Predictive Maintenance Strategies in Agriculture Using Survival Analysis

Rodrigo Hermont * **Gilberto Reynoso-Meza** **

* *Polytechnical School, Pontifícia Universidade Católica do Paraná (PUCPR), (e-mail: rodrigo.hermont@pucpr.edu.br).*
** *Industrial and Systems Engineering Graduate Program (PPGEPS), Pontifical Catholic University of Paraná (PUCPR), (e-mail: g.reynosomeza@pucpr.br)*
Control Systems Optimizations Laboratory, LOSC, Pontifícia Universidade Católica do Paraná (PUCPR)

Abstract: This study explores the application of survival analysis models—Kaplan-Meier, Cox Proportional Hazards, Aalen's Additive, and Random Survival Forests—to predict failure times and identify critical factors influencing the reliability and efficiency of agricultural machinery. By integrating Directed Acyclic Graphs (DAGs), the research enhances the understanding of causal relationships among operational variables, thus providing a robust framework for predictive maintenance strategies. Utilizing simulated data to ensure confidentiality and integrity, this study demonstrates how modern data science techniques can significantly optimize maintenance schedules and reduce unplanned downtime. The findings underscore the potential of survival models to revolutionize agricultural machinery management by facilitating data-driven decision-making, improving operational efficiency, and reducing maintenance costs. The research not only contributes to agricultural engineering but also offers broader implications for predictive maintenance practices across various industries.

Keywords: Survival analysis; predictive maintenance; agricultural machinery; data science; operational efficiency; Directed Acyclic Graphs.

Resumo: Este estudo explora a aplicação de modelos de análise de sobrevivência—Kaplan-Meier, Cox Proportional Hazards, Aalen's Additive e Random Survival Forests—para prever tempos de falha e identificar fatores críticos que influenciam a confiabilidade e eficiência de maquinários agrícolas. Integrando Diagramas Acíclicos Direcionados (DAGs), a pesquisa aprimora o entendimento das relações causais entre variáveis operacionais, fornecendo assim uma estrutura robusta para estratégias de manutenção preditiva. Utilizando dados simulados para garantir a confidencialidade e integridade, este estudo demonstra como técnicas modernas de ciência de dados podem otimizar significativamente os cronogramas de manutenção e reduzir tempos de inatividade não planejados. Os resultados sublinham o potencial dos modelos de sobrevivência para revolucionar a gestão de maquinários agrícolas ao facilitar a tomada de decisões baseada em dados, melhorando a eficiência operacional e reduzindo custos de manutenção. A pesquisa não apenas contribui para a engenharia agrícola, mas também oferece implicações mais amplas para práticas de manutenção preditiva em diversas indústrias.

Keywords: Análise de sobrevivência; manutenção preditiva; maquinário agrícola; ciência de dados; eficiência operacional; Diagramas Acíclicos Direcionados.

1. INTRODUCTION

Predicting failures in agricultural machinery is essential for improving productivity and reducing maintenance costs. Traditional maintenance schedules often fail to account for the complex and variable operational conditions that contribute to mechanical wear and tear. Machinery in agriculture operates under dynamic conditions, where factors such as load variations and environmental stresses can accelerate wear, increasing the risk of unexpected failures. Predictive maintenance, therefore, offers a proactive solution by identifying potential failures early, improving reliability, and minimizing downtime. This transition from reactive to proactive maintenance results in significant cost savings and productivity gains for operators and the industry.

Survival analysis provides robust methods for predicting "time to failure" events, especially in scenarios with censored data where exact failure times are uncertain. Models like Kaplan-Meier, Cox Proportional Hazards, Aalen's Additive, and Random Survival Forests offer complementary strengths: Kaplan-Meier for non-parametric simplicity, Cox for covariate adjustment, Aalen for dynamic effects, and Random Survival Forests for complex interactions. These models were selected to handle various aspects of operational data, from basic survival probabilities to complex variable interactions.

To better understand the causal relationships between factors, Directed Acyclic Graphs (DAGs) are used to illustrate the interactions among operational variables. This combination of survival models and DAGs allows for more effective identification of risk factors, offering data-driven insights for optimizing maintenance strategies.

The next section reviews the latest published papers on predictive maintenance for agricultural machinery. Section 3 details the methodologies used, section 4 presents the findings, and section 5 outlines future research directions.

2. LITERATURE REVIEW

This section reviews recent advances in predictive models for agricultural machinery, focusing on survival analysis and traditional maintenance methods, with an emphasis on integrating analytical tools to improve failure prediction

The use of survival analysis in machinery failure prediction has gained traction due to its ability to model time-to-event data, particularly with censored data where exact failure times are unknown. Durczak et al. (2022) applied the Kaplan-Meier estimator to assess tractor reliability, providing a foundational approach for optimizing maintenance schedules. Similarly, the Cox Proportional Hazards model has been widely used in industrial settings to account for covariate influences on failure times, offering a more comprehensive understanding of machinery breakdown factors.

Machine learning techniques are also advancing predictive maintenance. Kataev et al. (2022) used neural networks to predict engine failures, demonstrating the efficacy of machine learning with large diagnostic datasets. Wang et al. (2023) applied self-coding neural networks for fault prediction in harvesters, showing high accuracy in forecasting both occurrence and types of failures.

In addition to these models, vibration analysis techniques, such as CEEMDAN and Approximate Entropy, have been employed to enhance Remaining Useful Life (RUL) prediction by detecting subtle mechanical changes before failure Yang et al. (2020). These methods complement survival analysis by providing real-time diagnostics and early warnings.

Traditional approaches often use expert systems and diagnostic parameters. For example, Zotov et al. (2020) integrated expert systems into failure detection, improving accuracy, while Semenenko et al. (2022) explored machinery reliability under extreme conditions, emphasizing how environmental factors impact performance.

The integration of real-time data streams into predictive maintenance frameworks has also evolved. Gabitov et al. (2021) and Purnachand et al. (2021) emphasized the importance of real-time diagnostics, allowing dynamic adjustments to maintenance schedules based on machinery performance, enhancing reliability and reducing downtime.

Recent hybrid models combine machine learning with traditional techniques. Cao et al. (2020) proposed a hybrid framework combining clustering and ontology reasoning, while Azim et al. (2023) and Dixon et al. (2022) used deep learning, including CNNs and LSTMs, to analyze vibration and acoustic signals, further improving failure prediction accuracy.

These studies highlight the growing interest in combining survival analysis and machine learning. While machine learning excels in real-time diagnostics, survival analysis provides a robust framework for time-to-event modeling, together offering improved reliability and efficiency in agricultural machinery.

3. METHODOLOGY

This study employs various survival models to predict failures of agricultural machinery, using simulated data that reflects realistic operational conditions and machine characteristics.

3.1 Survival Models

Kaplan-Meier Model The Kaplan-Meier estimator, a non-parametric approach for estimating survival over time, is particularly useful for analyzing censored data.

$$S(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

where d_i and n_i represent the number of events and subjects at risk at time t_i, respectively.

Cox Proportional Hazards Model The Cox model assesses the impact of several covariates on the time to an event, without assuming a specific survival time distribution.

$$h(t, X) = h_0(t) \exp(\beta X)$$

where $h(t, X)$ is the hazard function, $h_0(t)$ the baseline hazard, and βX the linear predictor.

Aalen's Additive Model Aalen's model offers a flexible approach by allowing the effects of covariates to vary over time, useful for dynamic risk scenarios.

$$dN(t) = Y(t)(\beta_0(t) + \beta(t)^\top X(t))dt + dM(t)$$

where $dN(t)$ is the differential event count, $Y(t)$ the risk set, $\beta(t)$ time-varying coefficients, and $dM(t)$ a martingale increment.

Random Survival Forests This model extends traditional random forests to survival data, capable of managing complex interactions and high dimensionality.

$$\hat{S}(t|X) = \exp\left(-\int_0^t \hat{h}(s|X)ds\right)$$

Here, $\hat{S}(t|X)$ estimates the survival function, with $\hat{h}(s|X)$ representing the hazard function.

Aqui está a redução das subseções solicitadas, mantendo as informações essenciais:

3.2 Data Simulation

The data were simulated to reflect real-world scenarios based on agricultural machinery telemetry. Due to confidentiality agreements, the exact patterns of real data were not disclosed, and the original data were modified and anonymized. The simulated data captured realistic behaviors of parameters from the CAN telemetry network.

The models was performed in R using specialized packages for survival analysis, including **randomForestSRC**, **survival**, **flexsurv**, **SurvMetrics**, and **survminer** for plotting survival curves.

4. RESULTS AND DISCUSSIONS

The analysis of simulated telemetry data reveals key factors driving machine failures, such as machine age, maintenance frequency, and operational conditions, which are modeled in a Directed Acyclic Graph (DAG) to highlight potential causal pathways (Figure 1).

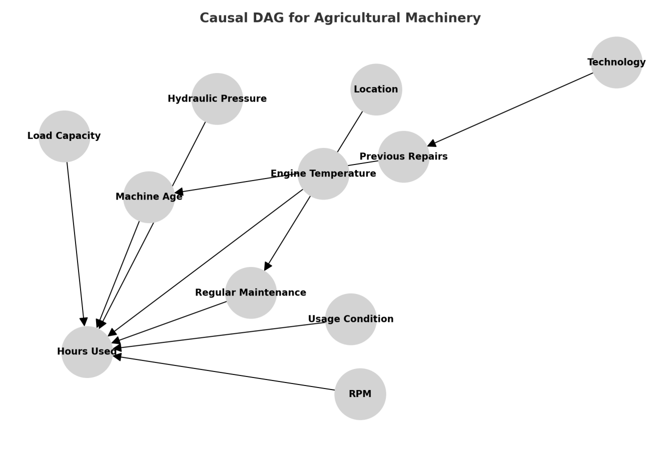

Figure 1. Directed Acyclic Graph for telemetric sensors for causal modelling

The DAG illustrates how *Machine_Age* directly influences *Previous_Repairs* and *Hours_Used*, with *Technology* affecting repair frequency due to improved reliability. Variables like *Regular_Maintenance* and *Load_Capacity* extend machine life, while *Usage_Condition* moderates the wear rate. Operational factors, such as *Engine_Temperature*, *Vibration*, *Hydraulic_Pressure*, and *RPM*, serve as early indicators of failure risks.

These causal pathways provide a solid framework for developing predictive maintenance strategies, enhancing machine lifespan and optimizing resource allocation.

4.1 Kaplan-Meier Survival Estimates

The Kaplan-Meier estimator provided an initial understanding of the survival probabilities of agricultural machinery over time based on operational hours. The survival probabilities were calculated for several time milestones, as shown in Table 1.

Table 1. Survival probabilities at different operational hour milestones using the Kaplan-Meier estimator.

Time (hours)	No. at Risk	No. of Events	Survival Prob.
3000	990	6	0.994
4000	746	131	0.855
5000	143	312	0.366
6000	—	—	—

The survival probability of the machines steadily declines as the operational hours increase, illustrating the expected degradation of machine performance over time. Notably, the survival probability at 3000 hours is very high (99.4%), indicating that most machines are likely to operate without failure up to this point. However, there is a significant decrease in survival probability by 5000 hours, dropping to 36.6%, which underscores the increased likelihood of machine failures as they age and undergo more extensive use.

These results highlight critical timeframes where preventive maintenance could be most beneficial to extend the lifespan of the machinery and prevent unexpected failures.

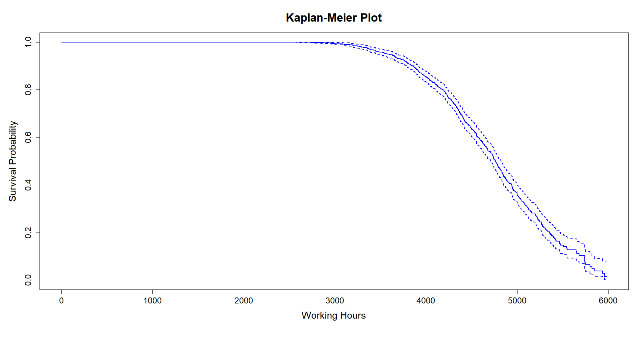

Figure 2. Kaplan-Meyer survival function curve
he survival probabilities were calculated for several time milestones,

The Kaplan-Meier survival analysis indicates a clear trend of decreasing survival probabilities with increased usage time, as depicted in Figure 2. This decline can be associated with various factors such as wear and tear, the need for more frequent maintenance as machines age, and potentially more severe operational conditions. Future interventions could focus on these higher-risk periods to implement more targeted maintenance strategies.

Further insights into the survival of agricultural machinery are gleaned by segmenting the Kaplan-Meier estimates based on chassis models. The median operational hours until failure and confidence intervals for each model are summarized in Table 2.

Table 2. Kaplan-Meier median survival times and confidence intervals for different chassis models.

Chassis Model	No.	Events	Median Hours	95% CI
Chassis 1	89	40	4863	(4646, 5394)
Chassis 10	109	67	4641	(4487, 4896)
Chassis 2	88	46	4712	(4638, 4949)
Chassis 3	109	59	4875	(4554, 5031)
Chassis 4	87	49	4742	(4602, 5023)
Chassis 5	93	47	4820	(4593, 4989)
Chassis 6	90	58	4612	(4400, 4833)
Chassis 7	122	53	5090	(4683, 5331)
Chassis 8	108	63	4766	(4526, 4846)
Chassis 9	105	45	4822	(4724, 5278)

Each chassis model in the dataset represents a different type of agricultural machinery, differentiated by various factors such as design specifications, intended operational load, and maintenance requirements. For instance, some chassis models are designed for heavier loads and more intense operational conditions, while others are built for lighter, more specialized tasks. The differences in median survival times reflect these inherent design and operational characteristics.

The variability in median survival times across different chassis models indicates differential reliability and performance characteristics inherent to each model as shown in Figure 3. Notably, Chassis 7 shows a higher median operational hour indicating potentially better durability or design suited for longer operational life. In contrast, Chassis 6 exhibits a shorter median operational time, suggesting quicker degradation or higher susceptibility to operational stress.

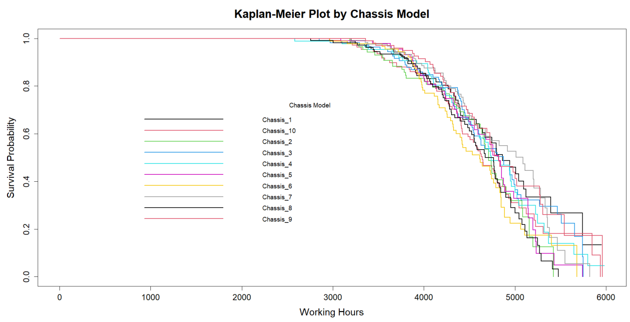

Figure 3. Kaplan-Meyer survival function for different chassis

This analysis assists in identifying specific chassis models that may require more frequent monitoring or preventive maintenance interventions to enhance their operational lifespan.

The Kaplan-Meier survival analysis further extends to the impact of operational locations on the lifespan of agricultural machinery, summarized in Table 3.

The Kaplan-Meier curves indicate relatively similar survival probabilities across different geographical locations, with slight variations in median operational hours. Machines in the East have the highest median survival time, which could suggest better operational conditions or maintenance practices. In contrast, the South has the lowest median survival time, possibly due to more severe environmental conditions or different usage patterns affecting machinery wear and tear.

These insights highlight the importance of considering environmental and operational conditions when planning maintenance schedules and machinery deployment strategies.

4.2 Cox Proportional Hazards estimatives

The Cox Proportional Hazards model offers a comprehensive analysis of how various factors influence the time until failure in agricultural machinery. This model provides a hazard ratio for each variable, quantifying how each one either increases or decreases the risk of machine failure.

The variables such as Regular Maintenance and Technology show significant protective and risk-enhancing effects, respectively (Here for parcimony we only report the significant variables at the model fit estimatives). The model coefficients reveal how each unit increase in predictors alters the hazard of failure, with Regular Maintenance showing a significant reduction in risk (exp(coef) = 2.0973), implying maintenance is crucial in reducing failure risk. Variables like Machine Age and Technology were also significant, indicating older machines and lower tech levels are associated with increased risks.

$$\begin{aligned}
\text{HR}_{\text{Regular Maintenance}} &= 2.0973,\ p < 0.0001 \\
\text{HR}_{\text{Technology}} &= 1.5556,\ p < 0.0001 \\
\text{HR}_{\text{Machine Age}} &= 1.0927,\ p < 0.0001 \\
\text{HR}_{\text{Load Capacity}} &= 0.9354,\ p = 0.0018
\end{aligned}$$

These results were derived from a Cox regression analysis indicating the relative hazard (HR) associated with each variable, showcasing the effects on the time to event (failure) based on operational and machine characteristics.

4.3 Aalen's Additive Model Analysis

The Aalen's additive model are illustrated in Figure 4, was applied to the dataset to explore the dynamic relationships between machine characteristics and their operational lifespan.

Significant findings include:

- Machine Age ($p = 1.08 \times 10^{-17}$) and Regular Maintenance ($p = 1.13 \times 10^{-11}$) both showed significant effects on the time to failure, indicating that older

Table 3. Kaplan-Meier median survival times and confidence intervals by location.

Location	Number	Events	Median Hours	95% CI
East	268	139	4800	(4712, 4946)
North	235	125	4748	(4620, 4945)
South	236	122	4736	(4619, 4868)
West	261	141	4769	(4651, 4875)

machines are more likely to fail, whereas regular maintenance can substantially reduce this risk.

- Technology showed a protective effect ($p = 1.05 \times 10^{-6}$), suggesting that higher levels of technological integration are associated with increased risks, potentially due to more complex systems.
- Load Capacity ($p = 0.0057$) and Previous Repairs ($p = 0.0407$) were also significant, highlighting their roles in affecting machine durability.

These results reinforce the importance of considering both mechanical and operational variables in the management of agricultural machinery.

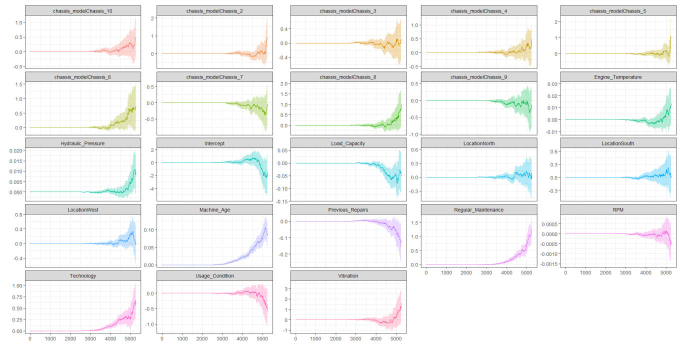

Figure 4. Cumulative Coefficient Plots for Variables in the Aalen Model

The Aalen model graph shows the cumulative effect of each variable on the survival probabilities of agricultural machinery. Each plot represents a different variable, where the $y-$axis depicts the cumulative coefficient, and the $x-$axis represents operational hours.

- Chassis Models: Different chassis models exhibit varying influences on failure rates, with Chassis 6 showing a particularly significant impact.
- Machine Age and Regular Maintenance: Both have substantial positive slopes, indicating a strong effect on increasing failure risk as machines age and a significant reduction in risk with regular maintenance.
- Load Capacity and Previous Repairs: Display negative effects, suggesting that higher load capacities and frequent past repairs increase the risk of future failures.
- Technology: Shows a positive trend, indicating that newer technologies might increase the failure risk, possibly due to the complexity or teething problems of new systems.
- Operational Factors (Hydraulic Pressure, RPM): These factors have smaller, but noticeable effects on failure rates, showing the importance of mechanical conditions in predicting machinery failures.

This visualization helps in understanding how each factor dynamically influences machinery survival over time, emphasizing areas for targeted maintenance and operational adjustments.

4.4 Random Survival Forest Analysis

The Random Survival Forest model was configured with 500 trees, a sample size of 1000, 12 independent variables, and a split rule based on extratrees. The mtry parameter was set to 4, with a target node size of 3, and variable importance assessed using permutation. The model identified 527 unique death times, reflecting the complexity of the dataset and focusing on key factors impacting machinery failure.

This model setup underscores the complexity of the dataset, with a focus on identifying the significant factors impacting machinery failure through permutation importance. Variables such as *Machine Age*, *Regular Maintenance*, and *Technology* were identified as critical.

Harrell's c-index (0.3552821), a generalization of the ROC curve, measures the model's predictive accuracy. A lower prediction error suggests that the model effectively captures the time-dependent nature of the survival data, essential for designing predictive maintenance strategies.

Important remark is that ranger models does not address the time varying coefficients. This apparently is a challenge and many researchers are concerned with this limitation of ensembled forest models.

The Random Survival Forest model provides insights into the variables most influential in predicting machine survival. The analysis reveals that *Machine Age* is the most significant predictor, with an importance score of 0.019734, closely followed by *Regular Maintenance*, which scores 0.019294. *Technology* also plays a notable role, with an importance of 0.007926. Other factors include *Load Capacity* with a score of 0.003760, *Previous Repairs* at 0.002969, and *RPM* scoring 0.001812. These findings underscore the critical impact of maintenance and technology on the operational lifespan of agricultural machinery.

These insights from the Random Survival Forest model provide actionable information for enhancing agricultural machinery management through targeted maintenance interventions.

4.5 Comparative Analysis of Survival Models

In our comprehensive study, we performed a comparative analysis of three different survival models—Kaplan-Meier (KM), Cox Proportional Hazards (Cox), and Random Survival Forest (RF)—to understand their performance in predicting the failure times of agricultural machinery. The results, depicted in the accompanying figure 5, illustrate the failure probabilities over time for each model.

The Kaplan-Meier model shows the highest survival probabilities initially, which steeply decline as operational hours increase. The Cox model, while initially similar to KM, demonstrates a more gradual decline, reflecting its ability to adjust for multiple covariates simultaneously. In contrast, the Random Survival Forest model, which incorporates complex interactions between variables, shows the earliest onset of increased failure probability, suggesting a more sensitive detection of risk factors.

These differences highlight the strengths and limitations of each model: KM for its non-parametric simplicity, Cox for its semi-parametric robustness in handling covariates, and RF for its non-linear handling of high-dimensional data, providing a deeper insight into variable importance and interactions.

Upon analysis, it becomes evident that the Random Survival Forest (RF) model, depicted by the blue curve, predicts higher failure probabilities earlier on compared to the other models. This suggests that the RF model is more sensitive to early risk factors and captures more complex interactions between variables. The RF model's sensitivity to machine failures at earlier operational hours may make it the most suitable for use when prioritizing early detection of machinery wear.

This plot demonstrates how each survival model predicts machinery failure across operational hours, illustrating their distinct predictive capabilities and sensitivities.

This comparative analysis is crucial for selecting the appropriate model based on the specific requirements of predictive maintenance strategies, with each model offering unique benefits depending on the complexity of data and the need for accuracy in failure prediction.

5. CONCLUSION

This study applied survival analysis and Directed Acyclic Graphs (DAGs) to predict and analyze failures in agricultural machinery. Models such as Kaplan-Meier, Cox Proportional Hazards, Aalen's Additive, and Random Survival Forests were utilized, offering insights into machinery lifespan and reliability, with a focus on regular maintenance and machine age.

The comparative analysis of Kaplan-Meier, Cox, and Random Survival Forest models revealed their different strengths. Kaplan-Meier provided a simple, nonparametric view, while Cox adjusted for covariates, offering more precise risk estimates. The Random Survival Forest captured complex interactions but was less interpretable. The choice of model depends on the specific data and objectives of failure prediction.

Limitations include the use of simulated data, which limits real-world validation. Future work should integrate realtime data and machine learning to enhance predictive accuracy. Moreover, DAGs, while useful, require refinement for time-varying systems.

In practice, these findings offer actionable insights for optimizing maintenance schedules in agriculture, extending machinery lifespan, and reducing breakdowns. The approach can also be adapted to other industries, demonstrating broader applicability.

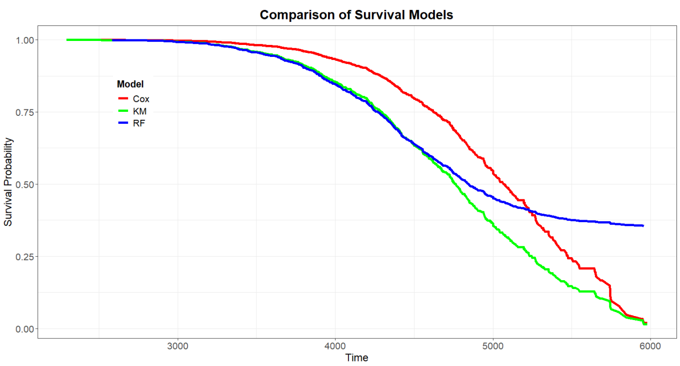

Figure 5. Comparative failure probabilities by model over time/working hours of agricultural machines.

Ultimately, this research marks a significant advancement in agricultural engineering by enhancing maintenance effectiveness and contributing to the sustainability and profitability of agricultural operations worldwide.

ACKNOWLEDGEMENTS

The authors would like to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (Grants number: PQ2-310195/2022-5, UNIV-408164/2021-2) for financial support of this work.

REFERENCES

Azim, M., Afzal, M., Ariz, M., and Samad, A. (2023). Predictive maintenance in rotating machinery using deep learning techniques. *Journal of Fluid Mechanics and Mechanical Design*. doi:10.46610/jfmmd.2023.v05i03.001.

Cao, Q., Samet, A., Zanni-Merk, C., Beuvron, F.D., and Reich, C. (2020). Combining evidential clustering and ontology reasoning for failure prediction in predictive maintenance. doi:10.1016/j.engappai.2020.103730.

Dixon, S., Doster, T., and Emerson, T. (2022). To fail or not to fail: an exploration of machine learning techniques for predictive maintenance. doi:10.1016/j.procs.2022.01.204.

Durczak, K., Selech, J., Ekielski, A., Żelaziński, T., Waleński, M., and Witaszek, K. (2022). Using the kaplan–meier estimator to assess the reliability of agricultural machinery. *Agronomy*. doi:10.3390/agronomy12020428.

Gabitov, I., Insafuddinov, S., Kharisov, D., Gaysin, E., and Farhutdinov, T. (2021). Diagnostic method and device for evaluating and forecasting the technical condition of farm machinery in operation. *Journal of Agricultural Engineering*. doi:10.4081/jae.2021.1158.

Kataev, Y., Zagoruiko, M., Tishaninov, I., and Gradov, E. (2022). Prediction of failures in agricultural machinery engines using digital technologies. *The Agrarian Scientific Journal*. doi:10.28983/asj.y2022i2pp79-82.

Purnachand, K., Shabbeer, M., M, P.R., and Babu, C.M. (2021). Predictive maintenance of machines and industrial equipment. *IEEE International Conference on Communication Systems and Network Technologies*. doi:10.1109/CSNT51715.2021.9509500.

Semenenko, L., Tarasov, O., Vasylenko, S., Cherep, V., and Polishchuk, V. (2022). Mathematical models for assessing the reliability of agricultural machinery, taking into account the influence of military factors in the frontline territory. *Scientific Horizons*. doi:10.48077/scihor.2022.3-4.03.

Wang, X., Zhang, G., Yao, J., Lian, J., and Yang, X. (2023). Fault prediction model of corn grain harvester based on self-coding neural network. *INMATEH Agricultural Engineering*. doi:10.35633/inmateh-64-11.

Yang, J., Chang, Y., Gao, T., and Wang, J. (2020). Failure prediction of rotating machinery based on ceemdanapen feature and ar-ukf model. *Applied Sciences*. doi:10.3390/app10062056.

Zotov, A., Gritsenko, V., and Gazizov, A. (2020). About some possibilities of diagnosing partial failures arising in complex agricultural machinery. *E3S Web of Conferences*. doi:10.1051/e3sconf/202016301024.