and decision support. The time frame for the review was limited to the period 2020–2025 in order to include current architectures and the governance issues linked with edge computing, federated learning, digital twins and AI-enabled IoT. The searches were based on combinations of terms such as Internet of Things, predictive analytics, real-time analytics, decision-making, edge intelligence, digital platform, smart city, predictive maintenance, machine learning, Saudi Arabia, Vision 2030, privacy, reliability and cybersecurity. Preference was given to peer-reviewed journal articles and in-depth review studies that had traceable digital object identifiers.

The selection process prioritised relevance over the use of a narrow, domain-specific classification system. A study was included if it provided evidence for at least one of the six areas: sensing and connectivity; data engineering and interoperability; predictive modelling; edge or cloud execution; decision support and workflow integration; or governance and assurance. Papers that were specific to Saudi Arabia were incorporated in order to provide context regarding the adoption within the country, and international studies were included when their technical results could reasonably be applied to digital service architectures. Those studies which dealt solely with unrelated algorithmic benchmarks and had no implications for connected services or operational decision-making were excluded. The older foundational works were not included in the final set of evidence since the review had been deliberately limited to the specified period of publication.

The evidence base that has been finalised includes 30 references which have been verified by DOI and are peer-reviewed. The two reference papers provided were used as starting points for defining the terminology and for establishing the analytical framework: one of them looks at IoT-enabled real-time decision-making in organisations [1], the other assesses the selection of IoT applications in the context of Saudi Vision 2030 [2]. The evidence from these papers was supplemented by research on reliable IoT, smart-city frameworks, industrial automation, digital twins, federated learning, sustainable transport, supply chains and security [4–30]. Each of the sources was qualitatively coded according to its architectural layer, the decision role it plays, the latency implications, the trust

requirements and its relevance to service operations. Since the studies have heterogeneous outcomes, datasets and methodological designs, the coding enabled a thematic synthesis rather than a statistical meta-analysis.

The evaluation placed a strong emphasis on methodological transparency, the direct relevance to decisions made in the context of IoT, the clarity of the architecture or analytical method used, and the inclusion of a persistent DOI. Any evidence was assessed in a conservative manner. The associations that were reported were not regarded as indicating universal causal effects, and the algorithmic performance recorded in one domain was not numerically transferred to another. Rather, design principles that reappeared in different contexts were identified. This method is suitable for a review aimed at achieving architectural and managerial synthesis rather than at estimating pooled effects. Hence, the framework that results reflects convergent insights from the recent literature and does not amount to a statement that any one model or platform setup is the best option for every kind of service in Saudi Arabia.

4 From connected data to predictive service intelligence

IoT systems achieve operational visibility by connecting physical or software-observable events to digital platforms, but the predictive value relies on a series of trustworthy transformations. Device data may be delayed, duplicated or semantically inconsistent, and network conditions can change the patterns of arrival. Hence, real-time prediction should be regarded as an accountable pipeline and not merely as the end point of a model.

The initial step involves transforming a raw signal into a trustworthy event. Whether or not telemetry can be used to support time-sensitive applications depends on communications performance, latency, identity and security [18]. As a result, issues such as timestamp integrity, unit normalization, handling of missing data and local buffering are included in the decision architecture. When cloud connectivity is limited, edge processing can be used to filter out noise, extract features and maintain essential service continuity.

The next stage involves a shift from a trusted event to a contextual state. IoT telemetry becomes meaningful only when it is combined with service tickets, asset histories, geospatial information, transaction data or environmental conditions. According to smart-city research, interoperability and the use of shared data structures are necessary if coordinated services are to be provided [19,20]. By means of feature stores, event schemas and stable semantic identifiers the same context is able to support a number of predictive services without the need to rebuild the data logic repeatedly.

The fourth stage involves moving from a contextual state to a predictive estimate. It is possible for models to forecast demand, estimate the remaining useful life, detect anomalies or classify risk. As shown by research into digital twins, there is value in continuously linking the physical state to its computational representation as a means of achieving predictive maintenance [26], and industrial IoT studies have shown how machine learning can interpret the data from connected operations [21]. The prediction horizon, the level of confidence and the frequency of refresh therefore need to be selected on the basis of the intervention window.

The last step is to go from prediction to action. Decision methods are useful in cases where an IoT intervention has to deal with competing factors such as reliability, energy, usability, security and cost [2,11]. A decision layer can combine the model's outputs with service policies, confidence thresholds

and the effect on the user in order to decide whether to observe, recommend, escalate or act automatically. The policy layer thus turns the analytics into controlled operational behaviour.

5 Reference architecture for real-time predictive decisioning

The architecture for Saudi digital service platforms must include five operational layers and at the same time ensure end-to-end observability. Figure 1 shows the suggested structure. The sensing layer consists of connected devices, meters, software agents, mobile endpoints and gateways. Its main role is to ensure trustworthy data capture, not carrying out analytics. Device authentication, clock synchronization, local buffering and health monitoring should all be designed at this layer since the quality of any prediction is limited by the quality of the source.

The next stage is known as edge intelligence. At the edge nodes it is possible to carry out denoising, compression, protocol translation, feature extraction and lightweight inference. This layer cuts down the amount of data that is sent to the central system and ensures resilience if the wide-area connectivity fails. Research into federated learning shows a similar idea: it is possible to train or adapt models across a number of distributed nodes without having to centralise all the raw data, which in turn improves privacy and reduces the amount of data that has to be moved unnecessarily [27]. Edge intelligence is therefore not merely a means of reducing latency; it also functions as a form of governance whenever sensitive or high-volume telemetry should stay near its source.

The third layer is the streaming and context fabric, which takes in events, checks their quality, links them to identities and business context, and then makes them available to various consumers. A modern version of it should offer event-time processing, schema management, lineage tracking, the ability to replay events, and idempotent handling. These features are important since real-time decisions must still be auditable after the event has occurred; if an operator later wants to know why the platform changed the traffic routing, increased capacity or issued an alert, the system must be able to reconstruct the data and the state of the rules that were in place when the decision was taken. This need is what sets enterprise decision streams apart from single-use sensor feeds.

The fourth layer involves predictive analytics, and the models should be deployed taking into account latency, data locality and computational intensity. Simple anomaly detectors can be operated at the edge, while more complex ensemble or deep-learning models might be run in regional or central cloud environments. As shown by AIoT research, combining AI with IoT can lead to improved efficiency and better operational intelligence, but it also brings about problems relating to model management and data quality [28]. It is therefore better to have a tiered model portfolio than to require each application scenario to use a single technical stack. For high-frequency operational models, stability and response time should be given priority, while planning models can make use of longer data windows and more computationally intensive methods.

The fifth layer consists of decision services, which integrate the prediction scores with service policies, business constraints, user impact and confidence thresholds; the actions it takes can involve automated API calls, the creation of workflows, capacity scaling, the generation of recommendations or escalating the matter to a human. It is necessary for the service to record both the prediction and the reasons for the decision so that the results can be assessed at a later stage; closed-loop learning occurs when the actual result is fed back into model monitoring and process improvement.

In all five layers, the trust plane must offer support in the areas of identity, cybersecurity, privacy, model governance and observability. Research into IoT security has on numerous occasions demon-

strated that connected environments increase the size of the attack surface and therefore call for specific assurance measures [4,7,15,18,30]. It is for this reason that the trust plane should incorporate device certificates, least-privilege access, encrypted transport, secure update procedures, model versioning, drift monitoring and records of decisions. The horizontal control plane of this kind is essential for Saudi organisations running platforms that span different sectors since the same analytics infrastructure can be used to support services which have very different risk profiles.

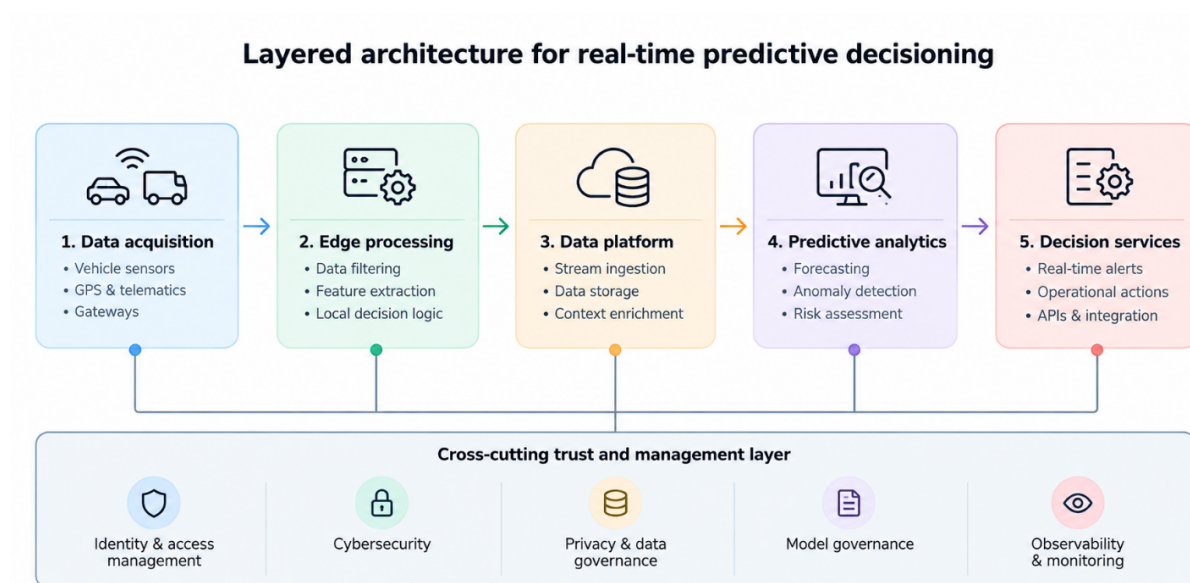


Figure 1. Reference architecture for IoT-enabled predictive decisioning in digital service platforms.

6 Predictive analytics patterns for real-time decisions

Predictive analytics on digital service platforms may be divided into four categories: forecasting, anomaly detection, condition prediction and decision ranking. Forecasting involves estimating future demand or load, for example the expected volume of transactions, the number of transport requests, energy usage or the number of service tickets. The operational value lies in the lead time; if a forecast is available early enough it can be used for capacity allocation, workforce planning or for proactive communication. Accuracy must therefore be assessed together with the horizon and the decision utility. A model that is only slightly less accurate but which gives stable predictions thirty minutes earlier may be more useful than a highly accurate model which only becomes reliable a moment before a service breach occurs.

The detection of anomalies involves identifying behaviour that differs from an expected baseline. IoT streams are particularly well adapted to this approach since they continuously record the current state of a system. Nevertheless, anomaly scores do not mean that an incident has occurred. Service platforms must use contextual thresholds which take into account seasonality, maintenance states and known periods of high demand. The architecture has to tell the difference between new observations, known exceptions and actual operational risk. Doing so helps to reduce alert fatigue and stops automated workflows from escalating harmless deviations.

Estimates of condition prediction involve assessing degradation, failure probability, or remaining useful life. As shown in the literature on digital twins, continuous state representations can aid predictive maintenance by connecting sensor observations to models of asset behaviour [26]. Other

research into maintenance stresses that digitalisation is successful to the extent that technical capabilities are combined with maintenance processes and ownership [12]. This applies likewise to digital service platforms in relation to their software infrastructure, network equipment, field devices and service dependencies. A failure-risk score should be associated with a maintenance or resilience action which has a defined owner, response window and evidence trail.

The process of ranking decisions applies to cases in which more than one action is possible and in which trade-offs are important. Multi-criteria methods have been applied in order to assess IoT applications, service providers, and choices relating to sustainable transport [2,11]. In operational platforms, ranking may take into account the predicted impact, cost, urgency, security risk, energy effects, and customer experience. An advantage of this approach is that it provides transparency, since stakeholders can see the reason why one course of action was chosen. The disadvantage is that the weights and thresholds may end up being political or organizational decisions disguised as mathematical ones. Governance should therefore record who is responsible for setting the criteria, how the weights are reviewed, and when automatic ranking has to yield to human judgment.

The choice of model should take into account the particular decision environment; for instance, tree-based models are suitable when dealing with heterogeneous tabular features and for the purpose of explainable risk scoring, recurrent or temporal models can be useful in cases where there is a strong degree of sequence dependence, probabilistic models allow uncertainty to be expressed directly, and deep learning can be considered appropriate when complex sensor patterns are the main factor. Research into industrial automation indicates the potential of machine learning when applied to IoT data, whereas reviews on AIoT stress that data acquisition and model complexity have to be aligned with the deployment constraints [21,28]. The architecture should allow for multiple models rather than favouring a single algorithm.

Real-time deployment also involves model operations, since version control, ensuring feature consistency, performance monitoring and detecting drift are all essential, as a model which performed well during development may deteriorate if its behaviour changes. Drift can result from the introduction of new devices, a redesign of the service, seasonal demand, changes in policy or alterations in user behaviour. Monitoring should therefore distinguish between data drift, concept drift and operational drift. Operational drift is especially important in this regard: even if the statistical performance stays the same, the business value of a prediction can change when there is a change in service capacity or intervention costs.

Figure 2 shows the final closed loop. The platform carries out observation, validation, prediction, decision-making, action, and learning. At every cycle, decision assurance checks the latency, confidence, fairness, and resilience. This approach draws a clear distinction between real-time analytics and real-time decision-making. It is only when the platform has the authority or a clear way to alter service behaviour that analytics becomes decision-making. As a result, the key performance indicators should comprise action lead time, the number of incidents prevented, service availability, the false intervention rate, and recovery speed, as well as predictive accuracy.

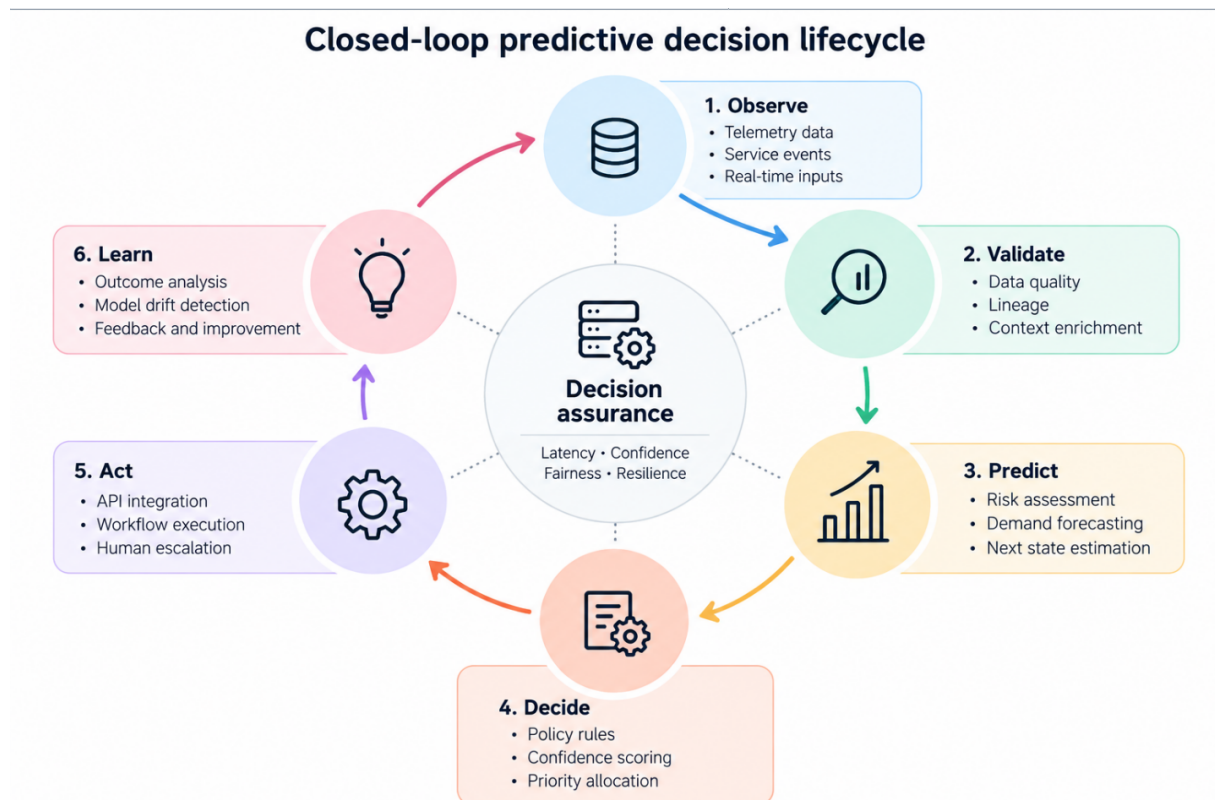


Figure 2. Closed-loop predictive decision lifecycle with decision assurance at the center.

7 Saudi digital service platforms: Strategic fit and use cases

Saudi Arabia is well-suited to the implementation of IoT-driven predictive decision-making since digital transformation is progressing hand in hand with improved connectivity, the modernization of its infrastructure, and the development of new service platforms. Saudi research has connected IoT, cloud capabilities, and modern communications to the objectives of Vision 2030 [2], and evidence from Saudi companies shows that real-time analytics can enhance decision-making processes when connected data are combined with operational activities [1]. The key opportunity is thus to make digital services anticipatory rather than just connected.

Public digital services may use application, network and user-journey telemetry to predict demand surges, anticipate likely degradation and increase their resources before service quality declines. The key aim is to ensure the continued availability of a service that is facing the public, and therefore technical measures should be combined with completion rate, wait time and the importance of the transaction. Mobility platforms can also use sensor and location data to forecast demand on a zone-by-zone basis, predict congestion and improve the positioning of their fleets. Research into smart cities indicates that connected transport and IoT infrastructure can enable coordinated mobility [19,29], while studies on sustainable transport stress the need to consider availability, cost, security and privacy when choosing IoT services [11].

Utilities and smart facilities can use prediction for demand, for the condition of assets and for energy management. Research based on IoT shows the value of connected data for operational optimisation [17], and studies involving digital twins show how continuous representations of the state allow failure-risk and remaining-life to be estimated [26]. In the operating environments of Saudi Arabia, predictive models must take into account weather conditions, the age of the assets, their level of

usage and their maintenance history, while edge processing should ensure a quick local response for protective actions.

Since incorrect actions in healthcare and emergency services could have more serious consequences, greater assurance is needed here. Security, privacy and the reliability of decisions are thus at the heart of connected-health architectures [4,7,30]. Platforms must differentiate between advisory predictions and autonomous interventions and should impose higher confidence levels and more stringent human review requirements as the consequences become greater.

The situation in Saudi Arabia is particularly noteworthy with regard to large events and pilgrimages. Studies on Hajj and Umrah show how IoT technology is being used to enhance connected services and enable situational awareness [3]. Predictive capabilities could be used to estimate crowd pressure, transport demand, queue formation or resource requirements. It is necessary to have distributed resilience since decisions might have to be made even when network conditions are variable or there is an excessive load on the central system.

In every situation, action should start with a specific, measurable decision rather than with a broad intention to make use of artificial intelligence. The organisation must first determine the intervention needed, the timing requirement, the minimum amount of data required, the acceptable trade-off between error and the course of action to be taken before it chooses a model. Adopting this approach makes it easier to justify the technical decisions and connects the investment in analytics with the actual service outcomes.

8 Governance, cybersecurity, privacy and reliability

Although predictive platforms enhance an organization's capabilities they also increase the level of risk since connected sensors widen the attack surface, data streaming platforms collect sensitive operational information and automated decisions can spread errors more quickly than manual ones. Governance should therefore be incorporated into the architecture. According to research into Reliable-IoT, security and latency are identified as essential requirements not as optional improvements [18], and studies on IoT security highlight the need to systematically assess the controls, the privacy requirements and the effectiveness of intrusion detection [4,15,30].

A necessary first step is to ensure trustworthy identity; devices, gateways, services and models must have verifiable identities and be given controlled privileges. If telemetry is compromised it can corrupt a prediction even without the model itself being attacked. The second requirement is data minimisation; edge processing and federated methods can reduce the amount of raw data that is moved while still allowing learning or inference to take place [27]. This is advantageous in cases where platforms combine personal, location or operationally sensitive data. The third requirement is model accountability; each model in production should have an owner, a approved purpose, a version, a validation record and a retirement condition.

Decision governance should also specify the different levels of autonomy. Automating actions of a low risk—such as cache expansion or noncritical notifications—can be done when confidence is high. For actions of medium risk, policy checks or confirmation from a supervisor are required. With regard to high-risk actions, human authority should be retained unless a formally validated safety case justifies automation. This approach of gradual escalation stops the organisation from mistaking technical confidence for institutional authority.

Reliability goes beyond just ensuring uptime; a reliable predictive service should still be useful even if some sensors fail, data arrive late, or the model becomes uncertain. Redundant signals, fallback rules, safe defaults, and observability should therefore be tested. Security and resilience exercises must also involve manipulated inputs, loss of connectivity, failure of the model or service, and sudden shifts in the distribution. The issue from an operational point of view is not whether failure can be eliminated, but rather whether the platform fails in a controlled and recoverable way.

9 Synthesis: A decision-centric implementation framework

The literature all supports an interpretation of IoT value that is focused on decision-making. Connected infrastructure generates evidence; analytics turns this evidence into estimates; governance decides which of the estimates can affect operations; and then the workflows turn the approved decisions into service outcomes. The capability chain thus formed is observe, validate, predict, decide, act and learn.

To observe it is necessary to have telemetry coverage and health; in order to validate it data quality, lineage, semantics and context are required; for prediction an operationally meaningful horizon and calibrated uncertainty are needed; for decision-making policies are required which combine the model output with business constraints; and to act it is necessary to integrate with APIs, workflows and accountable humans; for learning outcome capture is required so that both the models and the decision rules can improve.

Table 1 shows how the evidence examined relates to these capabilities. The studies involving smart cities, reliable IoT and industrial applications place particular emphasis on interoperability, connectivity, security and operational integration [18–21]. Research into decision-making indicates that it is necessary for the criteria and priorities to be explicit when alternatives are being compared [2,11]. The studies on digital twins and maintenance highlight the importance of feedback between the observed state, the analytical models and the operational processes [12,26]. Thus, prediction accuracy is only one element of service value.

Saudi organisations are said to follow a four-stage pathway towards maturity. In the first stage, observable services are set up and they produce consistent telemetry and outcomes. In the second stage, predictive models are introduced but actions are still mostly carried out by people. In the third stage, assisted decision-making is offered, with recommendations being prioritised and operator responses being recorded. In the fourth stage, bounded autonomy is allowed for certain actions within specified confidence, risk and rollback limits. Advance along the pathway should be based on proven reliability.

Table 2 sets out the deployment priorities in relation to this maturity pathway. When allocating time within end-to-end latency budgets it is necessary to distribute it among sensing, transmission, feature calculation, inference, policy evaluation and execution; equal importance should be given to data-freshness and recovery budgets since a fast model cannot make up for a slow or stale pipeline.

Model monitoring ought to connect statistical performance with service outcomes. Although precision, recall and forecasting error are still necessary, organisations should also monitor avoided downtime, reduced wait times, prevented escalations, resource savings or service completions. In the case where the model metrics improve but the operational outcomes do not, the decision policy or workflow may be the limiting factor.

Above all, the capabilities of the platform should be reusable. By providing shared services—which cover secure device onboarding, event management, feature calculation, model deployment, decision logging and observability—many different applications can be supported, while the sector-specific applications continue to keep their own data and autonomy rules. It is because of this that scale can be achieved without assuming that the various service sectors in Saudi Arabia have the same level of risk tolerance.

Table 1. Evidence synthesis across the predictive decision lifecycle.

Capability	Primary technical concern	Evidence synthesis	Operational implication
Observe	Coverage, identity, timestamps, connectivity	Reliable IoT and Saudi IoT studies emphasize dependable sensing and communications [2,3,18].	Monitor telemetry health and define data-freshness targets.
Validate	Quality, lineage, semantics, context	Smart-city studies stress interoperability and integrated data structures [19,20,29].	Apply schema control, deduplication, contextual joins and replay.
Predict	Horizon, uncertainty, drift, computational placement	Industrial automation, AIoT and digital-twin studies show value from predictive models over connected data [21,26,28].	Select models by decision horizon, not algorithm novelty.
Decide	Thresholds, trade-offs, autonomy, explainability	IoT decision studies show that reliability, cost, usability and security can require explicit multi-criteria evaluation [2,11].	Document policy thresholds and human escalation rules.
Act	Workflow integration, rollback, accountability	Maintenance digitalization shows that analytical capability must be integrated with operational processes [12].	Connect predictions to APIs, workflows and named owners.
Learn	Outcome capture, drift, policy review	Real-time IoT decision research links analytics to organizational performance and continuous improvement [1].	Monitor both model metrics and service outcomes.

Table 2. Design priorities for operational deployment.

Design priority	Recommended control	Suggested KPI
Latency	Allocate an end-to-end latency budget across sensing, processing, inference and action.	P95 decision completion time; percentage within intervention window
Data quality	Track freshness, completeness, duplication and device health before inference.	Valid-event ratio; stale-event rate; missing-feature rate
Model assurance	Version models, calibrate confidence, monitor drift and preserve rollback.	Calibration error; drift alerts; rollback time
Decision governance	Define autonomy tiers and escalation requirements by consequence.	Human override rate; false intervention rate; policy exceptions
Security and privacy	Use strong identity, least privilege, encrypted transport and minimized data movement.	Unauthorized-access events; sensitive-data transfer volume
Business value	Link every prediction to an operational outcome and accountable owner.	Avoided downtime; completion rate; cost or capacity benefit

10 Research agenda

There are still some research gaps. First, further evidence is required regarding the link between prediction latency and the actual service value. Although a number of studies give information on model accuracy, they do not measure how the delay affects the usefulness of a decision. Future research carried out in Saudi Arabia should provide data on end-to-end decision time and compare it with the intervention window. Second, model uncertainty should be treated as an operational variable. While confidence estimates can aid dynamic autonomy, the way in which organizations set these thresholds has not been adequately explored.

Third, there should be more focus on cross-platform interoperability. Since smart-city and enterprise ecosystems usually include data sources that are separately governed, predictive models might need information from a number of different platforms. Research should look into semantic interoperability, event contracts and common feature definitions within the various service areas of Saudi Arabia. Fourth, privacy-preserving analytics should progress from being just a matter of discussion to actual implementations. Although federated learning provides a valuable approach, its communication costs, the issue of model heterogeneity and the governance consequences need to be evaluated in a way that is specific to each context [27].

Next, further research needs to look into how humans interact with predictive recommendations. Even if a model is technically sound, it may fail if the operators do not understand the reasons behind it or if the alerts occur at an inappropriate stage in the workflow. The studies should examine trust calibration, the way in which users override recommendations, and the impact of different explanation designs on the quality of decisions. Lastly, sustainability assessments should also take into account the computational and network costs associated with the analytics itself. Although AIoT can lead to improved efficiency [28], continuous sensing, data transmission and model inference all require resources. Research must therefore assess the net operational benefit rather than assuming that digitalisation is inherently sustainable.

11 Conclusion

Predictive analytics enabled by IoT can enhance real-time decision-making in Saudi digital service platforms only when it is adopted as a governed operational system rather than being used as a standalone modelling exercise. Evidence examined between 2020 and 2025 indicates that successful implementations involve the integration of reliable sensing, processing at the edge or in the cloud, contextual data engineering, predictive models, clearly defined decision policies, workflow execution and continuous feedback. Reliability, cybersecurity, privacy and model accountability must all be considered as essential requirements in the design process.

The most practical architectural approach for Saudi Arabia is a federated edge-cloud model which allows for a variety of service domains while at the same time maintaining local responsiveness and having central governance. Implementation should start with clear decisions and measurable results, and then move on to observability, followed by prediction, assisted decision-making and bounded autonomy. The suggested architecture and its lifecycle offer a foundation for this kind of progression. Their main implication is simple: predictive intelligence only generates value when a service is able to act on it safely, swiftly and in an accountable manner. Future research must therefore assess not just whether the model predicts accurately, but also whether the entire decision process leads to improved service quality, resilience, sustainability and user outcomes within the Saudi digital ecosystem.

An additional consequence is that the service architecture must retain the provenance of sensor observations all the way through to the final action, enabling auditors and operators to reconstruct not only the model's predictions but also the particular contextual data and the policy state that influenced the response. Operational teams should also make the distinction between model refresh and model replacement, since frequent retraining can enhance adaptation at the same time as it leads to instability if the validation gates, the champion-challenger comparisons, and the rollback procedures are inadequate. In the case of cross-agency platforms, common technical components should be standardised while decision-making authority should remain specific to each domain, as the presence of a shared event fabric does not mean that health, mobility, utilities and commercial services should apply the same autonomy thresholds. This well-structured operating model ensures that predictive analytics remain in line with the purpose of the service while at the same time improving transparency, resilience, reuse and accountability.

References

- [1] Malik S. Data-driven decision-making: Leveraging the IoT for real-time sustainability in organizational behavior. *Sustainability*. 2024;16:6302. <https://doi.org/10.3390/su16156302>
- [2] Alojaiman B. A multi-criteria decision-making process for the selection of an efficient and reliable IoT application. *Processes*. 2023;11:1313. <https://doi.org/10.3390/pr11051313>
- [3] Shambour MK, Gutub A. Progress of IoT research technologies and applications serving Hajj and Umrah. *Arabian Journal for Science and Engineering*. 2022;47:1253–1273. <https://doi.org/10.1007/s13369-021-05838-7>
- [4] Ansari MTJ, Al-Zahrani FA, Pandey D, Agrawal A. A fuzzy TOPSIS based analysis toward selection of effective security requirements engineering approach for trustworthy healthcare software development. *BMC Medical Informatics and Decision Making*. 2020;20:236. <https://doi.org/10.1186/s12911-020-01209-8>
- [5] Abdel-Basset M, Gamal A, Son LH, Smarandache F. A bipolar neutrosophic multi criteria decision making framework for professional selection. *Applied Sciences*. 2020;10:1202. <https://doi.org/10.3390/app10041202>
- [6] Yadav S, Garg D, Luthra S. Development of IoT based data-driven agriculture supply chain performance measurement framework. *Journal of Enterprise Information Management*. 2021;34:292–327. <https://doi.org/10.1108/JEIM-11-2019-0369>
- [7] Qahtan S, Sharif KY, Zaidan AA, Alsattar HA, Albahri OS, Zaidan BB, et al. Novel multi security and privacy benchmarking framework for blockchain-based IoT healthcare Industry 4.0 systems. *IEEE Transactions on Industrial Informatics*. 2022;18:6415–6423. <https://doi.org/10.1109/TII.2022.3143619>
- [8] Gardas BB, Heidari A, Navimipour NJ, Unal M. A fuzzy-based method for objects selection in blockchain-enabled edge-IoT platforms using a hybrid multi-criteria decision-making model. *Applied Sciences*. 2022;12:8906. <https://doi.org/10.3390/app12178906>
- [9] Mahmoudi A, Deng X, Javed SA, Zhang N. Sustainable supplier selection in megaprojects: Grey ordinal priority approach. *Business Strategy and the Environment*. 2021;30:318–339. <https://doi.org/10.1002/bse.2623>
- [10] Ijamaru GK, Ang LM, Seng KP. Transformation from IoT to IoV for waste management in smart cities. *Journal of Network and Computer Applications*. 2022;204:103393. <https://doi.org/10.1016/j.jnca.2022.103393>
- [11] Krishankumar R, Ecer F. Selection of IoT service provider for sustainable transport using q-rung orthopair fuzzy CRADIS and unknown weights. *Applied Soft Computing*. 2023;132:109870. <https://doi.org/10.1016/j.asoc.2022.109870>
- [12] Ahmed U, Carpitella S, Certa A, Izquierdo J. A feasible framework for maintenance digitalization. *Processes*. 2023;11:558. <https://doi.org/10.3390/pr11020558>
- [13] Deveci M, Gokasar I, Pamucar D, Chen Y, Coffman DM. Sustainable e-scooter parking operation in urban areas using fuzzy Dombi based RAFSI model. *Sustainable Cities and Society*. 2023;91:104426. <https://doi.org/10.1016/j.scs.2023.104426>

-
- [14] Alyami H, Nadeem M, Alharbi A, Alosaimi W, Ansari MTJ, Pandey D, et al. The evaluation of software security through quantum computing techniques: A durability perspective. *Applied Sciences*. 2021;11:11784. <https://doi.org/10.3390/app112411784>
- [15] Alyami H, Ansari MTJ, Alharbi A, Alosaimi W, Alshammari M, Pandey D, et al. Effectiveness evaluation of different IDSs using integrated fuzzy MCDM model. *Electronics*. 2022;11:859. <https://doi.org/10.3390/electronics11060859>
- [16] Khan AI, Alsolami F, Alqurashi F, Abushark YB, Sarker IH. Novel energy management scheme in IoT enabled smart irrigation system using optimized intelligence methods. *Engineering Applications of Artificial Intelligence*. 2022;114:104996. <https://doi.org/10.1016/j.engappai.2022.104996>
- [17] Alassery F, Alzahrani A, Khan AI, Irshad K, Kshirsagar SR. An artificial intelligence-based solar radiation prophesy model for green energy utilization in energy management system. *Sustainable Energy Technologies and Assessments*. 2022;52:102060. <https://doi.org/10.1016/j.seta.2022.102060>
- [18] Khan MZ, Alhazmi OH, Javed MA, Ghandorh H, Aloufi KS. Reliable Internet of Things: Challenges and future trends. *Electronics*. 2021;10:2377. <https://doi.org/10.3390/electronics10192377>
- [19] Alloui H, Mourdi Y. Exploring the full potentials of IoT for better financial growth and stability: A comprehensive survey. *Sensors*. 2023;23:8015. <https://doi.org/10.3390/s23198015>
- [20] Bellini P, Nesi P, Pantaleo G. IoT-enabled smart cities: A review of concepts, frameworks and key technologies. *Applied Sciences*. 2022;12:1607. <https://doi.org/10.3390/app12031607>
- [21] Al Shahrani AM, Alomar MA, Alqahtani KN, Basingab MS, Sharma B, Rizwan A. Machine learning-enabled smart industrial automation systems using Internet of Things. *Sensors*. 2023;23:324. <https://doi.org/10.3390/s23010324>
- [22] Farooq MS, Riaz S, Abid A, Umer T, Zikria YB. Role of IoT technology in agriculture: A systematic literature review. *Electronics*. 2020;9:319. <https://doi.org/10.3390/electronics9020319>
- [23] Khan Y, Su'ud MBM, Alam MM, Ahmad SF, Ahmad AYAB, Khan N. Application of Internet of Things (IoT) in sustainable supply chain management. *Sustainability*. 2023;15:694. <https://doi.org/10.3390/su15010694>
- [24] Bibri SE. The eco-city and its core environmental dimension of sustainability: Green energy technologies and their integration with data-driven smart solutions. *Energy Informatics*. 2020;3:4. <https://doi.org/10.1186/s42162-020-00107-7>
- [25] Singh S, Ru J. Goals of sustainable infrastructure, industry, and innovation: A review and future agenda for research. *Environmental Science and Pollution Research*. 2023;30:28446–28458. <https://doi.org/10.1007/s11356-023-25281-5>
- [26] van Dinter R, Tekinerdogan B, Catal C. Predictive maintenance using digital twins: A systematic literature review. *Information and Software Technology*. 2022;151:107008. <https://doi.org/10.1016/j.infsof.2022.107008>
- [27] Nguyen DC, Ding M, Pathirana PN, Seneviratne A, Li J, Poor HV. Federated learning for Internet of Things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*. 2021;23:1622–1658. <https://doi.org/10.1109/COMST.2021.3075439>
- [28] Matin A, Islam MR, Wang X, Huo H, Xu G. AIoT for sustainable manufacturing: Overview, challenges, and opportunities. *Internet of Things*. 2023;24:100901. <https://doi.org/10.1016/j.iot.2023.100901>
- [29] Rai HM, Atik-Ur-Rehman, Pal A, et al. Use of Internet of Things in the context of execution of smart city applications: A review. *Discover Internet of Things*. 2023;3:8. <https://doi.org/10.1007/s43926-023-00037-2>
- [30] Ansari MTJ, Baz A, Alhakami H, Alhakami W, Kumar R, Khan RA. P-STORE: Extension of STORE methodology to elicit privacy requirements. *Arabian Journal for Science and Engineering*. 2021;46:8287–8310. <https://doi.org/10.1007/s13369-021-05476-z>