JOURNAL OF BASIC SCIENCE AND ENGINEERING

Machine Learning-Based Failure Prognostication of flight Equipment utilizing Hybrid Techniques

Jalajakshi V[a], Dr. Myna A N[b]

[a] Research Scholar, Department of CSE, Navkis College of Engineering, Hassan
[b] Research Supervisor, Department of CSE, Navkis College of Engineering, Hassan

DOI: https://doie.org/10.0618/Jbse.2024131340

Abstract

The air travel sector has a lot of knowledge and general upkeep data that could be utilized to estimate future behavior and produce useful outcomes. This research seeks to use variable selection and information deletion-based machine learning approach for predicting aeroplane system problems. Over the duration of two years, upkeep and inability information for flight systems and equipment have been gathered, and nine input and one outcome value were painstakingly recognized. To increase the accuracy of malfunction count prognostication in 3 levels, a blended model development model is suggested. Among most efficient and unproductive parameters are identified in the first phase using the RF method of feature selection for feature assessment. To remove noisy or inconsistent data, a modified K-means algorithm and naive bayes is used in the stage 2 . SVM, Decision Trees, linear regression, Random Forest and multilayer perceptron such as neural networks and advanced algorithms, respectively, are used to assess the efficacy of the proposed model development model on the upkeep set of data of the appliances. Additionally, in the third stage the designs are assessed using performance measures like the coefficient of correlation, mean absolute error, and root average square error. The findings show that the hybrid feature extraction model is effective in estimating the device failure rate.

Keywords: **Random Forest, Neural networks, Dataset, RF algorithm, Regression**

1. Introduction

The accessibility and dependability of airplane parts has always been crucial factors in air transport. The consistency of aeroplane components and systems will boost with precise fault diagnosis. The overarching upkeep and revamp costs of aircraft structures are determined in part by the planning of regular maintenance repairs. An important component of the total operational cost for flight systems is repairs. Course correction, preventative measures, and preventative analysis are the three primary types of servicing for devices [1]. Corrective and preventative aids in the control of necessary repair procedures and unanticipated failure occurrences, such as hardware and hardware failure. When an aeroplane component breaks down being used, it is either substituted or repaired.

The number of unforeseen repairs can be decreased with maintenance work. To prevent machinery malfunctions or machines and equipment shutdowns, it is incorporated through regular maintenance. In order to avoid unanticipated leisure time and breakdowns that would necessitate repair work, tasks for this type of maintenance are planned.

Repair / Maintenance prognosticating, as the term indicates, makes predictions about possible failure dates based on variables that are assessed while the hardware is in use. By giving the maintenance staff more dependable scheduling options for preventive maintenance, it seeks to intervene with the scheme until flaws manifest [1, 2] and contribute to reducing the number of unforeseen events. It's critical to evaluate reliability before selecting the best maintenance program. A modern tech that is expected to expand in the future is subset of artificial intelligence. Mechanisms for forecasting and prevention, comms, reliability, power management, and other areas all use ml algorithms [3]. The central aspect of machine learning and the judgement system is the data preparation quality.

It controls the information so that it can be used in decision-making. The judgement process is based on equipment performance, malfunction events, and potential forecasts [4]. Data mining is a technique for categorizing and reducing data to understandable information. It extracts the given set from a sea of data and uses various strategies to unearth hidden data. Data analytics is the process of acquiring knowledge from unstructured information [5].

In data science and ML algorithms, which concentrate on the characteristics that are the most pertinent to the desired prognostication, feature extraction is a basic problem [6]. Not all characteristics gathered from of the inspection of a situation are equally important. Available data typically tend to just be inadequate, unaddressed, only partly impactful, or completely meaningless. A few of them might be annoying, pointless, or utterly pointless. The goal of showcase selection is to pick a set of functionality that really is appropriate for a particular duty. This issue is complicated and multifaceted [7]. A unique feature selection technique that utilizes the correlation coefficient clustering method was put forth by Hsu [8]. It concentrated on removing distracting, repetitious, or redundant features.

By eliminating the irrelevant information, it is possible to increase computational cost and feature selection. Data analysis methods aid in improving the database's performance and enhance information mining's precision, thereby increasing its efficiency [8]. Information quality is important for information extraction, identifying discrepancies, and forecasting and analysing data to make decisions [9]. To lower repair and equipment costs and determine equipment reliability, product failure prediction is crucial [1]. Businesses can profit from big data, which can also direct processes in the correct direction. It is crucial to extract useful data from the data set with respect to enhance effeciency in ML advanced techniques.

Whereas preserving the cluster centers tiny, it describes k amount of cluster center and assigns each scenario to the nearest cluster [10]. The set of data was averaged to determine the centroid as part of the K-means algorithm. So every circumstance is linked to a specific set by this method. The goal is to achieve both great similarities inside every cluster and low resemblance between them [11]. It is used for less complicated cluster formation that is more efficient and superior. Numerous studies have been done on practice known and predicting rate of failure. An essential step in the feature representation, data preprocessing has a significant impact on the performance of a machine learning models. Gurbuz and others [9]

In order to comprehend and clear the dataset from noised and to identify the connections between the i/p and o/p characteristics, [9] used a variety of data preprocessing together with feature extraction to sets of sets of data of a Turkish air carrier. This same aerospace firm's experts have noted their list of regulations for developing malfunction notifications to be helpful. To retrieve trends from the device information, a classification method has been used. Deep neural networks could be used to design rate of failure, according to Kutlylowska [12]. The output signals of malfunction regularity were calculated using information from a Polish water utility. The findings suggested that artificial networks might be a viable option for estimating the regularity of issues in water distribution systems Ramos and others.

The author in [13] conducted a study on an item of production machinery to predict system failures and perform preventative scheduled maintenance. In this research, neural network algorithms and ARIMA forecasting methods were effectively especially in comparison. According to the findings, both frameworks performed well at predicting when a defibrator disc would need to be replaced, however the ARIMA performed significantly better at predicting how far apart the discs would be. Trani et al[14] .'s introduction of a machine learning network-based basic method for determining aeroplane fuel usage. Using the information provided in the airplane's achievement handbook, a fuel usage prototype endorsed by a neural net was developed.

Utilizing the data collected first from resonance detection systems to forecast technical glitches, Ming et al. [15] looked into the use of the Ann technique in vibration monitoring. The ANN method was used to assess a mill's faults. The outcomes provided evidence for this methodology's effectiveness. To determine the requirement for replacement parts to be supplanted all through jet engine revamp, Kozik and Sep [16] used ANN prediction. The findings suggested that lean production must be implemented in maintenance, fixing, and overhaul facilities using an estimation approach based on the car's hardtime computation. In order to simulate an artificial convolutional neural network model breakdowns, Altay et al. [17] was using 500+ breakdowns of 85+ aeroplane. The suggested framework generated good correlation prices of prognostication between both the actual and the desired trials of planes. A technique to estimate the service life (RUL) in machines and equipment with bearing surfaces was put forth by Benkedjouh et al. The research teams used back - propagation algorithm and the isometric classification map - based reducing methodology (ISOMAP) for all of this (SVR). A comparison assessment to evaluate the SVM's efficiency in forecasting failure probability was proffered by Moura et al. [19]. SVM regression's effectiveness is contrasted with that of other learning strategies.

The feature extraction of factors in the upkeep data collected from a Turkish airline company is discussed in this study. The suggested system will assist businesses in gathering, extracting, and creating data to enhance periodic maintenance through much more accurate forecasts. By contrasting the outcomes of three methods, this study indicates a hybrid data technique for servicing information and forecasting machine breakdowns counts. In the current study, characteristics are chosen using the method of feature selection (RF heuristic), and data items is removed using an altered K-means method. The MLP machine learning methodology, the Gradient Boosting algorithm, and regression are used to initiate and contrast three different approaches for forecasting machine breakdowns rates. An overall view of the components and research methods is given in the following section, which is followed by the results and findings of the trial.

1.2 Procedure

An aviation business in Istanbul, Turkey, served as that of the setting for the current study case. The Science and technology Programmer upkeep agency's documents were utilized to gather the upkeep information. They covered things like machinery expulsion, maintenance work, operator encounter, device flying hours, as well as other specific instance study-related details.

A. Dataset collection

The Enterprise Resources Planning framework is used to create a programme that gathers data and formats the set of data for machine algorithms assessment. The variables were categorised as input variables, and the failure count was regarded as an outcome variable. The attribute selection RF method estimates the chosen variables to determine which ones contribute the greatest to mistakes. In Figure 1, the advanced project's flow logic is shown. First, the landing gear system's serial-numbered hardware for the raw material was chosen.

They identified their operations and maintenance information. In collaboration with engineers and maintenance crews, characteristics of the failure and upkeep data were determined.

To discover declared malfunction information, every example of a no fault found (NFF) condition was investigated. For a particular time period, the sum of flying hours for each equipment component on various

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

aircraft was calculated. There are nine input variables that were identified as influencing damage to equipment. An resulting value that was estimated was the failure count. The algorithms for machine learning employed in the current study were modelled using these various input variables and one outcome variable (Support vector, and Regression).

B. Parameter choosing and RF algorithm methodology

By eliminating unnecessary and incomplete data information from the original database, selection of features is a methodology for obtaining the necessary characteristics. It involves choosing a small subset of features that can accurately represent all sets of data. The main goal of feature selection is to mine the data to find the fewest number of characteristics necessary to achieve the highest level of accuracy. Data analysis, deep learning, and machine intelligence are all using methods for selecting features. They make models simpler and speed up method execution. Rf is an attribute selection methodology to use for feature weight calculation that randomly selects instances. In 1992, Kira proposed the Rf algorithm [20, 21]. According to how well they can distinguish between nearby models, it forecasts weight values iterative manner.

To discuss classification concerns, the RF methodology was expanded to come to terms with erratic, irrelevant, and incomplete information. RF is an improvement on the old existing hybrid algorithm, which in two-class classification problems is unable to eliminate irrelevant or insufficient features. In order to update weight values, the RF search algorithm one close call for every distinct class and averaged nearly their significance.

C. Methodology for Data Removal and Modified K-Means

In 1967, MacQueen created the K-means algorithm, which is now widespread utilized in a wide range of situations [23]. It enables every data point to belong to a consistent dataset. It has two restriction fields: a preliminary midpoint and a stationary K value. The Euclidean distance is calculated using the length set of criteria. The altered K-means algorithm that Fahad and Alam [10] proposed demonstrated to be less time-consuming but more effective at clustering. The choice of the initial centroid affects how well the clusters turn out. By removing the smallest class value represented in the cluster, the K-means algorithm enables the creation of a new data cluster.

To remove noisy and irrelevant data, Yilmaz et al. [11] proposed a system using an altered K-means algorithm. The adapted Kmeans methodology from [11] was employed in this study, and Algorithm 1's procedure was created. Study on machine learning and data science methodologies has already been conducted recently to examine defect prediction application forms. In order to simulate maintenance data and forecast the failure rate, the neural network, support vector machine, decision tree, random forest, and linear regression algorithms were evaluated in this research.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

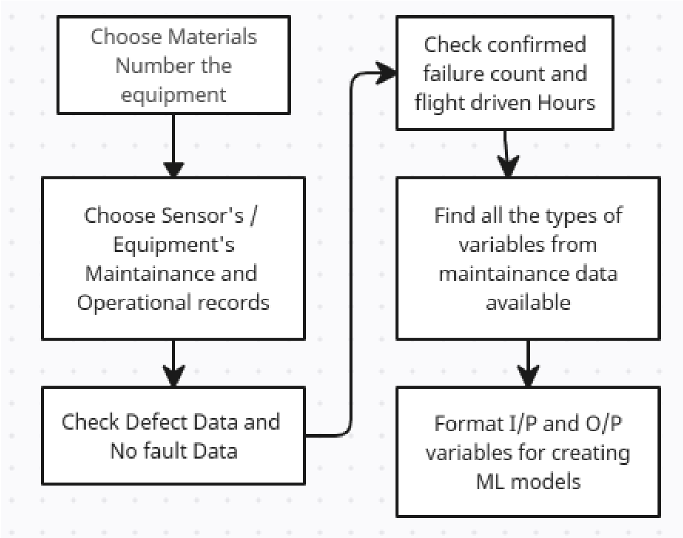

Figure 1: flowchart of the designed Methodology

D. Prognosticating Techniques

Investigation on data mining and artificial intelligence methods has been conducted recently to examine fault prediction application forms. In order to model average long - term and forecast the malfunction count, the neural network, Support vector machine, and regression algorithms were examined in this research.

i. Neural Networks

A mathematical formalism based on an innately inter - connected collection of synthetic neural nets is known as an Artificial Neural Network. The information-based method to arithmetic used by the brain is imitated by ANN [24, 25]. A non - linear statistical information model - based and ML technique is multi - layer perceptron. They can indeed be applied to model incredibly complex multidimensional connections among both data's inputs and its outcomes. With both the aid of schooling, learning, and testing procedures, they also portray any patterns or connections present in the information and aid in the forecasting of target value. A layer is made up of a set number of neurons, which are the neural network's equivalent of cells. A weighing scale is used to link nodes in different layers.

A feed forward method is used by ANN methodology to approximate the weights for the nerve cells in the input values, hidden layer, and output layer [24, 23]. The most common type of backpropagation is used to determine strength training in an Artificial Neural Networking. A supervised learning called back propagation algorithm aims to reduce the discrepancy between both the actual and desired outputs. The weights are changed to disperse the overall error among the different neural network neurons. Through backpropagating and feeding forward, the discrepancy is kept at a low level [17]. Neural networks' multi - layered framework is what gives them their capacity for prediction. An input data, one or more hidden units, and an output neurons are the three layers of a neural net.

The neurons' artificial neuron makes up neural net methodologies [21]. In this research, a back - propagation training methodology was combined with perceptron (MLP) feed-forward neural connections.

ii. SVM :

Cortes and Vapnik published the SVM Classification Models (SVM) method in 1993 [18]. It is a linear classifier applied to issues in classification and regression. The data is classified via a higher dimensional space created by the Svm classifier. A horizontal line divides the two main categories. Identification of the variables is a step in the method training phase [11]. Support Vector Regression (SVR) is a prediction technique that performs linear regression in a similar way to Support Vector Machine. SVR is an useful method that may be used for data analysis, forecasting, and regress issues. It is a supervised learning algorithm.

iii. Linear Regression

A machine learning algorithm called linear regression is one that relies on supervised learning and involves a regression job. It is employed to simulate the linear connection between dependent and independent variables. It assists in establishing the connection between factors and prediction. A prediction algorithm for a computing device, utilizing a linear regression with lowest optimization, was developed by Schuld et al. [19]. The computer intelligence problem of presuming the outcome corresponds to a new data was the main emphasis of its plan. The outcome of the prediction model can be used to other quantum communication processing procedures.

E. Assessment Methods for Evaluation

The effectiveness of each of the algorithms in this study was assessed using the mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (CC) criterion. There are several methods for measuring errors, and they are most frequently used to measure errors. The continuity equation, each adapted from [31], provide the error parameters:

$$\text{Mean} = \frac{1}{N} \sum_{i=1}^{n} |Xi - Yi| \qquad \text{………. I}$$

$$\text{Root Mean Square Error} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Xi - Yi)^2} \qquad \text{….II}$$

$$\text{Variability} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}}, \qquad \text{------III}$$

Xi is the observed value, Yi is the expected value, N represents the number of data, X is the average of the observational data, and Y is the average of the actual observations and anticipated data values.
CC calculates the correlation coefficient, which is the algorithm's estimate of the variance of the observational data.

2. Proposed Techniques

As mentioned in Section 2.1, the 625-line servicing reports from a international aviation firm over a five year period were employed in this investigation. eleven input parameters and one outcome variable

make up the data (failure count). The operational and external characteristics that may affect failure incidence and the amount of operating time before problems occur are the input different factors. Variables like flying hours and the amount of equip removal are examples of input parameters.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

the quantity of errors in planned and impromptu deletions. The variables were described using the corresponding domain categorization, as indicated in Table 1, and these data analyzed and expressed in a format suitable for modelling. The number of equipment failures is the outcome variable. Table 2 offers a sample of the dataset.

Parameter	Description
Aircraft Hours	The aggregate number of hours spent flying for a certain item on various aeroplanes within a chosen time frame
Rem	The quantity of machinery removal during the past 36 months
Planrem	The amount of scheduled machinery removal during the past 36 months
Unplrem	The quantity of unscheduled machinery removals during the last 36 months
Otherrem	The quantity of additional hardware removals in the last 36 months
Faultrem	The quantity of apparatus removal-related errors in the last 36 months
Faultplanrem	The amount of issues with anticipated machinery relocation in the last 36 months
Faultunplanrem	The amount of errors involving unexpected machinery removals in the last 36 months
Saferem	The quantity of material movements done safely during the past 36 months
avgtime	
nofail	The number of equipment failures in the last 24 months
Sensorstime	This calculates the average sensor lifetime

Aircraft Hours	Rem	Plan rem	Unpl rem	Other rem	Fault rem	Fault planrem
327	9	0	8	1	8	0
338.5	11	0	9	2	2	1
287	7	1	9	0	6	1
298	8	2	8	0	8	0
312	9	0	11	1	10	0

Table 2 : Maintenance Repair data

Aircraft Hours	The aggregate number of hours spent flying for a certain item on various aeroplanes within a chosen time frame
Rem	The quantity of machinery removal during the past 36 months
Faultrem	The quantity of apparatus

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

	removal-related errors in the last 36 months
Faultplanrem	The amount of issues with anticipated machinery relocation in the last 36 months
Faultunplanrem	The amount of errors involving unexpected machinery removals in the last 36 months

Table 3: Choose parameters.

These 13 criteria are used to choose characteristics. The analysis's primary focus is on potential breakdown rates. In order to do this, relationships between weighted coefficients and relations were discovered using the feature selection RF technique. The five best beneficial qualities were chosen based on the rated values (Table 3). The processed sets of data typically contain distorted and data inconsistency, which has a detrimental impact on predictions and lowers application performance. In order to improve the performance of the forecast, the noisy and inconsistent data were removed using the revised random forest method. It was created using the Protocol 1 pseudo - code. In this architecture, examples are correctly dispersed among the groups and set centres are originally assigned.

For each group, the record (N = 7) that was farthest from the centre was deleted. The Distance measure like euclidean or distance calculation served as the distance criterion. Figure 2 shows the occurrences that were deleted. The suggested data preprocessing model's syntax removed 93 records from the data, or around 15% of it. From the 625 records in the dataset, that were collected. Two processes make up our suggested hybrid data preprocessing paradigm, as indicated in Figure 3. In the first step, the attribute selection RF method decreased the eleven input qualities to four characteristics. The revised random forest method was used in the second stage to decrease the collection to 525 records. The 525 record acquired sample was used as input for the progression, vector machine, and Neural networs prediction algorithms.

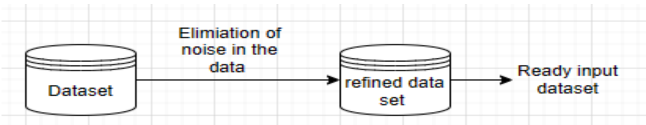

Figure 2 : Model for preprocessing data in the modified random forest approach

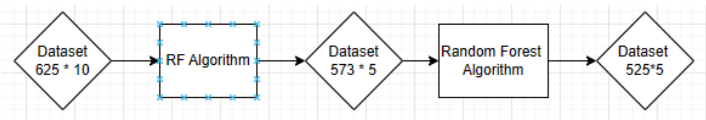

Figure 2: Hybrid data model

(1) **Method Dataset to be created and refined.**
(2) Obtain Clustured data, group length, and elimination number
(3) **For j = 1 to group_evaluation**
(4) find lenght of data_group(*j*) to group_evaluation(*i*) reducing (desc)
(5) **For k = 1 to deletion_count**
(6) eleminate *k*. data in group_evaluation(*j*)
(7) **End for**
(8) **End for**

(9) End procedure

Algorithm: Hybrid mechanism

1. Experimental Results

A software is created to gather data for automated algorithmic evaluation. The maintenance and operation information for a variety of machines was found. There were identified to be eleven input variables and one output variable. regression , vector machine, and neural networks models were trained and tested using clean 573 rows, eleven inputs, and 615* 10 output data. Tables 4-6, accordingly, give the values of the classifiers utilised in the study. Table 7 presents the prediction results again for data set, which consists of 615 records and 9 characteristics, to show how well the proposed hybrid system is performing.

Table 4: Tuned SVR parameters.

Parameters	Description
C	1.0
ε	0.002
Validation method	8-fold cross validation
Kernel function	Linear & horizontal

Table 5: Neural Networks

Parameters	Description
Number of neurons for the hidden layer	6
Hidden layers	4
Learning rate	0.3
Momentum	0.3
Epoch	900
Goal	0.008

Table 6: Regression parameters.

Parameters	Description
Batch size	100
Attribute selection method	sensor r
Ridge	$2.0E-8$

Table 7: Performance rating of models for (625 * 10) dataset (11 inputs 1 output).

Method	CC	MAE	RMSE
Regressio	0.92	0.8241	1.1546

n	67		
Neural Networks	0.90 25	0.8825	1.3292
Vector machine	0.93 08	0.881	1.4589

Table 8 : Random Forest.

Parameters	Description
Batch size	355
Attribute selection method	sensor r

Method	CC	Mean error	Root mean Error
Regression	0.956	0.682	1.08
Neural Network	0.962	0.686	1.077
Vector machine	0.942	0.703	1.088
Hybrid model	0.962	0.688	1.071

Table 7 Evaluation of the algorithms' efficiency for the (525 * 5) datasets (5 inputs and 1 output).

According to table 7, we can analyze and observe a new hybrid model which will make our system more refined with better efficiency, this model can be referred with respect to algorithm referred above which involves elimination of unrequired records and use the models of Neural networks, Regression and Vector machine as this creates a hybrid model with a better efficiency strategy which reduces the error rate of faulty sensors in turn increases the success rate.

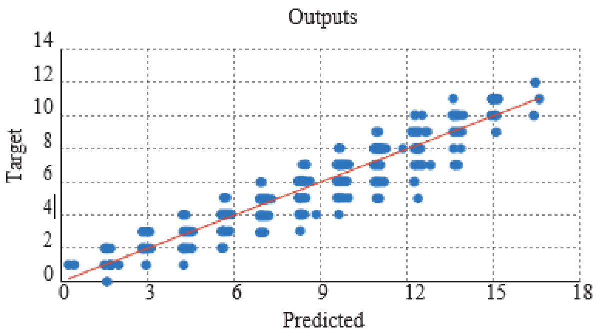

Figure 4 : connection between both the dataset's anticipated and desired results for regression.

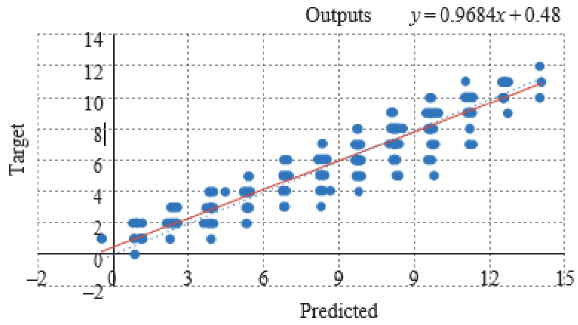

Figure 5 connection of the dataset's anticipated and desired value for super vector machines

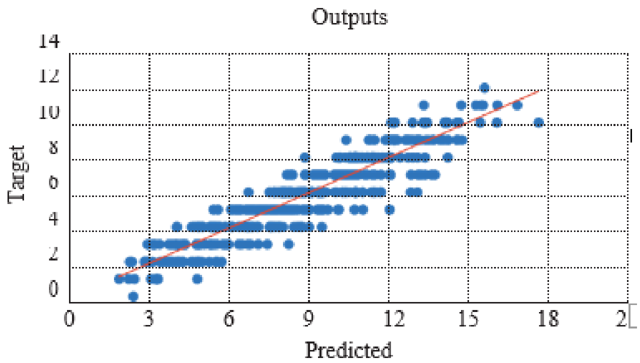

Figure 6 : connection of the dataset's anticipated and desired value for Neural Networks

According to Chart 7, the regression algorithm produced the best results based on the CC performance measure, whereas the regression method produced the best results based on the Mean square error and Root mean square performance characteristic. Then, the RF method was used to analyse the original data to see which characteristics are most beneficial for prediction. feature choice Nine input parameters were subjected to the RF method, and the last five of them were removed based on ranking values. Selected 573 rows, four inputs, and an output are used to construct MLP, LR, and SVR models. Both training and testing were done using the dataset (615*5). All of the results are superior than those obtained without feature selection, as shown in Table 8.

The data was subjected to the altered random forest algorithm at the analysis's concluding stage in order to remove loud and data inconsistency (615 * 5). It was discovered that (k = 13) was the ideal k value. The five parameters that were located the furthest away from each planet's centre were dropped. Therefore, 93 rows were removed. In order to increase the quality of the maintenance data, a hybrid model technique was used. The chosen 573 rows, four inputs, and one outputs were used to construct the regression, neural networks, and vector machine models. The chosen data (615 * 5) were tested on and sent to the test. In comparison to other findings produced without features selection or data compression, the results revealed that the model's efficiency was quite effective.

Based on the CC, Root mean square error, and Mean average error performance criteria, Table 9 demonstrates that the regression algorithm in the proposed hybrid hybrid model produced the best results. CC = 0.9485, Mean square error = 8868, and Root mean squared error = 0.910 were found for the test data. The linear correlation between the expected and desired outcomes is shown in Figures 4-6 for the testing data of Regression, Vector machine, and neural networks, respectively.

3. Conclusion

The utilization of repair data in the investigation of dependability and repair activities is crucial in the aviation industry. This is possible because estimations may be used to schedule predictive maintenance. Predictive maintenance's primary goal is to forecast technical glitches and create methods for replacement parts of system components while also evaluating the dependability and maintainability of a complex repairable system. In this work, the landing gear system maintenance dataset was subjected to a hybrid data preparation model employing feature selection RF algebra to choose characteristics and random forest algebra to **r**emove noise and data inconsistency. Through the use of regression, Vector machine, and neural networks models, the suggested hybrid data preprocessing strategy was put into practice.

The findings showed that the regression model performed better in failure count prediction than the neural networks and vector machine models. The findings also show that the suggested hybrid data preprocessing model greatly enhances failed count prediction performance. This paper might serve as a manual for applying hybrid data preprocessing techniques in data extraction and machine ML algorithms.

References

[1] Q. Fan and H. Fan, "Reliability analysis and failure prediction of construction equipment with time series models," Journal of Advanced Management Science, vol. 3, no. 3, pp. 203–210, 2015.

[2] P. Bastos, I. Lopes, and L. Pires, "A maintenance prediction system using data mining," in Proceedings of the Proceedings of the World Congress on Engineering, vol. 3, pp. 2–7, London, UK, July 2012.

[3] I. U. Din, M. Guizani, J. J. P. C. Rodrigues, S. Hassan, and V. V. Korotaev, "Machine learning in the Internet of)ings: designed techniques for smart cities," Future Generation Computer Systems, vol. 100, pp. 826–843, 2019.

[4] B. Jan, H. Farman, M. Khan, M. Talha, and I. U. Din, "Designing a smart transportation system: an internet of things and big data approach," IEEE Wireless Communications, vol. 26, no. 4, pp. 73–79, 2019.

[5] I. U. Din, M. Guizani, S. Hassan et al., ")e internet of things: a review of enabled technologies and future challenges," IEEE Access, vol. 7, pp. 7606–7640, 2019.

[6] S. Lecturer and A. Pradesh, "Feature selection using ReliefF algorithm," International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, no. 10, pp. 8215–8218, 2014.

[7] S. F. Rosario, "RELIEF: feature selection approach," International Journal of Innovative Research and Development, vol. 4, no. 11, pp. 218–224, 2015.

[8] H. Hsu, "Feature selection via correlation coefficient clustering," Journal of Software, vol. 5, no. 12, pp. 1371–1377, 2010.

[9] F. G¨urb¨uz, L. Ozbakir, and H. Yapici, "Data mining and ¨ preprocessing application on component reports of an airline company in Turkey," Expert Systems with Applications, vol. 38, no. 6, pp. 6618–6626, 2011.

[10] S. K. A. Fahad and M. Alam, "A modified K-means algorithm for big data clustering," International Journal of Computer Science Engineering and Technology, vol. 6, no. 4, pp. 129–132, 2016.

[11] N. Yilmaz, O. Inan, and M. S. Uzer, "A new data preparation method based on clustering algorithms for diagnosis systems of heart and diabetes diseases," Journal of Medical Systems, vol. 38, no. 5, 2014.

[12] M. Kutyłowska, "Neural network approach for failure rate prediction," Engineering Failure Analysis, vol. 47, pp. 41–48, 2015.

[13] P. Ramos, J. M. Oliveira, and P. Silva, "Predictive maintenance of production equipment based on neural network autoregression and ARIMA," in Proceedings of the 21st International EurOMA Conference-Operations Management in An Innovation Economy, pp. 1–10, Helsinki, Finland, June 2014.

[14] A. A. Trani, F. C. Wing-Ho, G. Schilling, H. Baik, and A. Seshadri, "A neural network model to estimate aircraft fuel consumption," in Proceedings of the AIAA 4th Aviation Technology, Integration, and Operations Forum, ATIO, vol. 2, pp. 669–692, Chicago, IL, USA, September 2004.

[15] M. Chen, R. Zhou, R. Zhang, and X. Zhu, "Application of artificial neural network to failure diagnosis on process industry equipments," in Proceedings 2010 6th International Conference on Natural Computation, ICNC 2010, vol. 3, pp. 1190–1193, Yantai, China, August 2010.

[16] P. Kozik, "Aircraft engine overhaul demand forecasting using ANN," Management and Production Engineering Review, vol. 3, no. 2, pp. 21–26, 2012.

[17] A. Altay, O. Ozkan, and G. Kayakutlu, "Prediction of aircraft failure times using artificial neural networks and genetic algorithms," Journal of Aircraft, vol. 51, no. 1, pp. 47–53, 2014. Scientific Programming 9

[18] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Remaining useful life estimation based on nonlinear feature reduction and support vector regression," Engineering Applications of Artificial Intelligence, vol. 26, no. 7, pp. 1751– 1760, 2013.

[19] M. D. C. Moura, E. Zio, I. D. Lins, and E. Droguett, "Failure and reliability prediction by support vector machines regression of time series data," Reliability Engineering & System Safety, vol. 96, no. 11, pp. 1527–1534, 2011.