Research Annals of Industrial and Systems Engineering

www.raise.reapress.com

Research Expansion Alliance

Res. Ann. Ind. Syst. Eng. Vol. 3, No. 1 (2026) 1–11.

Paper Type: Original Article

Evaluating the Performance of Machine Learning Algorithms in Predicting Industrial Equipment Maintenance Costs

Mohammad Mahdi Masoumian[1,*], Ali Cheraghalikhani[2], Ali Hossein Mirzaei Beni[2]

[1] Department of Industrial Engineering, Khajeh Nasir Toosi University of Technology, Tehran, Iran; mahdi.masoumian@email.kntu.ac.ir.

[2] Department of Industrial Engineering, Tafresh University, Business Model Simulation and Design Laboratory, Central Laboratory, Tafresh University, Tafresh, Iran; cheraghalikhani@tafreshu.ac.ir; hoseinmirzaei@tafreshu.ac.ir.

Citation:

Received: 16 September 2025 *Revised: 11 November 2025* *Accepted: 15 January 2026*	Masoumian, M. M., Cheraghalikhani, A., & Mirzaei Beni, A. H. (2026). Evaluating the performance of machine learning algorithms in predicting industrial equipment maintenance costs. *Research Annals of Industrial and Systems Engineering*, *3*(1), 1-11.

Abstract

Effective maintenance planning for agricultural machinery requires a reliable estimation of future repair and operating costs. Tractors are among the most important capital assets in agricultural operations, and their repair, oil, and fuel costs tend to increase as cumulative operating hours grow. Accurate prediction of these cost components can support maintenance budgeting, economic life assessment, and replacement-related decision-making. This study applies a polynomial regression-based machine learning approach to predict indices of agricultural tractor maintenance costs. The dependent variables include the cumulative repair cost index, cumulative oil cost index, and cumulative fuel cost index, while cumulative operating hours are used as the main explanatory variable. Cost values are standardized using cumulative cost indices to reduce the effects of inflation, differences in tractor purchase prices, and variations across tractor types. The predictive performance of second-degree and third-degree polynomial models and an exponential model is evaluated using goodness-of-fit and error measures. The results indicate that the third-degree polynomial regression model yields the lowest prediction error among the cost indices studied. The findings suggest that polynomial regression can provide a simple, interpretable, and practically useful tool for forecasting tractor maintenance cost trajectories and supporting machinery management decisions.

Keywords: Agricultural tractors, Maintenance cost prediction, Cumulative cost index, Polynomial regression, Machine learning, Repair and maintenance costs.

1 | Introduction

Agricultural tractors are essential assets in modern farming systems, and their availability directly affects the timeliness and efficiency of agricultural operations. Because tractors are capital-intensive machines, their

Corresponding Author: mahdi.masoumian@email.kntu.ac.ir

https://doi.org/10.22105/raise.v3i1.80

operating and maintenance costs constitute a significant share of total machinery ownership and use costs. As tractors age and their cumulative operating hours increase, repair needs, lubrication requirements, fuel consumption patterns, and downtime risks may also increase. Therefore, predicting future maintenance-related costs is an important requirement for machinery budgeting, maintenance planning, and replacement-related decision-making.

Repair and maintenance cost estimation has traditionally been studied in agricultural machinery management through empirical and regression-based cost models. Many classical studies present accumulated repair and maintenance costs as a percentage of the machine's purchase or list price and relate these normalized costs to cumulative use, working hours, or age. Such models are useful because they enable comparison of machines with different acquisition prices, operating conditions, and service lives. Previous studies on tractors and other agricultural machines have commonly evaluated linear, logarithmic, exponential, power, and polynomial regression models for predicting accumulated repair and maintenance costs.

In tractor-specific studies, cumulative operating hours have frequently been identified as one of the most important predictors of repair and maintenance cost. Studies on MF285, John Deere, Massey Ferguson, Universal, and other tractor models have shown that nonlinear models, particularly power and polynomial functions, often provide a strong fit to accumulated cost data. Power models are commonly preferred because of their simplicity and consistency with classical machinery cost equations. However, polynomial models can also provide accurate predictions, especially when maintenance costs grow more rapidly at higher cumulative use levels.

More recently, machine learning methods have been introduced into the prediction of maintenance costs for agricultural machinery. Artificial neural networks, including radial basis function networks, have demonstrated high predictive accuracy for components of tractor repair and maintenance costs. These methods can capture complex nonlinear relationships among cost variables, operating hours, and machine characteristics. However, neural network models are often less transparent than regression-based models and may require larger datasets, additional preprocessing, and more computational effort. For practical machinery management, especially in contexts where simple and interpretable decision-support tools are needed, regression-based machine learning approaches remain useful.

Among regression-based methods, polynomial regression provides a practical balance between predictive capability and interpretability. Polynomial regression can model nonlinear cost trajectories while preserving a clear mathematical form that can be implemented in spreadsheets or simple software tools. It is particularly useful for farm managers and maintenance planners who need understandable cost curves rather than black-box predictions. In addition, polynomial regression can be applied to standardized cost indices, enabling separate analysis of different cost components.

In this study, a polynomial regression-based machine learning approach is applied to predict indices of agricultural tractor maintenance costs. The study focuses on three cumulative cost indices: the cumulative repair cost index, the cumulative oil cost index, and the cumulative fuel cost index. These indices are used to standardize cost values and reduce the influence of inflation, differences in tractor purchase prices, and variations in operating context. Cumulative operating hours are used as the main explanatory variable because they represent the actual physical use of the tractor more directly than calendar age alone.

The main objective of this paper is to evaluate whether polynomial regression can provide an accurate and interpretable model for predicting tractor maintenance cost indices. Specifically, the study compares second- and third-degree polynomial and exponential models across different cost indices and evaluates their predictive performance using error and goodness-of-fit measures. The study's contribution is modest but practical: it applies a simple polynomial regression-based approach to multiple tractor cost indices. It discusses its usefulness for maintenance budgeting and for decision-making related to replacements.

The remainder of this paper is organized as follows. Section 2 reviews previous studies on agricultural machinery repair and maintenance cost modeling, tractor maintenance cost prediction, and machine learning applications in agricultural machinery maintenance. Section 3 introduces the polynomial regression model used in the study. Section 4 describes the data preparation procedure and the construction of cumulative cost indices. Section 5 presents the model evaluation criteria. Section 6 reports the case study results and compares the predictive performance of the selected models. Section 7 discusses the findings and their practical implications. Finally, Section 8 concludes the paper.

2 | Literature Review

2.1 | Repair and Maintenance Cost Models for Agricultural Machinery

Repair and maintenance cost modeling has long been an important topic in agricultural machinery management because decisions about machinery replacement, budgeting, and maintenance planning depend strongly on the ability to estimate future cost behavior. Classical machinery cost studies commonly express accumulated repair and maintenance costs as a percentage of purchase price, list price, or replacement value, and relate these normalized costs to cumulative operating hours, age, or accumulated use. This approach allows the costs of machines with different sizes, ages, and acquisition prices to be compared within a common analytical framework.

Early empirical studies in agricultural machinery economics established the foundation for this approach by modeling accumulated repair costs as a function of standardized use. These models later influenced machinery management standards and cost estimation procedures. In this tradition, repair costs are usually assumed to increase nonlinearly over the machine's life, starting slowly in the early years and increasing more rapidly as hours of use and machine age accumulate. Power-type relationships have therefore been widely used, although linear, logarithmic, exponential, and polynomial forms have also been evaluated in different contexts [1–3].

Rotz [3] proposed a standard model for repair costs of agricultural machinery, in which repair costs were related to purchase price and accumulated use. The study emphasized that repair cost equations should not be treated as fixed in the abstract, as operating conditions, maintenance quality, operator behavior, machine type, and local economic factors influence actual repair costs. As a result, many later studies attempted to recalibrate repair cost equations for specific countries, tractor models, and operating environments.

A common finding in this literature is that accumulated repair and maintenance costs are more meaningful when normalized by machine purchase price. This normalization reduces the effect of differences in tractor size, brand, age, and market price. It is also consistent with engineering economic analysis, where accumulated repair costs are often compared with machine ownership and replacement costs. Therefore, cost indices or percentage-based cost measures provide a practical basis for decision-making on machinery replacement and maintenance.

2.2 | Regression-Based Tractor Maintenance Cost Prediction Studies

Several tractor-specific studies have used regression analysis to predict repair and maintenance costs as a function of accumulated working hours or tractor age. Khoub Bakht et al. [4] investigated repair and maintenance cost models for MF285 tractors in central Iran. They analyzed 102 tractors and compared linear, logarithmic, polynomial, power, and exponential models. Their results showed that both polynomial and power models fitted the data very well. Although the polynomial model yielded a very high coefficient of determination, the power model was recommended for its simplicity and consistency with earlier ASAE-type machinery cost equations. This study is particularly relevant because it confirms that accumulated tractor repair and maintenance costs can be accurately modeled using simple nonlinear regression.

Keshavarzpour [5] studied repair and maintenance costs of John Deere 3140 tractors using long-term records from an Iranian agribusiness company. The study modeled accumulated repair and maintenance costs as a percentage of the initial purchase price against accumulated usage hours. The results suggested that a power model was appropriate for earlier service life, while a second-degree polynomial model provided a better fit at higher accumulated hours. This finding is important because it indicates that polynomial functions may be particularly useful when cost growth accelerates in the later stages of a tractor's life.

Khodabakhshian and Shakeri [6] examined repair and maintenance costs of three tractor models operating under a preventive maintenance program. Their study compared several regression forms, including linear, logarithmic, polynomial, exponential, and power models. The results indicated that power models generally provided reliable predictions of accumulated repair and maintenance costs across the tractors studied. The study also highlighted the role of preventive maintenance in stabilizing cost behavior, while still confirming the importance of accumulated working hours as the main explanatory variable.

Rashidi and Ranjbar [7] developed prediction models for Universal 650 tractors based on accumulated usage hours. Their results again supported the usefulness of nonlinear models. In particular, the study found that a power model was suitable for tractors with fewer accumulated hours, whereas a polynomial model was more appropriate for tractors with longer service lives. It supports the argument that polynomial regression can capture the increasing marginal cost at higher operating hours.

More recent evidence from other regions also supports the localization of tractor repair and maintenance cost models. Dahab et al. [8] estimated repair and maintenance costs for two sizes of agricultural tractors in Northern Sudan. They considered accumulated hours of use and tractor age and compared several regression models. Their results showed that power models provided a strong fit, but they also noted that local cost levels differed from those in models developed in the United States and Europe. Such differences were attributed to spare part prices, labor costs, maintenance practices, and the use of non-genuine spare parts. Therefore, the study reinforces the need for locally calibrated models rather than direct adoption of generic equations.

Overall, tractor-focused regression studies show that accumulated operating hours and age are the dominant predictors of repair and maintenance cost. They also show that power and polynomial functions often outperform simple linear, logarithmic, and exponential forms. However, most of these studies focus on total accumulated repair and maintenance costs rather than separately modeling cost components such as repairs, oil, and fuel.

2.3 | Machine Learning Approaches for Agricultural Machinery Maintenance Cost Prediction

Although regression models dominate the literature on tractor repair and maintenance costs, machine learning methods have also been applied to predict agricultural machinery costs. Rohani et al. [9] used artificial neural networks to predict tractor repair and maintenance costs. Their study examined cost components and cumulative cost indices and showed that artificial neural network models can accurately predict tractor repair and maintenance costs. The results indicated that the back-propagation algorithm with a declining learning-rate factor performed better than basic back-propagation. The study concluded that artificial neural networks can be a promising tool for predicting tractor cost components.

In a follow-up study, Rohani [10] evaluated radial basis function neural networks for predicting tractor repair and maintenance costs. The study reported strong predictive performance for cumulative cost indices, including repair, oil, fuel, and total. These results show that neural network models can capture nonlinear relationships between cumulative hours and cost components. However, neural networks require more data preparation, parameter tuning, and computational effort than traditional regression models. They are also less transparent for managers who need simple equations for budgeting, planning, and replacement-related decisions.

Numsong et al. [11] extended machine learning-based repair and maintenance cost estimation to rice combine harvesters. Their study proposed an artificial neural network-based model to estimate repair and maintenance costs for locally made rice combine harvesters and reported high estimation accuracy. Although the application is not tractor-specific, it underscores the growing interest in machine learning methods for estimating the costs of agricultural machinery.

The available machine learning studies suggest that artificial neural network and radial basis function neural network models may achieve high predictive accuracy, especially when multiple cost components and nonlinear relationships are involved. Nevertheless, these models are often less interpretable than polynomial or power regression models. For practical farm machinery management, interpretability is an important requirement because managers, farmers, and maintenance planners often need transparent cost curves that can be implemented in spreadsheets or simple decision-support tools.

2.4 | Research Gap and Contribution of the Present Study

Previous studies provide a strong empirical basis for predicting tractor and agricultural machinery repair and maintenance costs using regression and machine learning models. Classical and regional tractor studies have shown that accumulated repair and maintenance costs, often expressed as a percentage of purchase price, can be predicted with high accuracy using cumulative operating hours and tractor age. Power models have been widely used and often recommended because of their simplicity and consistency with traditional machinery cost equations. At the same time, several studies have shown that polynomial regression can provide an excellent fit, especially when cost growth accelerates at higher accumulated hours.

However, most regression-based studies focus on a single aggregate repair and maintenance cost measure and do not explicitly model separate cumulative indices for repair, oil, and fuel costs within the same framework. This argument is a relevant limitation because different cost components may follow different patterns over the tractor's life. Repair costs may increase due to aging and component wear, oil costs may follow service routines, and fuel-related costs may be affected by usage intensity and operating conditions. In addition, machine learning methods such as artificial neural networks and radial basis function neural network models provide high predictive accuracy but are less transparent and more difficult for practitioners to implement.

The present study modestly addresses this gap by applying polynomial regression to predict cumulative repair, oil, and fuel cost indices for agricultural tractors as functions of cumulative operating hours. The study does not introduce a new algorithm; rather, it uses a simple and interpretable machine-learning-based regression approach to support maintenance cost forecasting. By modeling separate cost indices, the proposed approach can help maintenance planners and machinery managers estimate future cost trajectories, support maintenance budgeting, and provide preliminary information for replacement-related decision-making.

Table 1 summarizes the selected previous studies.

Table 1. The summary of selected previous studies.

Ref.	Machinery/ Application	Data/Sample	Method	Cost Variable	Key Finding	Relevance to the Present Study
[1]	Tractors and combines	Farm repair cost records	Empirical regression/ power-type models	Accumulated repair cost as a percentage of list price	Established early mathematical repair cost relationships based on accumulated use	Provides a classical foundation for normalized repair cost modeling
[3]	Agricultural machinery	Machinery cost data	Standard repair cost model	Annual and accumulated repair cost	Related repair cost to purchase price and accumulated use	Supports the use of normalized cost functions
[4]	MF285 tractors, Iran	102 tractors	Linear, logarithmic, polynomial, power, exponential regression	Accumulated R&M cost as a percentage of list price	Polynomial and power models fitted the data very well; the power model was recommended for simplicity	Highly relevant; shows the suitability of nonlinear regression for Iranian tractor cost data

Table 1. Continued.

Ref.	Machinery/ Application	Data/Sample	Method	Cost Variable	Key Finding	Relevance to the Present Study
[5]	John Deere 3140 tractors, Iran	15 tractors, long-term records	Power and polynomial regression	Accumulated R&M cost as a percentage of purchase price	Power model suited early life; polynomial model suited longer service life	Supports polynomial modeling for high accumulated hours
[6]	JD-3140, MF-285, JD-3350 tractors	Farm records under preventive maintenance	Linear, logarithmic, polynomial, exponential, power regression	Accumulated R&M cost	Power models performed well under preventive maintenance conditions	Links cost prediction with maintenance policy
[7]	Universal 650 tractors	Usage and cost records	Power and polynomial regression	Accumulated R&M cost as a percentage of purchase price	The polynomial model better captured later-life cost behavior	Directly supports the use of polynomial regression
[8]	MF tractors, Sudan	44 tractors	Linear, logarithmic, polynomial, exponential, power regression	Accumulated R&M cost versus hours and age	Power models showed a strong fit; local cost models differed from foreign equations.	Shows the importance of local calibration
[9]	Tractors, Iran	Historical tractor cost data	Multilayer perceptron ANN	Repair, oil, fuel, and total cost indices	ANN predicted cost components accurately; BDLRF performed best	Demonstrates ML potential for cost-index prediction
[10]	Tractors, Iran	Tractor cost-index data	RBF neural network	Repair, oil, fuel, and total cost indices	RBFNN showed high predictive accuracy	Provides ML benchmark but with lower interpretability
[11]	Rice combine harvesters	Harvester R&M cost data	ANN	Estimated R&M cost	ANN achieved high estimation accuracy	Extends ML-based cost estimation to other agricultural machinery

3 | Polynomial Regression Model

In supervised machine learning, regression models are used when the target variable is numerical. The objective is to learn a mathematical relationship between one or more independent variables and a continuous dependent variable. In this study, the independent variable is cumulative operating hours, and the dependent variables are cumulative maintenance cost indices.

A simple linear regression model may not be sufficient when the relationship between tractor operating hours and cumulative cost is nonlinear. Maintenance costs often increase slowly in the early stages of a machine's life and then accelerate as components wear out and failures become more frequent. Polynomial regression is a practical method for modeling such nonlinear behavior while keeping the model mathematically simple and interpretable.

In polynomial regression, the independent variable is transformed into higher-order terms. For example, a third-degree polynomial regression model can be written as *Eq. (1)*:

$$y = \beta_0 + \beta_1 . x + \beta_2 . x^2 + \beta_3 . x^3 + \varepsilon \quad \textbf{(1)}$$

where y represents the maintenance cost index, x represents cumulative operating hours, β_0 is the intercept, β_1, β_2, and β_3 are model coefficients, and ε is the error term.

In this study, polynomial regression is used to estimate the relationship between cumulative operating hours and each maintenance cost index. Second-degree and third-degree polynomial models are compared with an exponential model. The objective is not to introduce a new algorithm, but to evaluate whether polynomial regression can provide an accurate and interpretable model for forecasting tractor maintenance cost indices.

4 | Data Preparation Methodology

Raw maintenance cost data cannot usually be used directly in predictive modeling. Several preparation steps are required before model estimation. First, cost values must be checked for consistency and adjusted when necessary. Second, cost values measured at different times may be affected by inflation and price changes. Third, tractors may differ in purchase price, model, age, operating conditions, and usage intensity. These factors can make the direct comparison of raw costs misleading.

To address these issues, this study uses cumulative cost indices. The use of indices allows repair, oil, and fuel costs to be standardized relative to the tractor's initial purchase price. This argument makes the cost variables more comparable across tractors and over time. The standardized cost indices considered in this study are:

I. Cumulative repair cost index (CCI_{repair}).

II. Cumulative oil cost index (CCI_{oil}).

III. Cumulative fuel cost index (CCI_{fuel}).

Cumulative operating hours are used as the main explanatory variable. Operating hours provide a practical representation of tractor use because they reflect the machine's physical workload more directly than calendar age alone. In this study, cumulative operating hours were estimated based on the frequency of monthly engine oil changes. Although this approach provides a useful proxy for tractor use, it should be considered an approximation and is noted as one of the study's limitations.

Before model fitting, the dataset was split into training and test sets. The training subset was used to estimate model parameters, and the testing subset was used to evaluate predictive performance. In the original implementation, 75% of the data were used for model training and 25% for testing. This procedure provides a basic method for evaluating how well the model predicts observations not used directly in parameter estimation.

5 | Model Evaluation and Error Metrics

To evaluate the predictive performance of the proposed models, the dataset was split into training and test sets. The training subset was used to estimate model parameters, while the testing subset was used to assess the models' ability to predict unseen observations. This procedure provides a practical basis for comparing different regression specifications and identifying the model with the most reliable predictive performance.

In this study, model performance was evaluated using three common regression metrics: Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2).R^2). MAPE expresses the average prediction error as a percentage of the actual value and is useful for comparing errors across cost indices with different scales. RMSE measures the average magnitude of prediction errors and gives greater weight to larger deviations. The coefficient of determination (R^2) measures the proportion of variance in the observed data that the regression model explains.

Let y_i denote the actual observed value of the cost index for observation i, $\hat{y}_i$ denote the predicted value obtained from the model, $\bar{y}$ denote the mean of the observed values, and (n) denotes the number of observations. The evaluation metrics are defined as *Eq. (2)-(4)*:

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}. \tag{4}$$

Lower values of MAPE and RMSE indicate better predictive accuracy, while higher values of R^2 indicate better goodness-of-fit. However, R^2 alone is not sufficient for model selection, especially when nonlinear or higher-degree polynomial models are considered. A model may achieve a high R^2 by fitting the training data very closely, but this does not necessarily guarantee strong generalization performance. Therefore, the reported prediction error was also considered when comparing candidate models.

6 | Case Study and Results

The case study focuses on predicting indices of agricultural tractor maintenance costs. The cost components considered in this study include repair, oil, fuel, and total maintenance costs. Because tractors may differ in purchase price, operating conditions, and cost levels, the raw cost values were standardized before model estimation. This standardization was carried out using cumulative cost indices, which express cumulative costs relative to the tractor's initial purchase price. For each cost component c, the cumulative cost index at time t can be represented as in *Eq. (5)*:

where $CCI_{c,t}$ is the cumulative cost index for cost component c at time t, $C_{c,k}$ is the cost of component c in

$$CCI_{c,t} = \frac{\sum_{k=1}^{t} C_{c,k}}{P_0}, \tag{5}$$

period k, and P_0 is the initial purchase price of the tractor. In this study, three main indices were considered: the cumulative repair cost index (CCI_{repair}), cumulative oil cost index (CCI_{oil}), and cumulative fuel cost index (CCI_{fuel}). Cumulative operating hours were used as the primary independent variable because they reflect the tractor's actual use more directly than calendar age alone. The regression models were estimated separately for each cost index. Second-degree and third-degree polynomial and exponential specifications were compared to identify the most suitable functional form for predicting each maintenance cost index.

6.1 | Model Comparison

Table 2. Predictive performance of regression models for tractor maintenance cost indices.

Cost Index	Model	β_1	β_2	β_3	R^2	MAPE/Reported Percentage Error (%)
CCI_{repair}	Second-degree polynomial	0.398	0.001	—	0.998	27.37
CCI_{repair}	Third-degree polynomial	0.046	0.005	1.1×10^{-5}	0.999	2.94
CCI_{repair}	Exponential	3.993	0.018	—	0.906	62.00
CCI_{oil}	Second-degree polynomial	0.054	0.00028	—	0.999	4.02
CCI_{oil}	Third-degree polynomial	0.047	0.00028	3.0×10^{-7}	0.999	2.89
CCI_{oil}	Exponential	1.169	0.014	—	0.848	39.71
CCI_{fuel}	Second-degree polynomial	0.029	0.00030	—	0.999	8.95
CCI_{fuel}	Third-degree polynomial	0.041	0.00011	5.2×10^{-7}	0.999	4.89
CCI_{fuel}	Exponential	1.009	0.015	—	0.902	28.00

The results presented in *Table 2* show that the third-degree polynomial regression model achieved the lowest reported percentage error for all three cost indices. For CCI_{repair}The third-degree polynomial model reduced the error to 2.94%, while the second-degree polynomial and exponential models produced errors of 27.37% and 62.00%, respectively. It indicates that the repair cost index follows a strongly nonlinear pattern and that the third-degree polynomial term improves the model's ability to capture accelerated cost growth at higher cumulative operating hours.

For CCI_{oil}, both second-degree and third-degree polynomial models achieved very high R^2 values. However, the third-degree polynomial model produced a lower reported error of 2.89%, compared with 4.02% for the second-degree model and 39.71% for the exponential model. It suggests that oil-related cumulative costs can be predicted with relatively high accuracy using polynomial regression. The improvement from the second-degree to the third-degree model is smaller than for repair costs, which may indicate that oil cost behavior is more regular and more closely related to scheduled service intervals.

For CCI_{fuel}The third-degree polynomial model again achieved the lowest reported error, at 4.89%. The second-degree polynomial model also showed a high R^2, but its reported error was higher at 8.95%. The exponential model performed less accurately than the polynomial models, with an error of 28.00%. This result indicates that the cumulative fuel cost index also follows a nonlinear pattern, which is better represented by polynomial regression than by a simple exponential specification.

Overall, the third-degree polynomial regression model performed best across the three cost indices. Nevertheless, the very high R^2 values should be interpreted with caution. In polynomial regression, increasing the model degree can improve the fit to the available data, but it may also increase the risk of overfitting. Therefore, the third-degree model should be considered appropriate for the present dataset, but its generalization to other tractor fleets, operating conditions, and regions should be validated using additional data.

7 | Discussion

The findings of this study are consistent with previous tractor repair and maintenance cost studies, which have shown that nonlinear regression models are generally more suitable than simple linear models for representing accumulated maintenance costs. Previous research on MF285, John Deere, Universal, and Massey Ferguson tractors has reported that accumulated repair and maintenance costs increase nonlinearly with cumulative operating hours and that power or polynomial models often provide a strong statistical fit. The results of the present study support this general conclusion by showing that polynomial regression can effectively model cumulative repair, oil, and fuel cost indices.

A key finding of this study is that the third-degree polynomial model outperformed the second-degree polynomial and exponential models for all three cost indices. The improvement was particularly clear for the repair cost index, where the third-degree model substantially reduced the reported percentage error. The nature of repair costs may explain this. Unlike fuel and oil costs, which are more directly tied to regular operation and service routines, repair costs may increase more sharply as the tractor ages and components wear. Therefore, a higher-degree polynomial may better represent the accelerating behavior of repair costs at higher cumulative use levels.

The oil cost index showed lower errors across polynomial models, suggesting that oil-related expenses may follow a more predictable trajectory. This argument is reasonable because oil consumption and oil filter replacement are often tied to service schedules and periodic maintenance routines. Fuel cost behavior was also better represented by the third-degree polynomial model. However, it may be affected by additional factors, such as load, field conditions, tractor efficiency, operator behavior, and fuel price variation.

From a practical perspective, the proposed polynomial regression approach has two advantages. First, it is simple and interpretable. The model coefficients can be reported, examined, and implemented in spreadsheet-based decision tools. Second, the use of cumulative cost indices allows different cost categories to be compared on a normalized basis. This argument is useful for maintenance planners and machinery managers because it allows them to identify which cost components grow more rapidly over the tractor's operating life.

At the same time, the results should not be interpreted as a complete replacement of the optimization model. The predicted cost indices can support maintenance budgeting and inform decision-making related to replacements. Still, a full replacement analysis would also require additional variables, such as resale value, downtime costs, productivity losses, interest and inflation rates, and the cost of alternative machines. Therefore, the present model should be viewed as a cost forecasting tool that can support, rather than fully determine, replacement decisions.

This study has several limitations that should be considered when interpreting the results. First, the analysis is based on a specific dataset of tractor maintenance costs, and the estimated coefficients may not be directly generalizable to other tractor models, regions, farming systems, or maintenance regimes. Local spare part prices, labor costs, fuel prices, operator skill, maintenance quality, and field conditions influence repair and operating costs. Therefore, the proposed models should be recalibrated before being applied to other contexts.

Second, cumulative operating hours were used as the main explanatory variable. Although operating hours are a meaningful indicator of tractor use, maintenance costs may also be influenced by other variables, including tractor age, engine power, workload intensity, soil conditions, operator behavior, and the quality of preventive maintenance. Including such variables may improve prediction accuracy in future studies.

Third, this study compares a limited set of functional forms, including second- and third-degree polynomials and exponential models. Although the third-degree polynomial model produced the lowest reported error in the present case, the comparison does not cover a broader set of machine learning methods, such as support vector regression, random forest, gradient boosting, or artificial neural networks. Future research can compare these methods using open or larger datasets while considering both predictive accuracy and interpretability.

Fourth, the very high R^2 values reported for the polynomial models should be interpreted carefully. High goodness-of-fit does not automatically imply strong generalization performance. Future studies should use larger datasets, cross-validation, and independent test samples to evaluate predictive reliability better and reduce the risk of overfitting.

Future research can extend this study in three directions. First, additional cost components, such as downtime, labor, spare parts, and resale value, can be incorporated into the modeling framework. Second, the predicted cumulative cost indices can be integrated with engineering economic models to estimate economic life and replacement timing more rigorously. Third, interpretable machine learning methods can be compared with polynomial regression to determine whether higher predictive accuracy can be achieved without sacrificing practical usability.

8 | Conclusion

This study applied a polynomial regression-based machine learning approach to predict indices of agricultural tractor maintenance costs. The analysis focused on three cumulative cost indices: the cumulative repair cost index, the cumulative oil cost index, and the cumulative fuel cost index. Cumulative operating hours were used as the main explanatory variable because they reflect the tractor's physical use and provide a practical basis for modeling cost growth over time.

The results showed that the third-degree polynomial regression model achieved the lowest reported percentage error for all three cost indices. The improvement was most notable for the repair cost index, suggesting that repair costs may follow a more strongly nonlinear pattern than oil and fuel costs. The results also showed that exponential models were less accurate than polynomial models for the studied indices.

The findings suggest that polynomial regression can provide a simple, interpretable, and practically useful method for forecasting trajectories of tractor maintenance costs. By modeling repair, oil, and fuel costs separately as cumulative indices, the proposed approach provides more detailed information than a single aggregate maintenance cost model. These predictions can support maintenance budgeting, cost monitoring, and preliminary decision-making for replacement. However, the results should be interpreted within the limitations of the study. The model was developed using a specific dataset, and its coefficients may not be directly transferable to other tractor fleets or operating environments. In addition, high R^2 values in polynomial regression may indicate a strong fit to the available data, but should be validated using additional datasets and cross-validation procedures. Future studies can extend this work by incorporating additional explanatory variables, comparing polynomial regression with other machine learning methods, and integrating predicted cost indices into full engineering economic replacement models.

Acknowledgments

The authors appreciate the valuable insights and contributions of the experts who participated in this study.

Funding

This research received no external funding.

Data Availability

The data used in this study are available from the corresponding author upon reasonable request.

References

[1] Bowers, W., & Hunt, D. R. (1970). Application of mathematical formulas to repair cost data. *Transactions of the asae*, *13*(6), 806–809. https://doi.org/10.13031/2013.38725

[2] Fairbanks, G. E., Larson, G. H., & Chung, D. S. (1971). Cost of using farm machinery. *Transactions of the asae*, *14*(1), 98–101. https://doi.org/10.13031/2013.38233

[3] Rotz, C. A. (1987). A standard model for repair costs of agricultural machinery. *Applied engineering in agriculture*, *3*(1), 3–9. https://doi.org/10.13031/2013.26633

[4] Khoub Bakht, G. K., Ahmadi, H., Akram, A., & Karimi, M. (2009). Repair and maintenance cost models for MF285 tractor: A case study in central region of Iran. *Advances in biological research*, *3*(1–2), 19–23. https://www.idosi.org/abr/3(1-2)/4.pdf

[5] Keshavarzpour, F. (2011). Prediction of repair and maintenance costs of John Deere 3140 tractors. *World engineering & applied sciences journal*, *2*(2), 11–14. https://idosi.org/weasj/2(2)11/1.pdf

[6] Khodabakhshian, R., & Shakeri, M. (2011). Prediction of repair and maintenance costs of farm tractors by using of preventive maintenance. *International journal of agriculture sciences*, *3*(1), 39–44. https://doi.org/10.9735/0975-3710.3.1.39-44

[7] Rashidi, M., & Ranjbar, I. (2016). Prediction of repair and maintenance costs of universal 650 tractors based on accumulated usage hours. *Agricultural engineering research journal*, *6*(1), 10–14. https://doi.org/10.5829/idosi.aerj.2016.10.14

[8] Dahab, M. H., Mohamed, A. O.,& Abdalla, N. O. K. (2021). Effect of hours of use and age in years in estimating repair and maintenance costs for two sizes of agricultural tractors in northern Sudan. *Journal of agronomy research*, *4*(2), 1–11. https://doi.org/10.14302/issn.2639-3166.jar-21-3875

[9] Rohani, A., Abbaspour-Fard, M. H., & Abdolahpour, S. (2011). Prediction of tractor repair and maintenance costs using artificial neural network. *Expert systems with applications*, *38*(7), 8999–9007. https://doi.org/10.1016/j.eswa.2011.01.118

[10] Rohani, A. (2013). Prediction of tractor repair and maintenance costs using RBF neural network. *World applied sciences journal*, *28*(12), 1929–1937. https://doi.org/10.5829/idosi.wasj.2013.28.12.2294

[11] Numsong, A., Posom, J., & Chuan-Udom, S. (2023). Artificial neural network-based repair and maintenance cost estimation model for rice combine harvesters. *International journal of agricultural and biological engineering*, *16*(2), 38–47. http://dx.doi.org/10.25165/j.ijabe.20231602.5931