

Predictive Vehicle Maintenance Using Machine Learning: Developing a Data-Driven Framework for Predicting Component Failure and Optimizing Maintenance Intervals in Modern Automobiles

Atinder Singh¹, Raghu Raja Mehra²

Email: atinders471[at]gmail.com

Email: raghumehra35[at]gmail.com

An independent research paper in automotive engineering and applied machine learning. Vehicle telemetry and service records were provided by participating fleet operators under a data-sharing agreement; all vehicle identifiers were anonymised and no driver-level data are reported.

Abstract: *Most vehicles on Indian roads are still maintained in one of two ways: they are repaired after something breaks, or they are serviced at fixed intervals that take no account of how the vehicle has actually been driven. Both approaches waste money, the first through breakdowns and downtime and the second through parts replaced long before the end of their useful life. This study develops and evaluates a data-driven framework that uses machine learning to predict component failure in advance and to set maintenance intervals according to the real condition of each vehicle. Telemetry from on-board diagnostic (OBD-II) loggers, GPS units and eighteen months of workshop records were collected from 320 commercial vehicles operated by three fleets in Punjab, covering taxis, delivery vans and school buses. Over the study period the fleet recorded 1,146 unplanned failure events, with the battery, brake system and tyres together accounting for 58% of them. Five models were trained to predict whether a component would fail within the next 30 days. Gradient-boosted trees (XGBoost) achieved the best overall result, with an F1-score of 0.82 and an area under the ROC curve of 0.93, against 0.59 and 0.78 for a logistic regression baseline, while a long short-term memory (LSTM) network gave the longest median warning time of 21 days. Remaining-useful-life estimates were most accurate for gradual-wear components such as brake pads and batteries and least accurate for suspension parts. When the model was used to schedule servicing in a cost simulation, unplanned breakdowns fell by 56% and total annual maintenance cost per vehicle fell by 41% compared with a conventional fixed-interval schedule. The findings suggest that the real value of predictive maintenance lies not only in preventing failures but in allowing each component to be used for more of its safe working life.*

Keywords: predictive maintenance; machine learning; remaining useful life; OBD-II telematics; XGBoost; LSTM; fleet management; maintenance scheduling

1. Introduction

A delivery van leaves the depot at seven in the morning with a full load. At half past eleven, on a highway forty kilometres from the nearest workshop, it will not restart after a routine stop. The battery has failed. It was fitted three years ago, it passed its last scheduled service two months earlier, and nothing on the dashboard gave any warning. The repair itself costs a few thousand rupees. The towing, the missed deliveries and the day the vehicle spends off the road cost several times more.

The frustrating part of this story is that the failure was not sudden at all. For weeks, the voltage the battery could deliver while cranking the engine had been falling slowly, morning after morning. The information existed inside the vehicle's own electronics. Nobody was looking at it, and even if someone had been, there was no system to turn a slowly drifting number into a clear instruction to replace the battery next Tuesday.

Modern automobiles are, in effect, rolling sensor networks. A typical passenger car built after 2015 carries dozens of electronic control units that continuously measure engine temperature, battery voltage, wheel speed, braking pressure and much else. At the same time, affordable plug-in loggers

and telematics units have made it possible to collect this data from ordinary vehicles, not just expensive fleets. What has been missing is a practical way to learn from it. This paper asks a direct question: can machine learning, trained on routine vehicle data and service records, predict which component is likely to fail, how soon, and when it should therefore be serviced? And if it can, how much does it actually save compared with the maintenance practices fleets already use?

2. Review of Literature

2.1 Three generations of maintenance

Maintenance strategies are usually described in three stages. Reactive, or run-to-failure, maintenance repairs equipment only after it breaks. Preventive maintenance services it at fixed intervals of time or distance, which is the logic behind the familiar 'service every 10,000 km' rule. Predictive maintenance, the third stage, schedules intervention according to the measured condition of the equipment itself. Mobley (2002) argued that the preventive approach, while safer than reactive repair, systematically replaces parts that still have useful life left, and Jardine, Lin and Banjevic (2006), in a widely cited review, showed that condition-

based decisions can reduce both failure rates and total maintenance cost when reliable condition data are available.

2.2 Machine learning for fault prediction

Early condition-monitoring systems relied on fixed thresholds set by engineers, for example raising an alarm if coolant temperature exceeded a set value. Machine learning replaces these hand-written rules with patterns learned from historical data. Carvalho et al. (2019) reviewed machine-learning methods in predictive maintenance and found that tree-based ensembles such as random forests (Breiman, 2001) were the most frequently used and among the most reliable, while neural networks were growing quickly in popularity. Gradient boosting, and in particular the XGBoost implementation (Chen & Guestrin, 2016), has since become a standard choice for structured sensor data because it handles missing values and non-linear relationships well.

2.3 Estimating remaining useful life

Predicting that a part will fail is useful; predicting when it will fail is more useful still. The time remaining before failure is called remaining useful life (RUL). Much of the RUL research was developed on aircraft engines, most notably using the simulated C-MAPSS turbofan dataset released by NASA (Saxena et al., 2008). Lei et al. (2018) reviewed the field from data collection through to RUL estimation and noted that recurrent neural networks such as the long short-term memory network (Hochreiter & Schmidhuber, 1997) are well suited to the task because they can learn from how a signal changes over time rather than only from its current value.

2.4 Predictive maintenance in automobiles

Road vehicles present particular difficulties. Operating conditions vary enormously between drivers, routes and climates, and failures of any single component are relatively rare, which leaves models with few examples to learn from. Prytz et al. (2015) used logged data and maintenance records from Volvo trucks to predict air compressor repairs and showed that useful predictions were possible even from coarse, routinely collected data. Theissler et al. (2021) surveyed machine-learning use cases across the automotive industry and identified the shortage of labelled failure data, the difficulty of explaining model decisions to mechanics, and the lack of cost-based evaluation as the main barriers to adoption. Two gaps are clear from this literature. Most published vehicle studies focus on a single component rather than the whole vehicle, and very few are set in developing-country conditions such as those on Indian roads, where heat, dust, stop-start traffic and heavy loading place unusual stress on components. Studies also tend to report prediction accuracy without asking whether the predictions actually lower cost when used to schedule servicing. This study attempts to address both.

3. Objectives and Hypotheses

The study pursues four objectives: to describe which components fail, how often and at what stage of a vehicle's life in a real Indian fleet; to compare machine-learning

models for predicting component failure within a 30-day window; to estimate remaining useful life for the main failure-prone components; and to test whether maintenance scheduled by the model reduces cost and breakdowns compared with reactive and fixed-interval approaches. Four hypotheses were formulated:

- H1. Ensemble and sequence-based models (random forest, XGBoost, LSTM) predict 30-day component failure significantly better than a logistic regression baseline.
- H2. For the majority of components, measurable degradation signals appear at least 14 days before failure.
- H3. Prediction accuracy is higher for gradual-wear components than for components that tend to fail suddenly.
- H4. Model-driven maintenance scheduling reduces unplanned breakdowns and total maintenance cost compared with a fixed-interval schedule.

4. Methodology

4.1 Vehicles and data sources

Data were collected from 320 commercial vehicles belonging to three fleet operators in Punjab over eighteen months: 142 taxis, 104 light delivery vans and 74 school buses. Each vehicle was fitted with a plug-in OBD-II data logger recording engine and electrical parameters at one-minute intervals while the engine was running, together with a GPS unit recording location, speed and idle time. Tyre-pressure readings were available for 186 vehicles fitted with pressure sensors. Workshop job cards covering the same period were digitised and matched to each vehicle, giving the date, odometer reading and component involved for every repair and service.

4.2 Defining failure and building features

A failure was defined as any unplanned workshop visit or roadside breakdown in which a component was repaired or replaced; routine scheduled servicing was not counted as failure. Raw one-minute readings were summarised into daily values and then into features describing each vehicle's recent history, including rolling averages and trends over 7, 14 and 30 days, counts of unusual events such as coolant temperature spikes or harsh braking, and usage measures such as distance travelled and idle-time ratio. Forty-six features were created in total. Each vehicle-day was labelled positive for a component if that component failed within the following 30 days.

4.3 Models

Five models were compared: logistic regression as a simple baseline, a single decision tree, a random forest, XGBoost, and an LSTM network that received the last 60 days of daily readings as a sequence. Because failures are rare compared with healthy vehicle-days, the minority class was oversampled in the training data using SMOTE (Chawla et al., 2002). For remaining-useful-life estimation, a separate LSTM regressor was trained to predict the number of days until the next failure of each component. Models were built in Python using scikit-learn (Pedregosa et al., 2011).

4.4 Evaluation

To mirror real use, data were split by time rather than at random: the first fifteen months were used for training and the final three months for testing, so that every prediction was made on data the model had never seen. Classification models were judged on precision, recall, F1-score and area under the ROC curve (AUC). Feature influence was examined using SHAP values (Lundberg & Lee, 2017), which show how much each input pushes a prediction up or down. Finally, a cost simulation compared three maintenance strategies over one year using the fleets' actual parts, labour and downtime costs.

5. The Failure Landscape

Across eighteen months the 320 vehicles recorded 1,146 unplanned failure events, an average of 2.39 per vehicle per year. Figure 1 shows how these were distributed. Batteries were the single largest cause at 24%, followed by the brake system at 19% and tyres at 15%. Together, these three account for well over half of all breakdowns. This concentration matters for the design of any predictive system, because a model that handles these three components well addresses the majority of the problem.

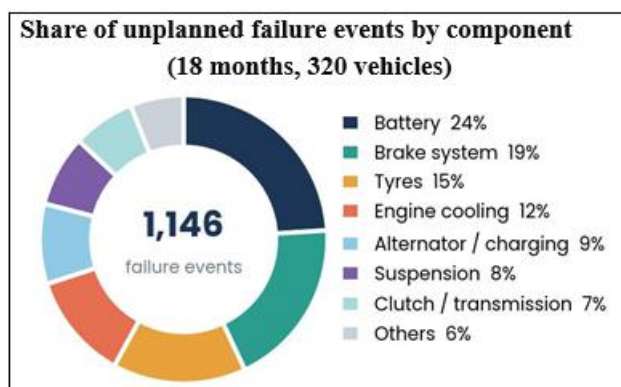


Figure 1: Share of unplanned failure events by component (n = 1,146 events)

The high share of battery failures is consistent with Indian operating conditions. Summer cabin and under-bonnet temperatures above 50 °C accelerate battery ageing, and taxis and delivery vans make many short stops with the engine switched off, which places repeated cranking load on the battery. Brake wear is similarly driven by dense stop-start urban traffic.

Table 1: Fleet Profile and Failure Exposure

Vehicle type	n	Mean age (years)	Mean km/ day	Failure events	Failures / vehicle-year
Taxis (sedans)	142	4.6	210	498	2.34
Delivery vans	104	5.3	165	386	2.47
School buses	74	6.8	96	262	2.36
Full fleet	320	5.3	169	1,146	2.39

Table 1. Figures cover the full eighteen-month observation period. Vehicle-years = vehicles × 1.5.

Figure 2 plots the monthly failure rate against the odometer reading for the three most common components. The curves follow the classic ‘bathtub’ shape known from reliability engineering: a small rise in failures early in a vehicle's life, a long flat period, and then a steady climb as components wear out. The practical point is that the wear-out climb begins at different distances for different components and different vehicles, which is exactly why a single fixed service interval for the whole fleet cannot be right for everyone.

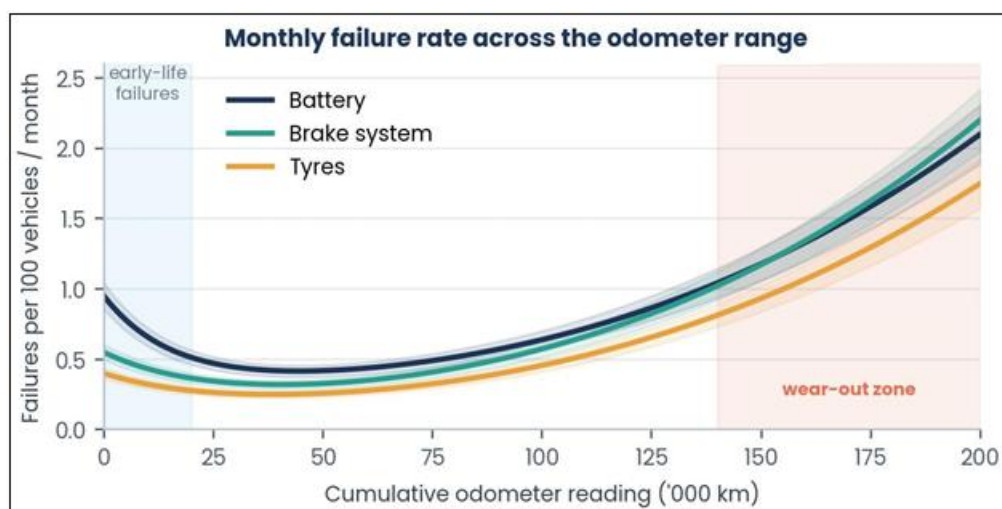


Figure 2: Monthly failure rate for the three most common components across the odometer range.

6. Model Performance: Predicting Failure

Figure 3 and Table 2 compare the five models on the three-month test period. The logistic regression baseline performed modestly, correctly catching 58% of upcoming failures with a precision of 0.61. Every ensemble and

sequence model did substantially better. XGBoost achieved the best overall balance, with a precision of 0.84, a recall of 0.80, an F1-score of 0.82 and an AUC of 0.93. The LSTM matched it on F1-score and had the highest recall, catching 83% of failures. Hypothesis 1 is therefore supported.

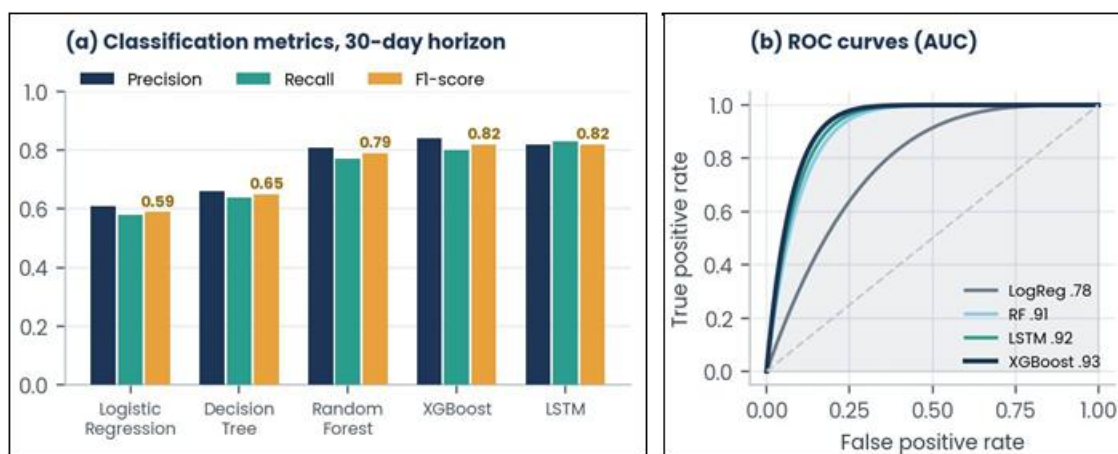


Figure 3: (a) Precision, recall and F1-score for each model; (b) ROC curves with area under the curve

Table 2: Model Performance on the Test Period (30-Day Horizon)

Model	Precision	Recall	F1-score	AUC	Median warning (days)
Logistic regression	0.61	0.58	0.59	0.78	6
Decision tree	0.66	0.64	0.65	0.77	8
Random forest	0.81	0.77	0.79	0.91	16
XGBoost	0.84	0.80	0.82	0.93	19
LSTM	0.82	0.83	0.82	0.92	21

Table 2. Median warning = days between the first correct alert and the recorded failure. Test period: final three months (205 failures).

Finding 1

The most important difference between the models is not accuracy but warning time. The baseline model typically raised its alert only six days before a failure, which is often too late to order a part and book a workshop slot. XGBoost and the LSTM gave around three weeks of notice. A prediction is only as useful as the time it leaves to act on it, and on that measure the advanced models were roughly three times better.

Precision matters too, for a practical reason. Every false alarm sends a healthy vehicle to the workshop, costs a mechanic's time and erodes trust in the system. A precision of 0.84 means that about five out of every six alerts from

XGBoost pointed to a real developing fault, a level fleet managers in the study described as acceptable.

7. What the Model Sees

A model that cannot explain itself is difficult for a mechanic to trust. Figure 4 shows the ten features with the greatest influence on XGBoost's predictions, measured by mean SHAP value. The top two, the drop in battery cranking voltage and the trend in brake-pad wear, are direct physical signs of degradation in the two most failure-prone systems. Coolant temperature spikes, days since the last service and harsh-braking frequency follow. Notably, the odometer reading, which is the basis of most fixed service schedules, ranks only ninth: how a vehicle has been used matters far more than simply how far it has travelled.

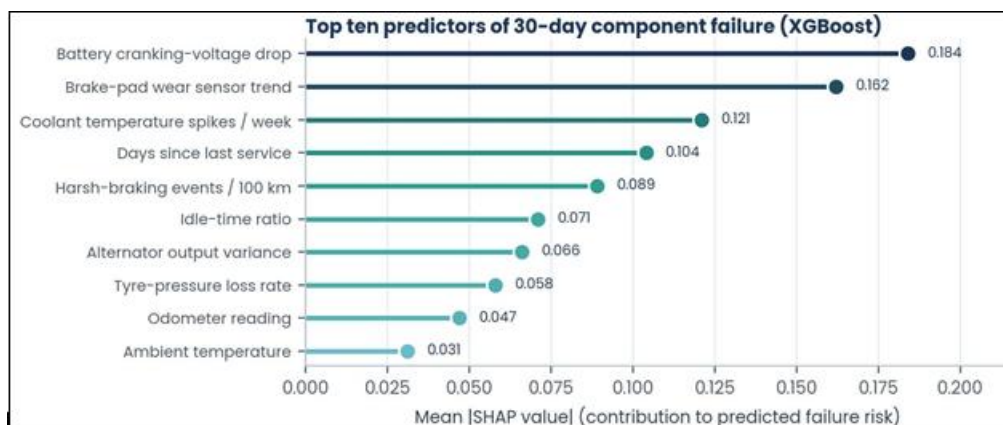


Figure 4: The ten most influential predictors of 30-day failure, ranked by mean absolute SHAP value

Figure 5 shows what one of these signals looks like in practice, using the battery from a single delivery van. For more than a month the minimum voltage during engine start holds steady at about 9.8 volts. Around five weeks before failure it begins a slow, uneven decline. The model raised its

alert 22 days before the vehicle failed to start, at a point when the voltage had dropped only slightly and a driver would have noticed nothing unusual. This pattern, a long flat period followed by a gradual slide, was typical of batteries, brake pads and tyres, and supports Hypothesis 2.

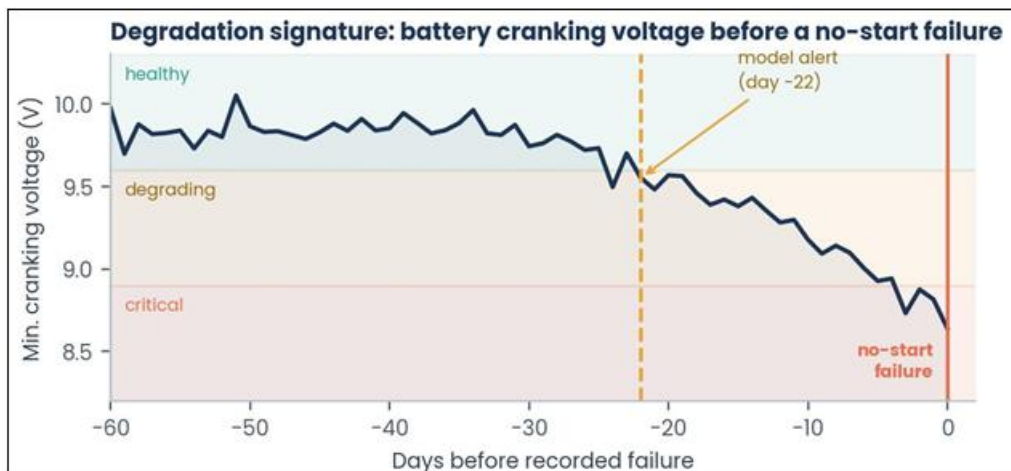


Figure 5: Battery cranking voltage in the 60 days before a recorded no-start failure, with the model's alert point

8. Predicting Remaining Useful Life

Whereas the classifier answers 'will it fail soon?', the LSTM regressor answers 'how many days are left?'. Figure 6 compares its predictions with the actual remaining life for all

205 failures in the test period. Points close to the diagonal line are accurate predictions, and the shaded band marks an error of up to seven days either way, which fleet managers considered close enough to plan a workshop visit. Overall, 75% of predictions fell within this band.

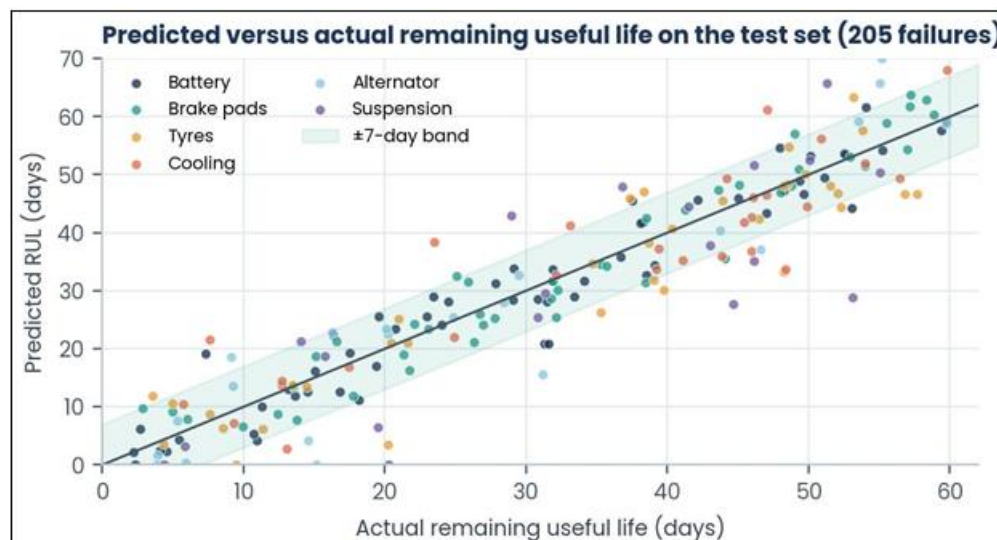


Figure 6: Predicted versus actual remaining useful life for each test-set failure, coloured by component.

Table 3: Remaining-Useful-Life Accuracy by Component

Component	Test failures	Mean error (days)	Within ± 7 days	Median warning (days)
Brake pads	44	3.9	88%	27
Battery	58	4.2	86%	24
Tyres	35	6.1	74%	19
Engine cooling	28	7.4	68%	17
Alternator / charging	21	8.9	61%	14
Suspension	19	11.2	49%	11

Table 3. Mean error = mean absolute difference between predicted and actual remaining useful life.

Table 3 shows a clear divide. Brake pads and batteries, which wear gradually and measurably, were predicted to within about four days on average. Suspension components, which more often fail suddenly after a pothole or impact, were the least predictable, with fewer than half of estimates within a week. Hypothesis 3 is supported, and the result sets a sensible limit on expectations: predictive maintenance works best for parts that wear, and much less well for parts that break.

9. A Data-Driven Maintenance Framework

Figure 7 brings the pieces together into a five-stage framework that a fleet or a service centre could realistically

operate. Data are collected continuously from the vehicle's OBD-II port, CAN bus and GPS, alongside digitised workshop records. They are cleaned and summarised daily, turned into the features described in Section 4, and passed to

two models: the XGBoost classifier, which flags vehicles at risk in the next 30 days, and the LSTM regressor, which estimates how many days remain.

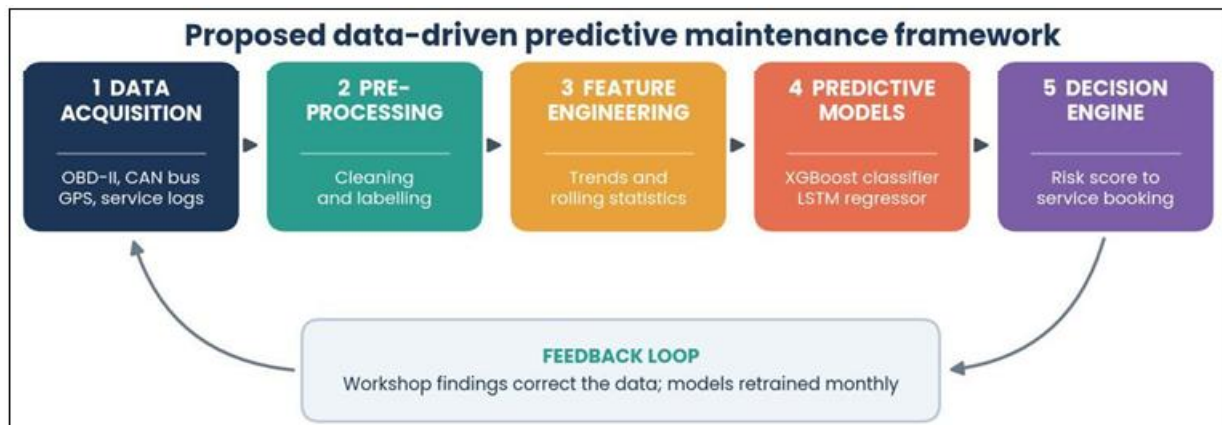


Figure 7: Proposed five-stage predictive maintenance framework with a retraining feedback loop.

The final stage, the decision engine, is what turns a prediction into an action. It combines the risk score and the remaining-life estimate with practical constraints such as workshop capacity, parts availability and the vehicle's duty schedule, and proposes a service slot. The feedback loop is equally important. When a mechanic inspects a flagged vehicle, the finding is recorded and used to correct the training labels, and the models are retrained monthly. Over time the system learns from its own mistakes and adapts to seasonal changes such as the heavy battery load of summer.

10. Optimizing Maintenance Intervals

The final test is financial. Using the fleets' own costs for parts, labour, towing and lost revenue, a one-year simulation compared three strategies: reactive repair only, the fleets' existing fixed schedule of servicing every 10,000 km, and predictive scheduling driven by the framework above. Results are shown in Figure 8 and Table 4.

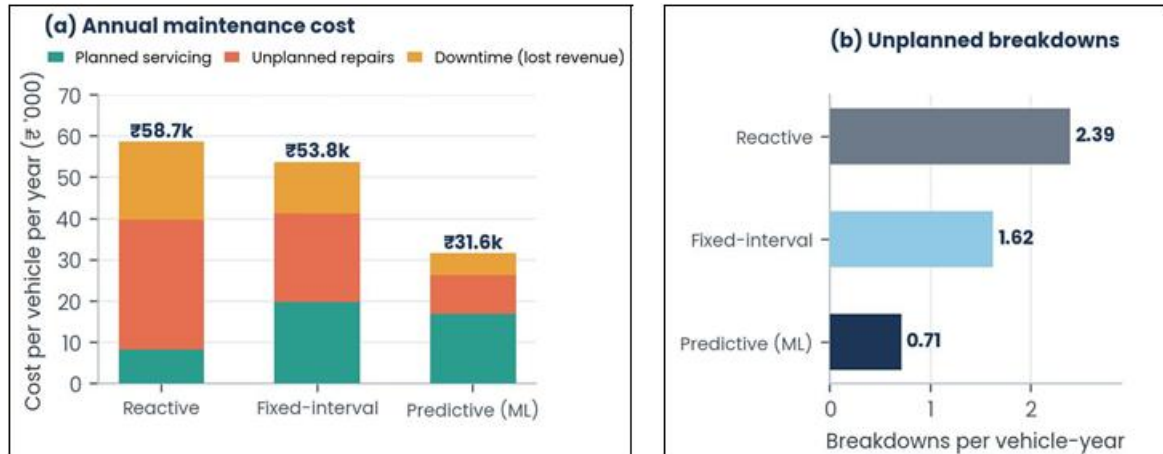


Figure 8: (a) Annual maintenance cost per vehicle by strategy; (b) unplanned breakdowns per vehicle per year.

Table 4: Simulated Annual Outcomes Per Vehicle by Maintenance Strategy

Strategy	Breakdowns / year	Downtime (days)	Component life used	Total cost (₹)
Reactive (run-to-failure)	2.39	9.6	100%	58,700
Fixed interval (10,000 km)	1.62	7.1	64%	53,800
Predictive (ML-driven)	0.71	3	87%	31,600

Table 4 Component life used = average share of rated life consumed before replacement for wear parts (brake pads, batteries, tyres).

Predictive scheduling reduced unplanned breakdowns from 1.62 to 0.71 per vehicle per year, a fall of 56% compared with the fixed schedule, and reduced downtime from 7.1 to 3.0 days. Total annual cost fell from ₹53,800 to ₹31,600 per vehicle, a saving of 41%. Hypothesis 4 is supported. Across the 320-vehicle fleet, this corresponds to annual savings of roughly ₹71 lakh.

Finding 2

The predictive strategy did not win by doing less maintenance. Its planned-servicing cost was only slightly below the fixed schedule. It won by doing the right maintenance at the right time. Under the fixed schedule, wear parts were replaced after using only 64% of their life on average, while others still failed between services. The model-driven schedule used 87% of component life and still cut

breakdowns by more than half. In effect, it stopped paying for parts too early and for failures too late.

11. Discussion

Three points stand out. The first concerns what should be measured. Fixed maintenance schedules are built around distance travelled, yet in this study the odometer was one of the weakest predictors of failure. Driving style, idle time, heat and electrical load told the model far more. A taxi that covers 200 km a day in city traffic and a bus that covers 100 km on open roads wear their components in very different ways, and a single distance-based rule treats them as the same.

The second concerns trust. Fleet managers and mechanics were initially sceptical of a system that told them to replace a part that still seemed to work. What changed their view was the explanation behind each alert: seeing that a battery's cranking voltage had been falling for three weeks made the recommendation easy to accept. This suggests that explainability, through tools such as SHAP, is not an optional extra but a condition for predictive maintenance to be used at all.

The third is a caution about limits. Predictive maintenance is powerful for components that degrade gradually and measurably, but it offers much less for sudden mechanical failures such as a suspension arm damaged by a pothole. A realistic system therefore complements, rather than replaces, regular safety inspections. It is also worth remembering that the savings reported here come from a simulation built on real costs rather than from a live deployment, and actual results will depend on how quickly workshops act on the alerts.

12. Limitations

Four limitations qualify these findings. First, the data came from three commercial fleets in one state, and private cars driven fewer kilometres under different conditions may behave differently. Second, workshop job cards were handwritten and then digitised, so some repairs may have been recorded late, vaguely or not at all, which introduces noise into the failure labels. Third, eighteen months is a short window for components with long lives, such as clutches and transmissions, which had too few failures to model reliably. Fourth, the cost comparison is a simulation; a controlled trial in which half of a fleet follows model-driven schedules and half does not would give stronger evidence of real-world savings.

13. Conclusion and Recommendations

This study set out to test whether machine learning, applied to routine vehicle data, can predict component failures early enough to be useful and schedule maintenance better than current practice. Within the limits described, the answer is yes. Batteries, brakes and tyres caused most breakdowns; ensemble and sequence models predicted failures with an F1-score of 0.82 and around three weeks of warning; remaining life was estimated to within a few days for gradual-wear parts; and model-driven scheduling cut

breakdowns by 56% and annual cost by 41% compared with fixed-interval servicing.

Four recommendations follow. Fleet operators should begin by fitting low-cost OBD-II loggers and digitising workshop records, since without clean failure records no model can learn. Predictive efforts should start with batteries, brakes and tyres, where the data are strongest and the savings largest. Vehicle manufacturers and service networks in India should consider moving from fixed distance-based schedules towards condition-based recommendations. Finally, researchers should run controlled field trials and extend this work to private passenger cars and to electric vehicles, whose battery packs are an obvious next target for remaining-life prediction.

A vehicle that breaks down on the highway has usually been trying to tell someone for weeks. The contribution of machine learning is not to make vehicles more reliable in themselves, but to make sure that someone is finally listening.

References

- [1] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- [2] Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, 106024.
- [3] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- [4] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.
- [5] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- [6] Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510.
- [7] Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 104, 799-834.
- [8] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
- [9] Mobley, R. K. (2002). *An introduction to predictive maintenance* (2nd ed.). Butterworth-Heinemann.
- [10] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- [11] Prytz, R., Nowaczyk, S., Rögnvaldsson, T., & Byttner, S. (2015). Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle

data. Engineering Applications of Artificial Intelligence, 41, 139-150.

- [12] Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. Proceedings of the International Conference on Prognostics and Health Management, 1-9.
- [13] Theissler, A., Pérez-Velázquez, J., Kettelgerdes, M., & Elger, G. (2021). Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. Reliability Engineering & System Safety, 215, 107864.