ADVANCES IN CYBER-PHYSICAL SYSTEMS
Vol. 9, No. 2, 2024

AGRICULTURE VEHICLES PREDICTIVE MAINTENANCE WITH TELEMETRY, MAINTENANCE HISTORY AND GEOSPATIAL DATA

Anton Shykhmat, Zenoviy Veres

Lviv Polytechnic National University, 12, Bandera Str, Lviv, 79013, Ukraine.
Authors' e-mails*: anton.o.shykhmat@lpnu.ua, zenovii.y.veres@lpnu.ua*

https://doi.org/10.23939/acps2024.02.134

Submitted on 03.10.2024

***Abstract*: Timely detection and prevention of agriculture vehicles malfunctions are key approaches to reducing maintenance costs, as well as updating and replacing equipment, and reducing the cost of growing agricultural crops. In this article an approach for Remaining Useful Life (RUL) prediction that utilizes a combination of telemetry, maintenance, and geospatial data (such as weather and terrain information) as input to a Long Short-Term Memory (LSTM) algorithm has been considered. The results have shown that the models trained on the dataset enriched with geospatial data outperformed the models that relied solely on telemetry and maintenance data, demonstrating the benefits of including location-specific factors. However, the model's RUL prediction applicability for electric and hydraulic systems needs further exploration due to the current dataset limitations.**

***Index Terms*: Agriculture vehicles, GIS, IoT, LSTM, Predictive Maintenance, RUL**

I. INTRODUCTION

The value of agricultural machinery depends on its reliability and operational efficiency, as it is manufactured to withstand thousands of hours of service. However, breakdowns still occur, and many repairs take place on the farm itself. In recent years, the complexity of modern equipment has made this practically impossible. The most important aspect of vehicle management in agriculture is timely and high-quality repair and maintenance. According to a study [1] about 95% of the tractors had at least one malfunction that occurred during fieldwork and made it impossible to continue using the tractor. Still, the frequency of agricultural machinery breakdowns varied among regions. It can also be concluded that only a quarter of the malfunctions can be resolved quickly. All other malfunctions will require more time for repair, as the use of service centers imposes such delays: transportation of agricultural machinery to and from the service center, availability of necessary parts at the service center, availability of resources that can perform maintenance at the service center, or the length of the queue for service appointments. During seasonal work, which is accompanied by serious resource shortages that limit the availability of service workers and spare parts, such delays can be critical for farmers, as this can reduce the amount of harvest that can ultimately be collected. Costs of owning and operating farm machinery represent 35% to 50% of the costs of agricultural production when the land is excluded [2].

In addition, the large-scale aggression of the Russian Federation has significantly worsened working conditions and reduced the possibilities for exporting agricultural products by sea routes, which negatively affects the financial performance of farms. Additionally, it is necessary to note the problem of loss of agricultural machinery due to military actions or theft by Russian Federation military personnel. At present, the main vegetable-growing regions of Ukraine, which produced and sold over 35% of vegetables on an industrial scale, are still partially occupied or are in close proximity to the combat zone. Currently, as a result of military intervention, about 20% of gross commercial vegetable production and 46% of melon production has been lost [3].

II. LITERATURE REVIEW AND PROBLEM STATEMENT

In the industrial use of machines and equipment, maintenance and repair are carried out according to one of three strategies - reactive maintenance, preventive maintenance, or predictive maintenance - with the latter two approaches aimed at reducing the number of unexpected failures [4].

In the case of reactive maintenance, machines, and equipment are operated until a defect or malfunction occurs, which is then corrected. For this strategy, maintenance is not planned, and components are used for as long as possible, which reduces the cost of spare parts but makes machines more vulnerable to becoming inoperable when a malfunction occurs. This approach results in significant unpredictability in fleet management, ultimately driving up maintenance expenses [5]

In the preventive maintenance strategy [6], equipment is replaced before a defect occurs. The usage interval is usually determined in relation to operating hours, based on experience or predetermined maintenance intervals set by manufacturers. Thus, equipment or its components may be replaced before reaching the end of their service life, increasing costs compared to reactive maintenance. On the other hand, preventive maintenance will reduce unpredictable downtime, as maintenance work can be planned before a defect occurs. This maintenance is based on theoretical failure frequency rather than the actual performance of specific equipment, so there may be situations where a malfunction occurs earlier than the next scheduled maintenance, or a malfunction does not occur even if maintenance has not been performed. The frequency of scheduled inspections requires careful calibration. Overly extended intervals between checks can result in increased incident rates, while excessively frequent inspections may lead to unnecessarily high maintenance expenses. [7]

According to the predictive maintenance strategy, repairs are planned based on the condition of the machine or component. Typically, the machine's condition is monitored and analyzed using data from sensors installed in the vehicle, while some approaches additionally use data from enterprise resource planning (ERP) systems to predict work interruptions. Predictive maintenance (PdM) allows for effective planning of maintenance while simultaneously reducing spare parts costs. Unlike preventive maintenance, predictive maintenance is more complex to implement, but it is cheaper to use and requires less maintenance time, as it occurs only, when necessary, that is when a malfunction is predicted.

Recent research [6,7,8] has seen a surge in the application of advanced machine learning techniques to predict Remaining Useful Life (RUL) in various systems. However, these studies primarily focused on analyzing telemetry data received from vehicle built-in sensors or from custom sensors, having a notable gap in the lack of consideration for geospatial data and factors that could potentially play a crucial role in providing impact on the lifecycle of farm machinery.

III. SCOPE OF WORK AND OBJECTIVES

In this paper, the RUL modeling approach for agriculture vehicles predictive maintenance using a combination of telemetry, maintenance history, and geospatial data is proposed. The research addresses the following key areas: combining telemetry, maintenance, and geospatial data (including weather and terrain information) to create a comprehensive dataset for analysis; utilizing Long Short-Term Memory (LSTM) networks to process the integrated data and predict RUL of agricultural equipment; evaluating the performance of models trained on the enriched dataset (including geospatial data) against models using only telemetry and maintenance data. The aim is to make a significant contribution to the ongoing discourse on remote health monitoring of agriculture vehicles, by developing more accurate and precise algorithm for RUL prediction.

IV. MATERIALS AND METHODS

The proposed approach consists of 4 parts: data collection (maintenance, telemetry, weather, and terrain data), data integration and preprocessing (combining single dataset from input data), deep learning modeling (using LSTM algorithm), and RUL prediction. Fig. 1 shows the proposed RUL prediction model.

As part of data collection, the dataset with agriculture vehicles telemetry and maintenance data was received from a cereal farming enterprise based in the US. That grain production farm has a strong interest in the prediction of the maintenance time of machinery to reduce maintenance costs and the cost of crop cultivation in order to maintain the ability to compete with other market players. The datasets were collected in 2015-2017 years and collected data about 4 different tractor models, 30 total tractors that worked in 3 different regions. All tractors passed maintenance prior to providing telemetry data that was analyzed, which means that their health index was 1 or close enough to 1 at the beginning of telemetry collection. The following features were extracted from the provided datasets: maintenance data features (Table 1), telemetry data features (Table 2).

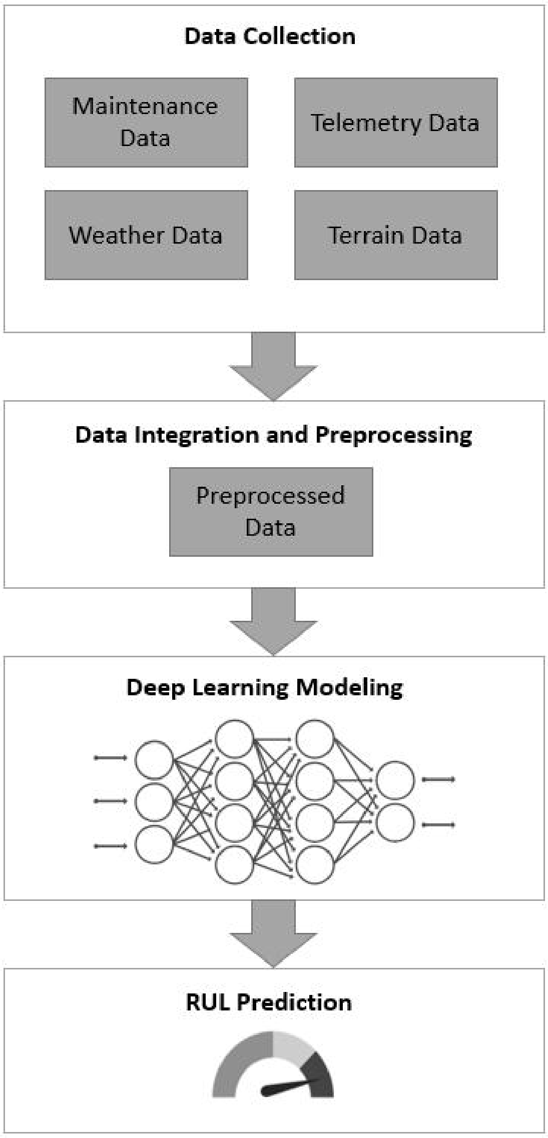

Fig. 1. Proposed RUL prediction model

Geospatial data, such as weather and terrain, was gathered for the areas where the agricultural machinery operated. Weather data was collected using VisualCrossing (https://www.visualcrossing.com) solution and the terrain data used in this study was identified using CalcMaps (https://www.calcmaps.com) solution. The following features were extracted: weather data features (Table 3), terrain data features (Table 4).

Table 1

Maintenance Data Features

Name	Description
Timestamp	The exact date and time when the maintenance entry was recorded
TractorID	A unique identifier for each tractor in the fleet
Model	The specific make and model of the tractor.
Region	The geographical area where the tractor is operating
Age	The number of years since the tractor was manufactured or put into service
Mileage	The total distance traveled by the tractor from the beginning of its operational life (km)
ComponentFailure	Indicates which specific part or system of the tractor failed

Table 2

Telemetry Data Features

Name	Description
Timestamp	The exact date and time when the telemetry data was recorded
TractorID	A unique identifier for each tractor in the fleet
EngineRPM	Indicates how fast the engine is spinning
FuelProductivity	The rate of fuel consumption
EngineLoad	The current engine load as a percentage of maximum capacity
CurrentSpeed	The current speed of the tractor
Power Take Off Load	The amount of power being transferred through the Power Take Off (PTO) system to operate attached implements
Transmission Load	The current stress on the transmission system
Lub Oil Temperature	The temperature of the engine's lubricating oil (Celsius)
Coolant Temperature	The temperature of the engine's coolant (Celsius)
DTC	Diagnostic Trouble Codes, which are standardized codes indicating specific issues detected by the tractor's onboard diagnostics system
GeoLocation	The precise geographical coordinates of the tractor

Table 3

Weather Data Features

Name	Description
Year	The calendar year during which the weather data was recorded
Month	The calendar month during which the weather data was recorded
Day	The calendar day during which the weather data was recorded
AvgTemperature	The average temperature that day (C)
MinTemperature	The minimum temperature that day (C)
MaxTemperature	The maximum temperature that day (C)
Precipitation	The total amount of precipitation that day (mm)
Wind	The average wind speed (Km/h)
Region	The geographical area for which data is collected

Table 4

Terrain Data Features

Name	Description
MeanElevation	The average elevation of the terrain (m)
MinElevation	The minimum elevation of the terrain (m)
MaxElevation	The maximum elevation of the terrain (m)
StdElevation	The standard deviation of elevation of the terrain (m)
MeanSlope	The average slope within the region (degrees)
MinSlope	The minimum slope within the region (degrees)
MaxSlope	The maximum slope within the region (degrees)
StdSlope	The standard deviation of slope within the terrain (degrees)
Region	The geographical area for which data is collected

After datasets have been collected, they were preprocessed and integrated for further usage in predictive maintenance scenario. First of all, telemetry data was combined with maintenance data using TractorID column, and each telemetry record was enriched with Time between failures (TBF) calculated for every failed component using maintenance history. The maintenance dataset contained the history of failures, which is why it was possible to calculate TBF for every component as a difference between the current component failure date and the previous one. The output model contained all columns from the telemetry dataset, columns Model, Region, Age, and Mileage from maintenance dataset and columns EngineTBF, ElecticSystemTBF, HydraulicSystemTBF, and TransmissionTBF. When the first two datasets were merged, the output was enriched with weather and terrain data using Region column and timestamps. It is noteworthy to mention, that telemetry data for specific dates and regions contained the same weather data, as weather data sampled per day, but telemetry is sampled every 30 minutes. Terrain data is a constant as it is stable for years.

Fig. 2 shows the merged datasets with calculated TBF. The merged dataset was split into multiple subsets targeting specific components of each tractor model. The dataset originally contained information about 4 tractor models and 4 components, as a result, 16 distinct datasets were created, each focused on a specific component of a particular tractor model. Within each of these 16 datasets, the data was cleaned to contain only the relevant component's features. The features related to other components, as well as any information related to the other tractor models, were removed. Such data segmentation allowed the development of component-specific predictive models for each tractor model, ensuring the models were trained on the most relevant data for each target component.

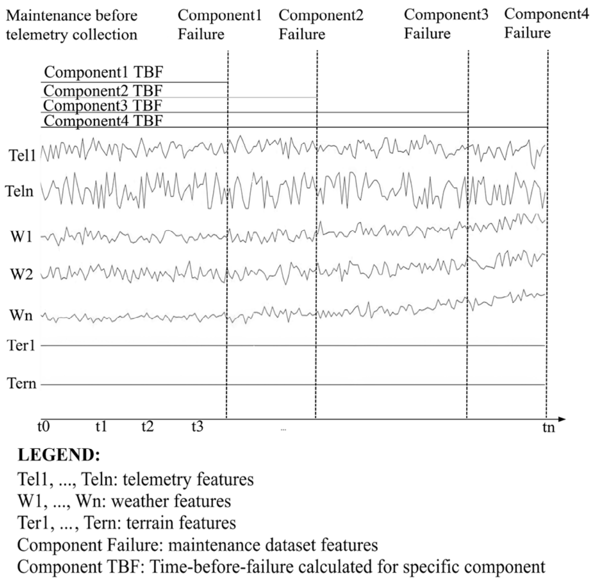

Fig. 2. Example of merged datasets with calculated TBF

The LSTM network excels at capturing sequential patterns within data. To uncover the underlying patterns in sequential data that were combined in the previous step, a two-layer LSTM sub-network is implemented. Additionally, a flattened layer is utilized to reshape the LSTM output for subsequent data integration. The network architecture begins with an input layer that branches into two parallel LSTM paths, each followed by a dropout layer to prevent overfitting. The dropout layers randomly deactivate a portion of neurons during training, which helps reduce overfitting and improves the model's ability to generalize. The parallel LSTM layers are designed to learn temporal dependencies from different perspectives of the sequential data. The outputs from both paths are then consolidated through a flatten layer, which transforms the multi-dimensional data into a format suitable for further processing. Subsequently, two dense layers perform additional feature extraction and transformation before the final output layer produces the network's predictions. This architectural design enables comprehensive sequential pattern recognition while maintaining model generalization through strategic dropout implementation. A structure for the LSTM network described above is shown in a Fig. 3.

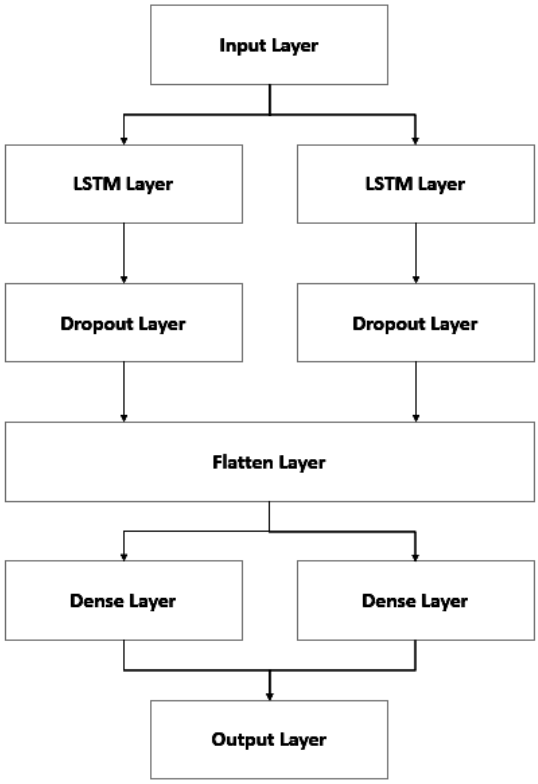

Fig. 3. LSTM network structure

To evaluate the model's accuracy, Mean Absolute Percentage Error (MAPE) and Coefficient of Determination (R-squared) metrics were used. It expresses the forecast error as a percentage of the actual values, providing a relative error metric, and is calculated as the sum of the absolute differences between the predicted (y_i) and actual (x_i) RUL values, divided by the actual (x_i) RUL values, and then divided by the total number of samples (n), and finally multiplied by 100 to express the result as a percentage:

$$\mathrm{MAPE} = \frac{\sum_{i=1}^{n} \left| \frac{x_i - y_i}{x_i} \right|}{n} \cdot 100 . \qquad (1)$$

R-squared gives a measure that represents how close the data is to the fitted regression line. It ranges from 0 to 1, where 0 indicates that the model explains none of the variability in the target variable around its mean, and 1 indicates that the model explains all of the variability in the target variable around its mean. It is calculated using the formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(x_i - y_i\right)^2}{\sum_{i=1}^{n}\left(x_i - x^-\right)^2}, \quad (2)$$

where y_i – predicted RUL, x_i – actual RUL, and x^- is the mean of the actual RUL values that is calculated using the formula:

$$x^- = \frac{\sum_{i=1}^{n} x_i}{n}. \quad (3)$$

V. RESULTS

To evaluate the performance of the RUL prediction models, 32 models were developed and used as train and test sets for LSTM. The first set of 16 LSTM models was trained and evaluated for each specific component across the 4 different tractor models by utilizing the full set of available data, including maintenance and telemetry records, as well as geospatial factors such as weather and terrain information. A second set of 16 models was created to predict the RUL using only the maintenance and telemetry data, without considering the geospatial data. These models allowed to analyze the impact of usage of the geospatial variables on the model's predictive accuracy. The modeling results for each of the models are summarized in Table 5 (note, E means Engine Component, ES – Electric System, HS – Hydraulic System, and T – Transmission).

Table 5

Models Performance Indicators Comparison

Model	Component	MAPE GIS	MAE No GIS	R^2 GIS	R^2 No GIS
1	E	2.78	3.15	0.9099	0.8861
	ES	43.63	45.28	0.5243	0.4811
	HS	30.40	31.53	0.6525	0.6312
	T	2.67	3.03	0.9134	0.8911
2	E	2.61	3.02	0.9153	0.8845
	ES	43.98	45.62	0.5012	0.4886
	HS	29.97	31.11	0.6892	0.6587
	T	2.57	2.94	0.9217	0.9012
3	E	2.46	2.89	0.9281	0.8973
	ES	44.35	45.97	0.4786	0.4462
	HS	29.53	30.68	0.7234	0.7165
	T	2.41	2.84	0.9351	0.9224
4	E	2.28	2.86	0.9415	0.9112
	ES	44.73	46.37	0.4551	0.4329
	HS	29.09	30.25	0.7087	0.6843
	T	2.24	2.66	0.94	0.9234

It can be seen that the LSTM network that incorporated geospatial data provides lower MAPE and higher R-squared values in Remaining Useful Life (RUL) modeling compared to the LSTM network that did not utilize geospatial data. This confirms the hypothesis that incorporating location-specific factors, such as weather patterns, terrain characteristics, and environmental conditions, can enhance the accuracy of predictive maintenance models. The improved performance metrics suggest that geospatial features capture important contextual information about asset degradation patterns that might be missed in traditional time-series-only approaches. The integration of geospatial data not only improves model accuracy but also provides deeper insights into how geographic and environmental variables influence asset deterioration rates, potentially enabling more targeted and efficient maintenance strategies across different operational contexts. However, current dataset can be used only to predict RUL for Engine and Transmission components. Fig. 4 shows RUL prediction accuracy for Machine ID 1 and Engine Component.

RUL Prediction with and without GIS Data for Machine 1 and Engine Component

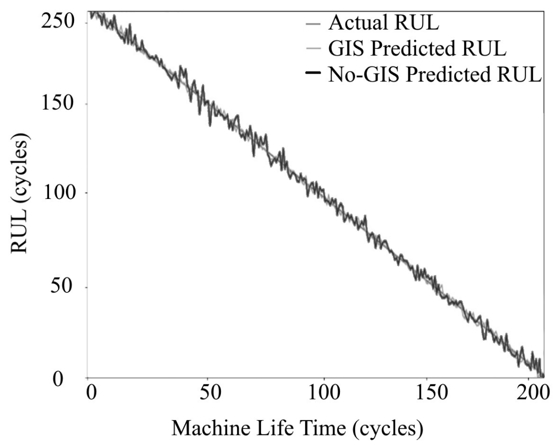

Fig. 4. RUL Prediction with and without GIS Data for Machine ID 1 and Engine Component

RUL Prediction with and without GIS Data for Machine 1 and Electric System Component

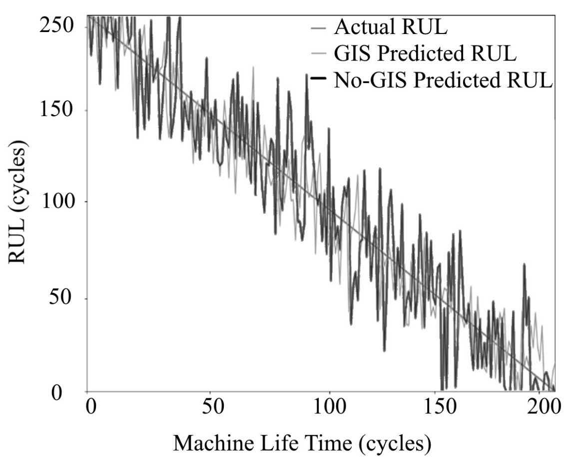

Fig. 5. RUL Prediction with and without GIS Data for Machine ID 1 and Electric System Component

The RUL prediction for the Electric System and Hydraulic System components consistently shows the poorest performance for both the GIS and NO GIS models (Fig. 5). This indicates that data available in telemetry and maintenance datasets may not be sufficient to accurately predict the remaining useful life of the ES and HS components.

The insights gained from this experiment can be used in future research and development efforts in the area of predictive maintenance. It is worth suggesting the need for further investigation of the features to build modeling approaches for accurate forecasting of the RUL of the Electric System and Hydraulic System components. Potentially, this required data can be obtained through the integration of additional data sources, such as CAN-bus or external sensors. It is also important to investigate the impact of geospatial data on the performance of other deep learning models, such as CNN, DBN, RNN, etc., as they can show different results.

VI. CONCLUSION

Existing studies do not consider geospatial data, such as weather and terrain conditions, while modeling predictive maintenance approaches for agriculture vehicles, and use only data received from tractor or maintenance data. In this paper, RUL prediction model using a combination of telemetry, maintenance, and geospatial data was researched. Datasets for experiments were obtained from a cereal farming enterprise based in the US and used for performance evaluation of the proposed approach. Results of the research revealed that the accuracy of the LSTM algorithm for RUL prediction that used weather and terrain information is higher compared to the same algorithm that does not utilize geospatial factors. However, the applicability of the model for RUL prediction of Electric System and Hydraulic System needs further exploration as datasets used for experiment did not contain enough data that can be utilized for prediction, therefore, it is important to identify features that can be suitable for it. Also, it is proposed to evaluate the impact of geospatial data on other deep learning algorithms for predictive maintenance of agriculture vehicles.

References

[1] Paman, Ujang & Khairizal, Khairizal & Sutriana, Selvia. (2022). Investigating Farm Machinery Breakdowns and Service Support System Conditions in Rainfed Rice Areas in Riau Province, Indonesia. *Asian Journal of Agriculture and Rural Development*, *vol. 12, no. 3*, pp. 182-191. DOI: https://doi.org/10.55493/5005.v12i3.4578.

[2] Dahab, Mohamed & Gafar, Montasir & Al-Gailani, Abdul Qader. (2021). Repair and Maintenance Cost Estimation for Two Power Sizes of Agricultural Tractors as Affected by Hours of Use and Age in Years: A Case Study, Dongola Area, Sudan. *Journal of Engineering Research and Reports, vol. 20, no. 10*, pp. 113-121. DOI: https://doi.org/10.9734/jerr/2021/v20i1017395.

[3] Vytoptova V.A. (2024). Study of the state and problems of agriculture in Ukraine in wartime conditions. *Agrarian Innovations, no. 23 (2024),* pp. 210-213. DOI: https://doi.org/10.32848/agrar.innov.2024.23.30.

[4] Dube, Aryaa. (2024). Application of Deep Learning in Predictive Maintenance of Aircraft Engines. *Darpan International Research Analysis*, *vol. 12, no. 3,* pp. 83-100. DOI: https://doi.org/10.36676/dira.v12.i3.58.

[5] Crespo del Castillo, Adolfo & Parlikad, Ajith Kumar. (2024). Dynamic Fleet management: integrating predictive and preventive maintenance with operation workload balance to minimise cost. *Reliability Engineering & System Safety, vol. 249.* DOI: https://doi.org/10.1016/j.ress.2024.110243.

[6] Spinelli, Raffaele & Eliasson, Lars & Magagnotti, Natascia. (2019). Determining the repair and maintenance cost of wood chippers. *Biomass and Bioenergy*, *vol. 122*, pp. 202-210. DOI: 10.1016/j.biombioe.2019.01.024.

[7] A. Bukhsh, Zaharah & Saeed, Aaqib & Stipanovic, Irina & Dorée, André. (2019). Predictive maintenance using tree-based classification techniques: A case of railway switches. *Transportation Research Part C: Emerging Technologies*, *vol. 101,* pp. 35-54. DOI: https://doi.org/10.1016/j.trc.2019.02.001.

[8] Xiao, Maohua & Wang, Weichen & Wang, Kaixin & Zhang, Wei Zhang, Hengtong. (2020). Fault Diagnosis of High-Power Tractor Engine Based on Competitive Multiswarm Cooperative Particle Swarm Optimizer Algorithm. *Shock and Vibration*. DOI: https://doi.org/10.1155/2020/8829257.

[9] Li, Dan & Zheng, Yi & Zhao, Wei. (2019). Fault Analysis System for Agricultural Machinery Based on Big Data. *IEEE Access*, *vol. 7,* pp. 99136-99151. DOI: https://doi.org/10.1109/ACCESS.2019.2928973.

[10] Gouvea, Carlos & Sá, José. (2019). Diagnostic of Failure in Transmission System of Agriculture Tractors Using Predictive Maintenance Based Software. *AgriEngineering*, *vol. 1, no. 1*, pp. 132-144. DOI: https://doi.org/10.3390/agriengineering1010010.

Anton Shykhmat – Solutions Architect at SoftServe Inc. He received his M.S. degree in Applied Mathematics and Informatics at Ivan Franko National University of Lviv, Ukraine, 2012. He focuses on various technologies including cross-platform cloud solutions, security, networks, IoT and computer engineering, and is currently working towards a Ph.D. in Computer Engineering.

Zenoviy Veres, PhD – an assistant professor of the Department of Computerized Automatic Systems of the Institute of Computer Technologies, Automation and Metrology at Lviv Polytechnic National University. In 2015 he received Ph.D. degree in Artificial Intelligence at Lviv Polytechnic National University. He is also a Senior Solutions Architect at SoftServe Inc. His research interests include distributed highly scalable microservice systems, IoT, cloud computing, and artificial intelligence.