© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722144

Sustainability, Energy Optimization and Mechanical System Downtime Reduction through Advanced HVAC Operation and Control

ABUBAKER MOHAMMED
ORCID: 0009-0000-5101-809X

Abstract-* Heating, ventilation and air-conditioning systems sit at the intersection of building decarbonisation, occupant wellbeing and asset reliability. Yet energy optimisation programmes often treat control efficiency and mechanical downtime as separate problems, even though fouling, sensor drift, valve leakage, unstable sequencing and degraded heat transfer simultaneously increase consumption and failure exposure. This review synthesises research published from 2020 to 2025 on advanced HVAC operation, predictive control, fault detection, digital twins and maintenance analytics. Its aim is to explain how an integrated operating model can minimise energy and carbon intensity while preserving comfort, equipment health and service continuity. A structured evidence review was undertaken across control, energy and building-engineering literature, with studies assessed according to operational scope, data requirements, verification quality and relevance to real facilities. The evidence indicates that forecasting, model-predictive control, reinforcement learning and occupancy-responsive ventilation can reduce avoidable load, but savings are sustained only when control actions are constrained by equipment limits and supported by reliable sensors. Machine-learning fault detection and predictive maintenance improve early recognition of degradation, yet their practical value depends on diagnostic precision, maintenance capacity and post-intervention verification. The paper proposes a unified architecture linking sensing, state estimation, optimisation, supervisory control, fault diagnosis and work management. The central conclusion is that advanced HVAC management should be evaluated as a resilience capability rather than a narrow energy-saving intervention. The strongest programmes jointly govern energy, comfort, emissions, mean time between failures, repair time and unplanned downtime, thereby converting intelligent control into durable operational performance.

***Keywords:* HVAC, Energy Optimisation, Sustainability, Predictive Control, Fault Detection, Predictive Maintenance, Mechanical Downtime, Digital Twin**

I. INTRODUCTION

Buildings are expected to deliver safe indoor conditions continuously while reducing energy use, operating expenditure and greenhouse-gas emissions. HVAC systems are central to this challenge because they combine large electrical and thermal loads with mechanically complex assets, distributed sensors and changing occupancy. Recent reviews show that artificial intelligence, advanced analytics and supervisory control can materially improve HVAC performance, but they also reveal an imbalance in the literature: control optimisation receives far more attention than maintenance and long-term reliability [1]. This imbalance matters because an apparently efficient control sequence cannot compensate indefinitely for leaking valves, clogged filters, refrigerant undercharge, failed dampers, drifting sensors or worn bearings. Such conditions alter the physical response of the plant, invalidate control assumptions and eventually produce service interruption.

The operational problem is therefore not simply to minimise instantaneous energy demand. It is to maintain acceptable temperature, humidity, ventilation and pressure while operating equipment within safe and efficient regions over its life cycle. That objective requires coordination across building management systems, maintenance teams, energy managers and occupants. Smart-building research describes this coordination as a data-rich control problem in which weather, occupancy, tariffs, equipment states and thermal dynamics must be interpreted together [3,5]. However, facilities commonly separate energy dashboards from computerised maintenance systems. Energy analysts identify abnormal consumption without access to work-order history, while technicians repair equipment without seeing how degradation changed energy intensity or comfort risk. The resulting fragmentation delays diagnosis.

Advanced operation and control can close this gap when it is designed around physical accountability. Model-predictive control uses forecasts and system models to calculate future control actions subject to constraints. Reinforcement learning can discover policies in uncertain and non-linear environments, while occupancy prediction and adaptive comfort models can reduce unnecessary conditioning [4,7,10]. Digital twins and virtual meters extend observability where direct measurements are incomplete [23,28]. In parallel, machine-learning fault detection can distinguish normal variation from persistent degradation, and predictive maintenance can convert diagnostic evidence into planned intervention [6,9]. Their combined value lies in a closed loop: detect plant condition, predict demand and failure risk, choose a safe action, verify the response and update control and maintenance decisions.

The sustainability implications are broader than reduced utility consumption. Avoided energy lowers operational emissions; stable control reduces peak demand and unnecessary cycling; early fault correction preserves coefficient of performance; and planned maintenance avoids emergency replacement, occupant disruption and premature disposal. Reliability is thus a sustainability variable. A chiller that repeatedly trips, an air-handling unit that operates against excessive pressure or a pump that short-cycles may consume more energy before failing, while emergency repair can create further material, labour and service impacts. Aggressive energy reduction can also increase downtime when set-points, starts or pressure resets violate mechanical constraints. The required balance is multi-objective.

This review addresses that problem by examining how recent research connects energy optimisation with mechanical-system continuity. It focuses on central plants, air-handling systems, terminal units, pumps, fans, heat exchangers, heat pumps and associated sensing and control infrastructure. The paper evaluates how forecasting, optimisation, diagnostics and maintenance governance interact in operating contexts and affect measured outcomes. The analysis develops an integrated framework and identifies evidence gaps in transferability, fault-aware control and avoided-downtime measurement.

II. AIM, OBJECTIVES AND REVIEW QUESTIONS

The aim of this review is to establish an integrated understanding of how advanced HVAC operation and control can improve environmental performance while reducing mechanical-system downtime. Four objectives guide the analysis. First, the review identifies control methods that reduce energy use without transferring risk to comfort, indoor air quality or equipment condition. Second, it evaluates how fault detection, diagnostic analytics and predictive maintenance contribute to reliability. Third, it examines the information architecture needed to connect building automation with maintenance decision-making. Fourth, it proposes performance measures and governance principles for durable implementation.

The review is organised around three questions. RQ1 asks which combinations of forecasting, supervisory control and optimisation provide the strongest balance between energy, comfort and mechanical constraints. RQ2 asks how operational data can detect degradation early enough to prevent inefficient operation and unplanned outages. RQ3 asks what organisational and technical conditions allow these methods to remain trustworthy after deployment. These questions integrate control and maintenance because shared observations can support both energy optimisation and health assessment.

The contribution is a synthesis centred on operating capability rather than algorithm ranking. Recent literature compares deep reinforcement learning, neural networks, random forests, support-vector machines and other techniques [1,4-6]. Such comparisons are useful, but practical selection depends on system observability, fault prevalence, available engineering knowledge, computing limits and tolerance for unsafe actions. This paper therefore evaluates methods according to whether they produce verifiable decisions, remain within physical boundaries, support maintenance workflows and improve measured outcomes over time.

III. REVIEW METHODOLOGY

A structured evidence-review method was used to synthesise studies published between 2020 and 2025. The search scope covered building energy management, HVAC supervisory control, model-predictive control, reinforcement learning, occupancy-responsive operation, fault detection and diagnosis, predictive maintenance, digital twins and virtual sensing. Priority was given to peer-reviewed engineering studies describing control architectures, diagnostics, operational datasets or field implementations. Studies concerned only with generic building-energy prediction were included only where their methods directly informed HVAC operation or equipment-health assessment.

Screening proceeded in three stages. Titles and abstracts were first examined for relevance to HVAC energy performance, mechanical reliability or integrated building operation. Full texts were then assessed for methodological transparency, including description of input variables, control or diagnostic outputs, baselines, constraints and evaluation periods. The synthesis classified evidence into six domains: sensing and data quality; forecasting; supervisory control; comfort and air quality; fault diagnosis; and maintenance planning. This classification reflects facility decisions rather than algorithm taxonomies.

The review used a comparative analytical approach. For control studies, attention was given to baseline selection, comfort constraints, weather and occupancy representation, action frequency, equipment cycling and the difference between simulated and field performance. For maintenance studies, the analysis considered fault coverage, class imbalance, diagnostic delay, false alarms, interpretability and whether detected faults resulted in verified corrective work. Evidence was also examined for transferability across buildings and climates. A method that performs well on one plant may fail when sensor naming, sequence logic, equipment capacity or operator practice changes.

The synthesis did not calculate a pooled effect size because reported outcomes were heterogeneous. Some studies measured percentage energy savings, others electricity cost, thermal discomfort, carbon reduction, fault-classification accuracy or operational stability. Instead, results were interpreted through mechanisms. Savings were considered more persuasive when mechanisms, constraints and persistence were demonstrated. Diagnostic accuracy was not equated with avoided downtime without maintenance action and post-repair verification.

To strengthen practical relevance, the evidence was mapped to an integrated architecture and a staged maturity model. The architecture defines information and control flows from instrumentation to verified action. The maturity model links technical functions to performance indicators such as energy use intensity, plant efficiency, comfort violations, mean time between failures, mean time to repair and unplanned downtime. This supports comparison while keeping the focus on operational outcomes.

IV. SUSTAINABILITY AND ENERGY-PERFORMANCE FOUNDATIONS

HVAC sustainability begins with reducing required load before optimising plant response. Mixed-mode operation, shading, envelope performance, ventilation strategy and occupancy schedules determine how much heating or cooling the system must provide. Reviews of natural-ventilation and HVAC coordination show that hybrid operation can reduce mechanical conditioning when outdoor conditions are suitable, but it requires careful control of humidity, air quality, noise and occupant behaviour [2]. Occupancy and window-opening prediction similarly allow systems to avoid conditioning empty zones and to anticipate disturbances that would otherwise trigger excessive reheating or recooling [16,19]. These measures reduce energy at source rather than merely improving conversion efficiency.

After load reduction, supervisory control must translate demand into efficient equipment operation. Traditional fixed schedules and proportional-integral loops are dependable but often respond to local variables without considering future weather, thermal storage, tariffs or interactions among plant components. Predictive strategies use forecasts to move operation away from reactive correction. Neural-network model-predictive control, for example, can estimate future zone conditions and select control actions that minimise energy while maintaining constraints [26]. The value of this approach is not only lower consumption. By avoiding large control errors, rapid valve movement and unnecessary starts, predictive control can reduce mechanical stress.

Energy performance should be assessed at component and system levels. A chiller may operate efficiently while the distribution system wastes energy through excessive differential pressure. A supply fan may meet airflow requirements while terminal boxes remain nearly closed because static-pressure set-points are too high. Reset strategies for chilled-water temperature, condenser-water temperature, duct pressure and ventilation can therefore produce substantial savings when they are based on actual demand. Swarm-intelligence optimisation has been applied to complex HVAC set-point selection, illustrating the usefulness of search methods where relationships are non-linear and constraints interact [8]. However, optimisation must include minimum flow, freeze protection, humidity limits, motor loading and equipment sequencing rules.

Comfort and indoor air quality cannot be treated as penalties added after energy minimisation. Adaptive comfort research shows that acceptable conditions vary with clothing, activity, exposure and expectation [10,17]. Learning-based carbon-dioxide prediction supports demand-controlled ventilation by anticipating occupancy-related pollutant build-up rather than reacting after thresholds are exceeded [20]. These methods permit a more precise service level, but they also depend on sensor placement and calibration. A biased carbon-dioxide sensor can reduce ventilation below safe requirements, while a drifting temperature sensor can create simultaneous heating and cooling. Data assurance is therefore part of sustainability: poor measurements produce both energy waste and service risk.

The most credible sustainability strategy combines energy, carbon, comfort and asset-health indicators. Large-scale evaluation of data-driven predictive control suggests that greenhouse-gas reduction depends on climate, grid carbon intensity, building type and baseline operation [22]. A control intervention that shifts load may reduce cost but not emissions if it moves consumption into a more carbon-intensive period. Similarly, an efficiency gain that increases cycling or operates equipment near unstable limits may shorten life. Sustainable operation requires transparent trade-offs and verification against the actual plant response.

V. ADVANCED HVAC OPERATION AND CONTROL

Advanced HVAC control is most effective when it separates decision layers. Fast local loops should continue to regulate pressure, temperature and flow with deterministic logic. A slower supervisory layer can then optimise schedules, set-points and equipment staging using forecasts and system models. This hierarchy protects stability while concentrating intelligence where uncertainty can be managed. Figure 1 follows this principle: validated sensing feeds state estimation; forecasting describes demand; optimisation proposes actions; safety checks constraints; supervisory control issues bounded commands; and verification compares predicted with realised performance.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722144

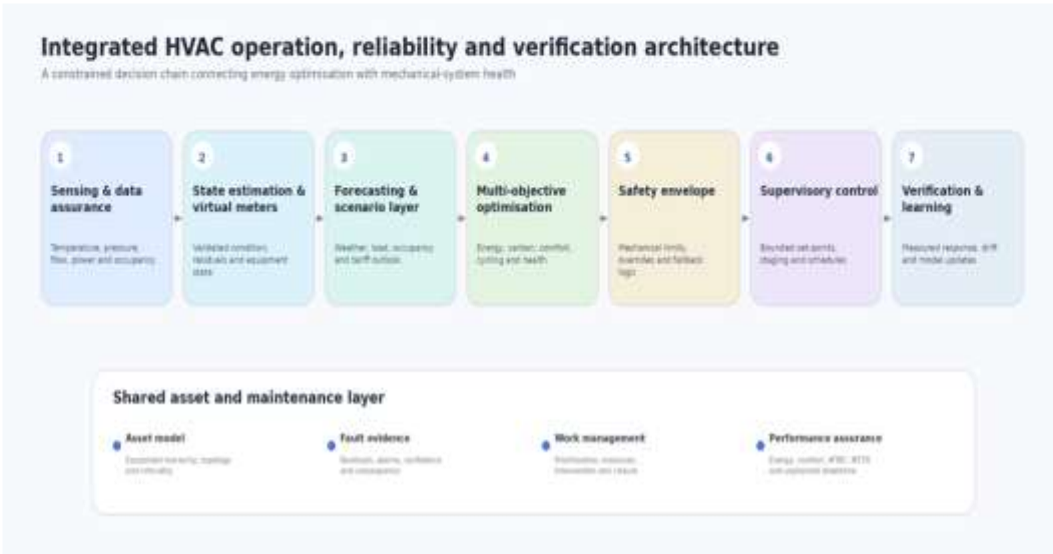

Figure 1. Integrated architecture for energy-efficient, fault-aware HVAC operation and verified maintenance.

Model-predictive control is attractive because constraints are explicit. It calculates a sequence of future actions over a receding horizon and applies only the first action before updating with new measurements. It suits precooling, storage, demand limiting and coordinated plant operation. The principal difficulty is model maintenance. Building thermal behaviour changes with occupancy, season, envelope alterations and equipment degradation. Grey-box virtual meters and inverse models can improve observability without requiring a sensor at every point [28]. Digital-twin co-simulation can further connect physical models with live data, but the model must be calibrated, versioned and checked for drift [23]. A twin that omits valve authority, fouling or actual sequence logic can produce confident but incorrect recommendations.

Reinforcement learning offers another path where detailed models are unavailable. Reviews show extensive experimentation with value-based, policy-gradient and actor-critic methods for building control [4,24]. Deep reinforcement learning has reduced heating energy while maintaining indoor temperature in simulation and field-oriented studies [11,14]. Multi-zone research demonstrates the value of coordinated policies where actions in one zone affect shared equipment and neighbouring conditions [15]. Multi-agent methods can distribute decisions across zones or components, allowing local responsiveness with system-level coordination [13]. Yet learning-based control introduces risks: unsafe exploration, sensitivity to reward design, unstable behaviour under unseen weather and difficulty explaining actions.

Hybrid control reduces these risks: physics-based rules define hard limits while learning algorithms optimise within an approved action space. Change-detection methods can recognise when a building becomes non-stationary and trigger adaptation rather than allowing a policy to operate on obsolete assumptions [25]. Predictive deep reinforcement learning can incorporate future load information to improve decision quality [27]. Deep Q-network studies similarly show the potential for balancing competing HVAC subsystems without a fully specified model [21]. The engineering question is how the system handles uncertainty, limits and failure.

Occupancy-responsive operation creates another important control layer. Sensors, access data, schedules and connected thermostats can estimate whether a zone is occupied and how demand is likely to evolve. Occupancy-based reviews report significant potential but also privacy, sensing and prediction challenges [19]. Control should distinguish presence from service need: a lightly occupied room may still require ventilation, humidity control or equipment protection. Occupancy predictions should not disable critical conditioning without fallback logic. Confidence thresholds, conservative minimum ventilation and graceful reversion to standard schedules are necessary.

Air-handling units illustrate the benefit of integrated control. Their energy use depends on supply temperature, airflow, outdoor-air fraction, coil condition, fan efficiency and zone demand. Recurrent neural networks can capture time-dependent behaviour, while reinforcement learning can select coordinated commands [18]. However, an algorithm may interpret a stuck damper as demand or compensate for a leaking valve by changing supply temperature. The controller must therefore consume diagnostic states, not only raw measurements. Fault-aware optimisation can exclude unreliable sensors, adjust confidence in model predictions and prevent the control system from masking degradation.

Cost and demand management add further objectives. Reinforcement-learning control has been used to reduce both energy cost and demand charges, showing that economic optimisation can be incorporated into building operation [30]. Such strategies should be evaluated against carbon and equipment impacts. Pre-cooling may lower peak demand while increasing total energy or compressor runtime; demand limiting may also reduce redundancy. A balanced objective function should include energy, emissions, comfort, cycling, start penalties and health indicators. Table 1 compares the principal control and maintenance strategies according to their energy and downtime mechanisms.

Table 1. Comparative mechanisms of advanced HVAC control and reliability strategies.

Strategy	Primary data and action	Energy mechanism	Downtime mechanism	Key implementation risk
Model-predictive control	Forecasts, thermal model and constrained set-point sequence	Anticipates load, resets plant variables and shifts demand	Avoids unstable correction and excessive cycling	Model drift, weak calibration or omitted constraints
Reinforcement learning	State observations, reward function and bounded actions	Learns sequential policies under non-linear conditions	Coordinates shared assets and adapts to changing demand	Unsafe exploration, opaque rewards and poor transferability
Occupancy / IAQ responsive control	Presence, schedules, CO2 and zone conditions	Conditions and ventilates according to verified service need	Reduces unnecessary runtime and limits overload	Privacy, biased sensors and insufficient fallback ventilation
Fault detection and diagnosis	Residuals, patterns, rules and equipment topology	Identifies inefficient operation before it becomes normalised	Provides earlier warning and probable cause	False alarms, sparse labels and combined faults
Predictive maintenance	Condition trends, history, criticality and work constraints	Restores heat transfer and efficiency before severe loss	Plans intervention before functional failure	Prediction without actionable lead time or resources

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722144

Strategy	Primary data and action	Energy mechanism	Downtime mechanism	Key implementation risk
Fault-aware supervisory control	Diagnostic confidence, redundancy and temporary constraints	Prevents optimisation from masking degraded performance	Maintains bounded service during degraded operation	Compensation becoming permanent or shifting stress elsewhere

Deployment should progress from advisory to supervisory and, where justified, closed-loop control. In advisory mode, the system proposes actions and records operator decisions. This exposes data problems and builds trust. Supervisory mode automates bounded set-point changes while retaining deterministic safeties. Closed-loop optimisation should be reserved for well-instrumented systems with tested fallback sequences, change control and continuous monitoring. Organisational maturity is as important as algorithm performance.

VI. DOWNTIME REDUCTION AND MECHANICAL RELIABILITY

Mechanical downtime usually develops through a chain of weak signals rather than a single sudden event. Heat-transfer degradation increases valve demand; rising pressure drop increases fan or pump power; bearing wear increases vibration and current; sensor drift causes control hunting; and refrigerant or water-side faults reduce capacity until trips occur. Conventional alarms often detect only the final threshold violation. Advanced fault detection seeks to identify residual patterns earlier by comparing measured behaviour with expected performance. Reviews of automated fault detection show that supervised, unsupervised and hybrid machine-learning methods can cover complex building systems, but evidence quality varies widely [6].

Supervised diagnosis requires labelled examples of each fault. Chiller research using generative adversarial networks demonstrates one way to augment sparse fault data and improve classification [12]. Yet synthetic examples may not reproduce combined faults, maintenance history and sensor noise. Unsupervised approaches instead learn normal behaviour and flag deviations. Time-series analysis has been used to detect building energy inefficiencies without exhaustive fault labels [29]. This is valuable, but anomaly detection does not identify root cause automatically. A high residual could represent equipment degradation, schedule change, unusual weather, sensor error or a legitimate operational event.

Diagnostic design should therefore follow a layered logic. The first layer validates sensor plausibility and communication quality. The second detects abnormal relationships, such as simultaneous heating and cooling, persistent valve saturation, excessive approach temperature or inconsistent flow and power. The third isolates probable causes using physics, topology and maintenance history. The fourth estimates consequence: energy loss, comfort risk, failure probability and affected zones. The fifth recommends action with evidence and confidence. This prevents an unprioritised stream of machine-generated alerts.

Predictive maintenance converts condition evidence into timing decisions. Machine-learning approaches in building facilities show potential to estimate degradation and prioritise work before functional failure [9]. Value depends on whether prediction changes action. A model that forecasts a fan failure but provides no lead time, component location or confidence may not improve maintenance planning. An interpretable trend in vibration, current and airflow can justify inspection during an available shutdown. Maintenance analytics should therefore report remaining opportunity, not only probability. Opportunity includes expected failure horizon, spare availability, access requirements, operational redundancy and the cost of delaying intervention.

Downtime reduction also requires fault-tolerant operation. When a component is degraded but not yet failed, the control system can reconfigure operation

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722144

to preserve service. Examples include redistributing load among chillers, reducing pressure demand, changing supply temperatures, isolating an unreliable sensor or shifting zones to available air-handling capacity. Such actions must be temporary and visible; otherwise control may conceal the fault, shift stress and remove the incentive for repair. A fault-aware controller should create a maintenance case, apply a bounded compensating strategy and track the cumulative cost of continued operation.

The feedback loop in Figure 2 links condition detection to verified recovery. Detection establishes evidence; diagnosis identifies mechanism; prioritisation evaluates risk and consequence; planning schedules resources; intervention restores the system; verification confirms that energy, capacity and stability returned to expected ranges; and learning updates models and thresholds. Verification is often missing. A closed work order does not prove that a valve was corrected or a replacement sensor calibrated. Post-repair analytics should compare performance before and after intervention under comparable conditions.

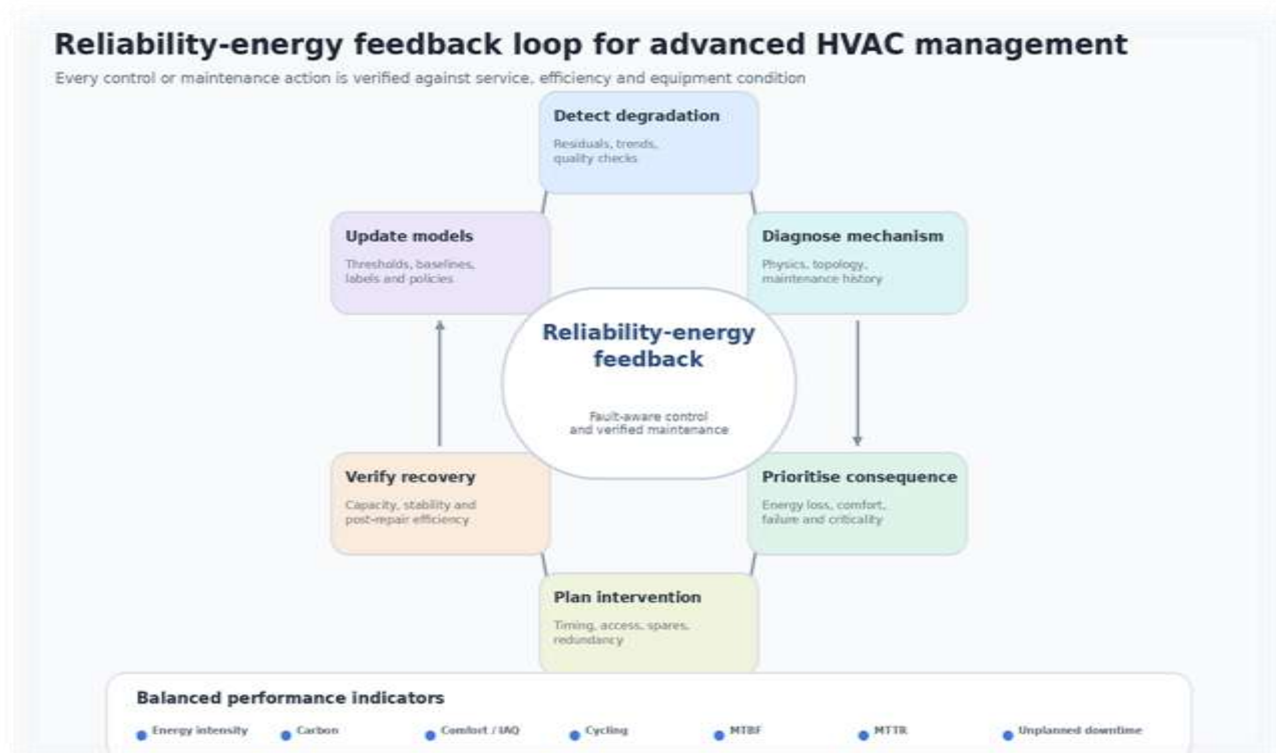

Figure 2. Reliability-energy feedback loop for fault-aware control and verified maintenance recovery.

Reliability indicators must be interpreted alongside energy data. Mean time between failures reflects recurrence, while mean time to repair reflects maintainability and response. Unplanned downtime captures service loss, but near-failures and degraded operation may have large energy costs without producing outages. The recommended indicator set therefore includes fault duration, diagnostic lead time, repeat-fault rate, alarm precision, avoided runtime in inefficient states and verified savings after repair. Table 2 presents a maturity model that connects these indicators to progressively stronger operating capabilities.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722144

Table 2. Maturity model for integrated HVAC energy and downtime management.

Maturity level	Operational capability	Control / maintenance integration	Core indicators	Decision gate
1. Instrumented	Reliable point naming, trend collection and asset hierarchy	Manual review links alarms with work history	Data completeness, sensor validity, schedule compliance	Proceed when critical measurements and sequences are trustworthy
2. Diagnosed	Rules, baselines, residuals and anomaly prioritisation	Energy exceptions create traceable maintenance cases	Alarm precision, fault duration, repeat-fault rate, energy residual	Proceed when alerts are actionable and false-alarm burden is controlled
3. Predictive	Load, occupancy and condition forecasts with advisory optimisation	Predicted demand and degradation inform planning windows	Forecast error, diagnostic lead time, avoided inefficient runtime	Proceed when models remain stable across seasons and operating modes
4. Supervisory	Bounded automatic resets, staging and fault-aware reconfiguration	Controls issue maintenance evidence and respect asset health states	Energy, carbon, comfort, cycling, MTBF, MTTR, downtime	Proceed when safeties, rollback and post-action verification are tested
5. Resilient	Portfolio learning, digital twins and verified closed-loop improvement	Interventions update baselines, models and control policies	Verified savings, avoided outages, service continuity, model drift	Maintain through governance, audits and continuous engineering review

Human expertise remains essential because technicians understand sounds, access limits, modifications and compromises absent from data. Advanced systems should structure this knowledge rather than replace it. Diagnostic explanations should show contributing measurements, expected behaviour and alternative causes. Operators should be able to confirm, reject or refine findings. These interactions create labelled evidence and reduce institutionalised automated error.

VII. INTEGRATED EVIDENCE SYNTHESIS

The literature supports a central proposition: energy optimisation and downtime reduction arise from the same disciplined loop. Both depend on trustworthy measurements, an expected-performance model, detection of deviation, constrained decision-making and verification. The difference is timescale: control responds over minutes or hours, while degradation management operates over days or months. An integrated platform allows these timescales to inform each other. A rising condenser approach temperature, for example, can alter chiller staging today while creating a cleaning recommendation for the next maintenance window.

The strongest architecture contains seven functions. First, instrumentation captures temperatures, pressures, flows, positions, electrical variables, occupancy and environmental conditions. Second, data-quality services detect missing values, bias, frozen points and naming inconsistencies. Third, state estimation combines measurements with physical relationships and virtual meters. Fourth, forecasting estimates loads, occupancy, tariffs and weather. Fifth, optimisation selects actions across energy, comfort, carbon and equipment-health objectives. Sixth, a safety and supervisory layer enforces mechanical limits, overrides and fallbacks. Seventh, fault and maintenance services preserve evidence, prioritise action and verify recovery. These functions should share an asset model linking data, control points and work orders to one equipment hierarchy.

The evidence also shows why apparently successful pilots fail to scale. Algorithms are frequently trained on clean research datasets, whereas operational systems contain sensor substitutions, manual overrides, inconsistent sampling and undocumented sequence changes. Models may deteriorate when season, occupancy or plant configuration changes. Transfer learning can reduce redevelopment effort, but it does not eliminate the need for local calibration. Governance must define who approves model changes, who can alter control bounds, how performance is reviewed and when the system reverts to a safe baseline.

A multi-objective scorecard is essential. Energy-only targets encourage actions that may compromise ventilation, comfort or asset life. Reliability-only targets can encourage excessive redundancy and conservative set-points. The scorecard includes weather-normalised energy, peak demand, carbon, efficiency, comfort and air-quality violations, cycling, alarm burden, diagnostic precision, maintenance lead time, failure interval, repair time and downtime. Financial measures should include avoided energy, avoided emergency work, deferred replacement and productivity effects where evidence permits.

Causal verification should accompany the scorecard. Simple before-and-after comparisons are vulnerable to weather, occupancy and tariff changes. Baseline models should estimate what would have occurred without the intervention, while uncertainty ranges should be reported. Maintenance analysis should test whether residuals, energy intensity and stability improved after repair; optimisation analysis should confirm that savings did not reduce service. This discipline converts analytics from persuasive visualisation into operational evidence.

The framework also clarifies technology selection. Neural networks and recurrent models are useful where complex time dependencies dominate; tree-based and support-vector methods can provide robust classification; reinforcement learning is suitable for sequential decisions; model-predictive control is preferable where constraints and models can be maintained; and rule-based analytics remain valuable for well-understood faults. Hybrid systems are generally more defensible than unrestricted black-box control because they combine engineering knowledge with adaptive capability [5,7]. The preferred design is the simplest architecture meeting the required decision quality and safety level.

VIII. IMPLEMENTATION, GOVERNANCE AND RESEARCH AGENDA

Implementation should begin with an asset and data readiness assessment. Facilities need a reliable equipment hierarchy, point list, sequence documentation, maintenance history and critical-service definition. Sensor coverage should be evaluated against decisions, not against a generic desire for more data. Flow measurement may be essential for plant efficiency, while vibration sensing may be justified only on critical rotating equipment. Existing valve, current and pressure signals can create useful virtual indicators before new hardware is installed.

The next step is to establish deterministic performance rules and baselines. These provide immediate value and a reference for testing advanced models. Automated checks for simultaneous heating and cooling, schedule drift, set-point override, excessive static pressure and abnormal cycling are understandable to operators. Predictive models can then be introduced for load, occupancy and equipment response. Control should remain advisory until data quality, model stability and operator workflows are proven. Automation should require documented acceptance tests and rollback procedures.

Cybersecurity and access control are integral because HVAC platforms connect operational technology, enterprise networks and cloud analytics. Control commands, model updates and maintenance recommendations require authentication, logging and least-privilege access. Remote optimisation should not bypass local safeties. Data governance should define retention, ownership and permissible use of occupancy information. Model governance should include version control, performance monitoring, bias checks and periodic engineering review.

Organisational design is equally important. Energy managers, controls specialists and maintenance teams need shared objectives and regular review. An energy-wasting fault should enter the maintenance backlog with a quantified consequence even without immediate service risk. A control change that saves energy but increases starts should be reviewed jointly. Procurement contracts should require interoperable data, documented interfaces and access to trend histories. Operating data and sequence logic should remain accessible to the owner.

Several research gaps remain. Field evidence is still weaker than simulation evidence, especially for multi-season operation and diverse building stocks. Comparative studies should report savings variance, comfort, cycling, faults and operator intervention. Fault-aware control requires further development so that optimisation can respond safely to uncertain sensors and degraded equipment. Diagnostic research needs better evaluation of combined faults, rare events and the economic cost of false alarms. Digital-twin studies should address model maintenance and configuration change rather than only initial calibration.

Future research should also examine portfolio-scale learning. Large owners may operate hundreds of similar air-handling units or rooftop systems. Shared models could identify common degradation patterns, while local adaptation would account for climate and usage. Privacy-preserving learning could support cross-site improvement without centralising sensitive data. Explainable optimisation is another priority: operators need to know which constraints and forecasts drove a recommendation. Finally, resilience studies should test operation during extreme weather, power constraints and partial equipment failure. These conditions reveal whether intelligent control improves continuity beyond normal operation.

IX. CONCLUSION

Advanced HVAC operation can contribute simultaneously to sustainability, energy optimisation and mechanical-system continuity, but only when control and maintenance are designed as one operating capability. Forecasting, predictive control, reinforcement learning and occupancy-responsive strategies can reduce avoidable conditioning and improve coordination. Fault detection, virtual sensing, digital twins and predictive maintenance can identify degradation before service is lost. Their value is highest when diagnostics alter control safely, generate prioritised maintenance work and verify recovery after intervention.

The review shows that algorithm performance alone is an insufficient basis for deployment. Durable outcomes depend on sensor quality, physical constraints, fallback logic, explainability, work-management integration and multi-season verification. Energy savings that reduce service or accelerate wear are not sustainable; reliability strategies that ignore energy waste are incomplete. A balanced scorecard should therefore govern energy, carbon, comfort, air quality, cycling, fault duration, repair time and downtime.

The proposed architecture and feedback loop provide a practical route from fragmented analytics to resilient building operation. Organisations should progress from data readiness and advisory diagnostics to bounded supervisory control, using evidence at each stage. Future work should prioritise field validation, fault-aware optimisation and portfolio-scale learning. The central lesson is clear: intelligent HVAC systems create lasting value when they continuously connect what the building needs, how the plant is actually performing and what maintenance action will preserve efficient service.

REFERENCES

[1] Aghili, S.A.; Haji Mohammad Rezaei, A.; Tafazzoli, M.; Khanzadi, M.; Rahbar, M. Artificial Intelligence Approaches to Energy Management in HVAC Systems: A Systematic Review. Buildings 2025, 15, 1008.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722144

[2] Peng, Y.; Lei, Y.; Tekler, Z.D.; Antanuri, N.; Lau, S.-K.; Chong, A. Hybrid System Controls of Natural Ventilation and HVAC in Mixed-Mode Buildings: A Comprehensive Review. Energy and Buildings 2022, 276, 112509.

[3] Alanne, K.; Sierla, S. An Overview of Machine Learning Applications for Smart Buildings. Sustainable Cities and Society 2022, 76, 103445.

[4] Sierla, S.; Ihasalo, H.; Vyatkin, V. A Review of Reinforcement Learning Applications to Control of Heating, Ventilation and Air Conditioning Systems. Energies 2022, 15, 3526.

[5] Tien, P.W.; Wei, S.; Darkwa, J.; Wood, C.; Calautit, J.K. Machine Learning and Deep Learning Methods for Enhancing Building Energy Efficiency and Indoor Environmental Quality: A Review. Energy and AI 2022, 10, 100198.

[6] Nelson, W.; Culp, C. Machine Learning Methods for Automated Fault Detection and Diagnostics in Building Systems: A Review. Energies 2022, 15, 5534.

[7] Halhoul Merabet, G.; Essaaidi, M.; Ben Haddou, M.; Qolomany, B.; Qadir, J.; Anan, M.; Al-Fuqaha, A.; Abid, M.R.; Benhaddou, D. Intelligent Building Control Systems for Thermal Comfort and Energy-Efficiency: A Systematic Review of Artificial Intelligence-Assisted Techniques. Renewable and Sustainable Energy Reviews 2021, 144, 110969.

[8] Miao, Y.; Yao, Y.; Hong, X.; Xiong, L.; Zhang, F.; Chen, W. Research on Optimal Control of HVAC System Using Swarm Intelligence Algorithms. Building and Environment 2023, 241, 110467.

[9] Bouabdallaoui, Y.; Lafhaj, Z.; Yim, P.; Ducoulombier, L.; Bennadji, B. Predictive Maintenance in Building Facilities: A Machine Learning-Based Approach. Sensors 2021, 21, 1044.

[10] Xiong, L.; Yao, Y. Study on an Adaptive Thermal Comfort Model with K-Nearest-Neighbors Algorithm. Building and Environment 2021, 202, 108026.

[11] Yu, L.; Xie, W.; Xie, D.; Zou, Y.; Zhang, D.; Sun, Z.; Zhang, L.; Zhang, Y.; Jiang, T. Deep Reinforcement Learning for Smart Home Energy Management. IEEE Internet of Things Journal 2020, 7, 2751-2762.

[12] Yan, K.; Chong, A.; Mo, Y. Generative Adversarial Network for Fault Detection Diagnosis of Chillers. Building and Environment 2020, 172, 106698.

[13] Yu, L.; Sun, Y.; Xu, Z.; Shen, C.; Yue, D.; Jiang, T.; Guan, X. Multi-Agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings. IEEE Transactions on Smart Grid 2021, 12, 407-419.

[14] Brandi, S.; Piscitelli, M.S.; Martellacci, M.; Capozzoli, A. Deep Reinforcement Learning to Optimise Indoor Temperature Control and Heating Energy Consumption in Buildings. Energy and Buildings 2020, 224, 110225.

[15] Du, Y.; Zandi, H.; Kotevska, O.; Kurte, K.; Munk, J.; Amasyali, K.; McKee, E.; Li, F. Intelligent Multi-Zone Residential HVAC Control Strategy Based on Deep Reinforcement Learning. Applied Energy 2021, 281, 116117.

[16] Dai, X.; Liu, J.; Zhang, X. A Review of Studies Applying Machine Learning Models to Predict Occupancy and Window-Opening Behaviours in Smart Buildings. Energy and Buildings 2020, 223, 110159.

[17] Gao, G.; Li, J.; Wen, Y. DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings via Reinforcement Learning. IEEE Internet of Things Journal 2020, 7, 8472-8484.

[18] Zou, Z.; Yu, X.; Ergan, S. Towards Optimal Control of Air Handling Units Using Deep Reinforcement Learning and Recurrent Neural Network. Building and Environment 2020, 168, 106535.

[19] Esrafilian-Najafabadi, M.; Haghighat, F. Occupancy-Based HVAC Control Systems in

Buildings: A State-of-the-Art Review. Building and Environment 2021, 197, 107810.

[20] Taheri, S.; Razban, A. Learning-Based CO2 Concentration Prediction: Application to Indoor Air Quality Control Using Demand-Controlled Ventilation. Building and Environment 2021, 205, 108164.

[21] Ahn, K.U.; Park, C.S. Application of Deep Q-Networks for Model-Free Optimal Control Balancing between Different HVAC Systems. Science and Technology for the Built Environment 2020, 26, 61-74.

[22] Deng, Z.; Wang, X.; Jiang, Z.; Zhou, N.; Ge, H.; Dong, B. Evaluation of Deploying Data-Driven Predictive Controls in Buildings on a Large Scale for Greenhouse Gas Emission Reduction. Energy 2023, 270, 126934.

[23] Mohseni, S.-R.; Zeitouni, M.J.; Parvaresh, A.; Abrazeh, S.; Gheisarnejad, M.; Khooban, M.-H. FMI Real-Time Co-Simulation-Based Machine Deep Learning Control of HVAC Systems in Smart Buildings: Digital-Twins Technology. Transactions of the Institute of Measurement and Control 2023, 45, 661-673.

[24] Weinberg, D.; Wang, Q.; Timoudas, T.O.; Fischione, C. A Review of Reinforcement Learning for Controlling Building Energy Systems from a Computer Science Perspective. Sustainable Cities and Society 2023, 89, 104351.

[25] Deng, X.; Zhang, Y.; Qi, H. Towards Optimal HVAC Control in Non-Stationary Building Environments Combining Active Change Detection and Deep Reinforcement Learning. Building and Environment 2022, 211, 108680.

[26] Elnour, M.; Himeur, Y.; Fadli, F.; Mohammedsherif, H.; Meskin, N.; Ahmad, A.M.; Petri, I.; Rezgui, Y.; Hodorog, A. Neural Network-Based Model Predictive Control System for Optimizing Building Automation and Management Systems of Sports Facilities. Applied Energy 2022, 318, 119153.

[27] Liu, X.; Ren, M.; Yang, Z.; Yan, G.; Guo, Y.; Cheng, L.; Wu, C. A Multi-Step Predictive Deep Reinforcement Learning Algorithm for HVAC Control Systems in Smart Buildings. Energy 2022, 259, 124857.

[28] Darwazeh, D.; Gunay, B.; Duquette, J. Development of Inverse Greybox Model-Based Virtual Meters for Air Handling Units. IEEE Transactions on Automation Science and Engineering 2021, 18, 323-336.

[29] Talei, H.; Benhaddou, D.; Gamarra, C.; Benbrahim, H.; Essaaidi, M. Smart Building Energy Inefficiencies Detection through Time Series Analysis and Unsupervised Machine Learning. Energies 2021, 14, 6042.

[30] Jiang, Z.; Risbeck, M.J.; Ramamurti, V.; Murugesan, S.; Amores, J.; Zhang, C.; Lee, Y.M.; Drees, K.H. Building HVAC Control with Reinforcement Learning for Reduction of Energy Cost and Demand Charge. Energy and Buildings 2021, 239, 110833.