

Predictive Maintenance of IoT-Enabled Biomedical Sensors Using Machine Learning to Enhance Patient Safety

Abhinandan Raj¹, Vaibhav Saunakiya²

¹ Master's Graduate, Software Engineering and Big Data, National Research Nuclear University MEPhI, Moscow, Russia

² Medical Student, National Research Nuclear University MEPhI, Moscow, Russia

Abstract - Internet of Things (IoT)-enabled biomedical sensors support continuous patient monitoring but may experience sensor drift, battery deterioration, communication interruptions, and hardware faults. Such failures can compromise measurement availability and reliability. This paper proposes a machine-learning-based predictive maintenance framework combining device-health telemetry, communication statistics, and verified maintenance records. The architecture comprises data acquisition, secure gateway communication, preprocessing, fault prediction, and maintenance alert management. Candidate methods include Random Forest, Isolation Forest, gradient boosting, and temporal models, compared against a transparent rule-based baseline. An evaluation protocol is defined using precision, recall, precision-recall area under the curve, false alarms per device-day, warning lead time, and inference latency. The work also addresses model drift, cybersecurity, and the separation of maintenance alerts from clinical decision-making. As a framework paper, it reports no measured model performance or patient outcomes. Experimental validation is necessary before operational or clinical claims can be made.

Keywords—Biomedical sensors; Internet of Things; predictive maintenance; machine learning; fault detection; patient safety.

I. INTRODUCTION

Connected pulse oximeters, electrocardiogram (ECG) devices, wearable monitors, and other biomedical sensors transmit physiological measurements through local networks or remote monitoring platforms. The reliability of these systems depends on sensing elements, batteries, firmware, communications, and data-processing infrastructure. A device can continue reporting while its measurements degrade, or it can fail to transmit observations altogether. Corrective maintenance responds to known faults, while scheduled preventive maintenance may miss degradation occurring between inspections. Predictive maintenance seeks to identify developing faults from operational signals before they disrupt service.

The objective is not to infer a patient's condition from device-maintenance data. Rather, it is to identify device faults that may reduce the reliability or availability of monitoring. Predictive alerts must remain subordinate to approved clinical alarm pathways and device-manufacturer procedures.

A. Objectives and contributions

This paper (1) defines a layered predictive-maintenance architecture, (2) maps plausible failure modes to measurable telemetry, (3) proposes supervised and unsupervised model baselines, (4) specifies a leakage-resistant validation plan, and (5) identifies safety and governance requirements. These are design contributions rather than empirically demonstrated performance gains.

II. RELATED WORK AND RESEARCH GAP

Healthcare IoT systems integrate sensors, network communication, and analytics, creating opportunities for remote monitoring as well as reliability, privacy, and interoperability challenges [1], [2]. Machine-learning approaches to predictive maintenance have been surveyed in

industrial settings, where label quality, imbalanced failures, and changing operating conditions are recurring limitations [3]. Isolation Forest is a widely studied unsupervised anomaly-detection method [4], and Random Forest provides a supervised classification baseline [5]. These methods are candidates rather than proven solutions for every biomedical device. Medical-device risk management and health software lifecycles impose additional requirements [6], [7].

A key research gap is the lack of validated, device-specific fault labels and cross-device generalisation evidence in many proposed systems. Clinical relevance also requires measuring alert burden and useful warning time, not only classification accuracy. A robust evaluation must distinguish sensor failures from actual physiological changes and network outages.

III. PROPOSED SYSTEM ARCHITECTURE

The system has five layers: (1) sensor and device telemetry, (2) authenticated IoT gateway, (3) data validation and feature engineering, (4) machine-learning fault estimation, and (5) maintenance alert management. Patient measurements should be separated from device-health data whenever feasible. Model outputs are advisory maintenance signals and must not suppress clinical alarms.

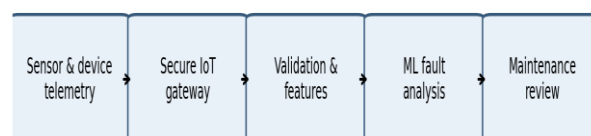


Fig. 1. Conceptual architecture of the proposed predictive-maintenance system.

TABLE I. PROPOSED TELEMETRY AND FAILURE INDICATORS

Subsystem	Candidate indicators	Possible failure mode
Sensing	Signal quality, noise, missing samples	Sensor degradation or poor contact
Battery	Voltage, discharge slope, charging events	Power deterioration
Communications	Packet loss, delay, disconnections	Link or interface instability
Device status	Error codes, internal temperature, resets	Hardware/firmware problem
Calibration	Deviation, last calibration date	Calibration drift
Service history	Confirmed faults and repair records	Fault labels and recurrence

IV. METHODOLOGY

A. Data acquisition and labelling

An empirical implementation should use an authorised, documented collection of device telemetry and independently verified fault events. For each device, record a pseudonymous identifier, device type, firmware version, timestamps, operating context, feature values, and maintenance outcomes. Distinguish naturally observed failures from controlled test faults. Do not inject faults into devices used for patient care.

B. Preprocessing and features

Deduplicate records, validate units and timestamps, flag missingness, align event logs, and calculate windowed statistics. For a telemetry feature x and a trailing window of w observations, the rolling mean is $\mu_t = (1/w) \sum_{i=0}^{w-1} x_{t-i}$. The rolling sample standard deviation is $s_t = \sqrt{[(1/(w-1)) \sum_{i=0}^{w-1} (x_{t-i} - \mu_t)^2]}$. Other candidates include battery-discharge slope, disconnections per hour, and error counts. Features must be computed only from data available at prediction time.

C. Prediction models

Train a rule-based threshold baseline and compare Random Forest [5], gradient-boosted trees, and Isolation Forest [4]. A temporal model may be added when longitudinal data volume justifies its complexity. For supervised learning, define $y_t = 1$ when a verified device fault occurs within a prespecified horizon H after time t , and 0 otherwise. A calibrated supervised model estimates $P(y_t = 1 \mid \text{telemetry available through } t)$. An Isolation Forest anomaly score is not automatically a calibrated fault probability.

TABLE II. CANDIDATE ANALYTICAL METHODS

Method	Primary role	Important limitation
Rules	Transparent baseline	Limited nonlinear interactions
Random Forest	Labelled fault classification	Needs verified labels
Gradient boosting	Structured telemetry prediction	Tuning and calibration needed
Isolation Forest	Novelty/anomaly detection	Anomaly is not synonymous with fault
Temporal model	Degradation trends	May require more data and compute

D. Alerts and safeguards

Alert thresholds should be selected using validation data, accounting for missed-fault severity and false-alarm burden. A high-risk alert must be reviewed under approved maintenance procedures. If monitoring integrity is uncertain, existing clinical escalation and backup monitoring pathways take precedence. Maintain auditable records of prediction, review, action, and post-service verification.

V. PROPOSED VALIDATION PROTOCOL

A reproducible evaluation should specify the device population, observation period, sampling frequency, number of independently verified failures, class prevalence, failure horizon, and exclusion criteria. Separate training, validation, and testing chronologically, with additional held-out devices where possible. Keep all observations from a failure episode within one partition. Fit imputation, scaling, feature selection, and hyperparameters only on training/validation data. Use the untouched test set for final evaluation.

TABLE III. EVALUATION METRICS

Metric	Interpretation
Precision = $TP/(TP+FP)$	Proportion of alerts that match verified faults
Recall = $TP/(TP+FN)$	Proportion of verified faults detected
$F1 = 2PR/(P+R)$	Precision–recall balance
PR-AUC	Performance across alert thresholds
False alerts/device-day	Operational alert burden
Warning lead time	Time from actionable alert to fault
Inference latency	Runtime per prediction
Energy/memory usage	Deployment feasibility

The fault-matching window, event-level counting rules, confidence intervals, and missing-data handling should be declared before reporting metrics. Performance should be stratified by sensor type and fault mechanism. Compare recall jointly with false-alert burden and warning lead time.

V-A. SYNTHETIC-DATA PROOF OF CONCEPT

To test software feasibility without claiming clinical validation, we generated 10,000 artificial telemetry observations from 100 simulated devices (100 time steps each). Features include battery percentage, a unitless signal-noise index, packet-loss rate, device temperature ($^{\circ}\text{C}$), and error count. A latent degradation variable increases over simulated time; fault labels are sampled from a logistic function of this latent variable and the observed features. This is a deliberately constructed simulation, not an empirical model of any specific medical device.

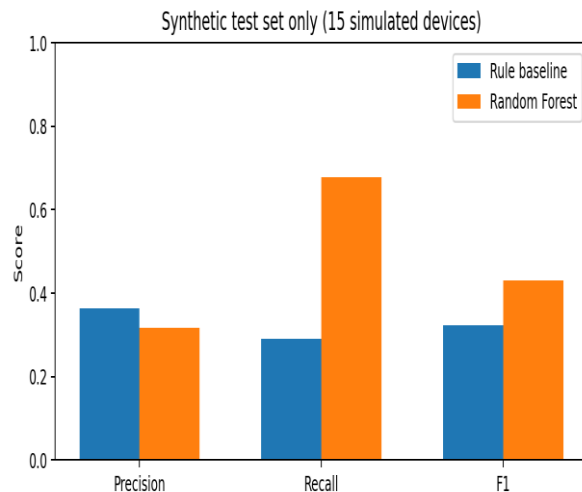
Devices 0–69 (7,000 observations) were used for training; devices 70–84 (1,500 observations) for threshold selection; and devices 85–99 (1,500 observations) for testing. A Random Forest (150 trees, maximum depth 9, minimum leaf size 8, random state 42) was compared with a transparent rule: flag if noise index > 0.55 , battery $< 25\%$, or packet loss > 0.25 . The classification threshold was selected on the validation set to maximise F1; no test labels were used for threshold selection.

TABLE IV. PERFORMANCE ON ARTIFICIALLY GENERATED TEST DATA

Method	Precision	Recall	F1
Rule-based	0.363	0.290	0.322
Random Forest	0.316	0.678	0.431

The synthetic test set contained 1500 observations, including 183 generated positive fault labels. The Random Forest validation-selected threshold was 0.18; its test-set average precision (PR-AUC) was 0.411. Because the label-generation process embeds feature relationships, these metrics must not be generalised to real sensors, real failure lead times, or patient outcomes.

Fig. 2. Precision, recall and F1 for the synthetic-data demonstration.



VI. DISCUSSION AND PATIENT SAFETY

The proposed system may enable earlier maintenance planning and help distinguish communication faults from sensing problems, but these benefits are hypotheses. A model can miss failures, misclassify environmental disturbances, or become unreliable following firmware or hardware changes. Consequently, performance monitoring, periodic recalibration, human oversight, and safe fallback mechanisms are essential. Device-maintenance alerts cannot independently certify that physiological readings are clinically correct.

Security controls should include access restrictions, authenticated transport, encryption where appropriate, software update integrity, audit logging, and data minimisation. Avoid collecting patient identifiers unless required by a defined, authorised purpose. A formal risk assessment should address false negatives, excessive false positives, communication loss, and model unavailability, consistent with applicable device safety and software lifecycle requirements [6], [7].

VII. LIMITATIONS AND FUTURE WORK

This study includes a reproducible synthetic-data demonstration, not measurements from actual biomedical sensors. The artificial fault labels are generated by a stipulated mathematical process; consequently, predictive scores reflect that process and do not establish performance on real devices or any improvement in patient safety. Future work requires real-world device data, independently verified failure labels, external validation, and clinical risk assessment.

VIII. CONCLUSION

A predictive-maintenance architecture for IoT-enabled biomedical sensors has been proposed, integrating device telemetry, secure acquisition, fault analytics, and reviewed maintenance alerts. The manuscript defines candidate models and a validation protocol centred on fault recall, alert burden, and useful warning time. Patient-safety requirements demand independent clinical safeguards and careful risk management. The framework offers a basis for reproducible experimentation; effectiveness remains unverified until empirical testing is completed.

REFERENCES

- [1] S. M. R. Islam et al., "The Internet of Things for Health Care: A Comprehensive Survey," *IEEE Access*, vol. 3, pp. 678–708, 2015, doi: 10.1109/ACCESS.2015.2437951.
- [2] M. M. Baig et al., "A Systematic Review of Wearable Patient Monitoring Systems—Current Challenges and Opportunities for Clinical Adoption," *Journal of Medical Systems*, vol. 41, article 115, 2017, doi: 10.1007/s10916-017-0760-1.
- [3] T. P. Carvalho et al., "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering*, vol. 137, article 106024, 2019, doi: 10.1016/j.cie.2019.106024.
- [4] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in *Proc. IEEE International Conference on Data Mining*, 2008, pp. 413–422, doi: 10.1109/ICDM.2008.17.
- [5] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [6] ISO 14971:2019, *Medical devices—Application of risk management to medical devices*, International Organization for Standardization, 2019.
- [7] IEC 62304:2006+A1:2015, *Medical device software—Software life cycle processes*, International Electrotechnical Commission.

DATA AVAILABILITY

No real biomedical device or patient data were collected. A fully synthetic, simulated telemetry dataset (10,000 rows from 100 simulated devices) was generated using a fixed random seed (20261028). The accompanying CSV and generation script are supplied as reproducibility supplements. These records are not clinical observations.