International Journal of Computers and Applications

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjca20

Machine failure prediction using joint reserve intelligence with feature selection technique

Amal Shaheen, Mustafa Hammad, Wael Elmedany, Riadh Ksantini & Saeed Sharif

To cite this article: Amal Shaheen, Mustafa Hammad, Wael Elmedany, Riadh Ksantini & Saeed Sharif (2023) Machine failure prediction using joint reserve intelligence with feature selection technique, International Journal of Computers and Applications, 45:10, 638-646, DOI: 10.1080/1206212X.2023.2260619

To link to this article: https://doi.org/10.1080/1206212X.2023.2260619

Published online: 02 Nov 2023.

INTERNATIONAL JOURNAL OF COMPUTERS AND APPLICATIONS
2023, VOL. 45, NO. 10, 638–646
https://doi.org/10.1080/1206212X.2023.2260619

Machine failure prediction using joint reserve intelligence with feature selection technique

Amal Shaheen[a], Mustafa Hammad[b], Wael Elmedany [a], Riadh Ksantini[a] and Saeed Sharif[c]

[a]University of Bahrain, Sakhir, Bahrain; [b]Mutah Universityät, Karak, Jordan; [c]University of East London, London, UK

ABSTRACT
A model with high accuracy of machine failure prediction is important for any machine life cycle. In this paper, a prediction model based on machine learning methods is proposed. The used method is a combination of machine learning algorithms and techniques. The machine learning algorithm is a data mining technique that has been widely used as a prediction model for classifying problems. Five algorithms have been tested including JRIP, logistic, KStar, Bayes network and decision table machine learning. The evaluation process is done by applying the algorithms on a predictive dataset using different performance measures. In the proposed model, the feature selection and voting techniques are used and applied in the classification process for each classifier. From the comparison of the result, the feature selection shows the best performance result. Paired *t*-test evaluation measures were considered to confirm our conclusion. The best accuracy result among the five classifiers shows that joint reserve intelligence classifier can be used to predict the failure with an accuracy high as 0.983. Applying classifier subset evaluation using the JRIP classifier can enhance the accuracy result to be 0.985. The finding shows that the proposed model improves the results of the classifiers.

ARTICLE HISTORY
Received 13 April 2023
Accepted 28 July 2023

KEYWORDS
Machine failure; prediction; machine learning; explainable artificial intelligence; total production maintenance

1. Introduction

Machine failure prediction is a critical process that plays a crucial role in ensuring a better understanding of a machine's progress. Accurately predicting machine failures is challenging due to various external factors and the uncertainty surrounding which component will fail, why, and when. The ability to estimate the probability of imminent machine failure is a significant challenge. Failures of machines can have a significant impact, especially in industries where tangible products are produced. Compromised product quality without immediate detection can lead to project delays or cancellations. Additionally, the manpower required to address these failures promptly far exceeds that needed for regular machine operations. Without an effective prediction technique, the costs incurred can extend beyond monetary and temporal factors, potentially resulting in the loss of customer loyalty and diminished market competitiveness. Ultimately, businesses may face severe consequences, including the risk of closure. Furthermore, unexpected machine failures can disrupt data flow, leading to substantial losses in critical or historical data.

Designing a model that can accurately predict machine failures is of utmost importance. Such a model can provide valuable insights into when, where, and what types of machine failures are likely to occur. It has the potential to significantly reduce maintenance and replacement costs, as well as minimize downtime. The ability to predict machine failures is a key factor in ensuring project success and the production of high-quality goods. Early warning systems and proactive measures are vital to mitigating the damages associated with failures, as they allow for sufficient time to save and back up data that may be at risk during the failure event. While numerous models have been proposed for failure detection, diagnosis, and prediction, further research is needed to explore the limitations of existing approaches. This paper presents a machine failure prediction model based on Machine Learning (ML) techniques.

In this study, we propose a prediction model for machine failure using a combination of machine learning algorithms and techniques. The model leverages five ML classifiers, namely decision table, Bayes network, KStar, logistic regression, and Joint Reserve Intelligence (JRIP). We also apply feature selection and voting techniques to these classifiers and compare the results with those obtained prior to applying feature selection and voting. The feature selection algorithms employed include Classifier Attribute Evaluation (CAE), Correlation Attribute Evaluation (COAE), Infogain Subset Evaluation (ISE), and Classifier Subset Evaluation (CSE). To confirm the superiority of our proposed approach, we perform paired *t*-tests. Specifically, we compare the performance of JRIP-CSE, which incorporates classifier subset evaluation on the JRIP classifier, with the performance of other methods by calculating Probability Values (*P*-Values). The results demonstrate that JRIP-CSE outperforms other methods and exhibits statistically significant differences. In the remainder of this paper, we provide an overview of related work in Section 2 and describe the dataset used in Section 3. Section 4 explains the methodology, followed by the evaluation criteria in Section 5 and the experimental results in Section 6. We discuss the threats to validity in Section 7 and conclude with the main highlights and contributions of this work in Section 8.

2. Related work

ML can be applied to predict machine failure. Matzaka [1] explained how a decision model can use the information of the training data and process the information through several intelligent algorithms. The work in Matzka [1] and Molnar [2] described two ML models.

CONTACT Saeed Sharif s.sharif@uel.ac.uk

Explainable model using decision trees and explanatory interface. Gunning et al. [3] explained that Explainable Artificial Intelligence (XAI) is one of the ML models that makes machine behaviors more intelligible to humans. Gunning et al. [3] mentioned that XAI should be capable to tell what it has done, what is doing now, and what will happen next. The work in Benkedjouh et al. [4] used Support Vector Machine (SVM) as ML model that can predict the Remaining Useful Life of the machine (RUL). Matzaka [5] proposed a method using ML to increase the efficiency of the screw fasting process, which has the benefit of early prediction in the process quality.

In prediction decisions, interpretability is needed for decision-makers to help in providing reasoned justifications. Matzaka [1] talked about the models that can be used as interpretabilities' models, such as linear regression, decision trees and decision rules. Matzaka [1] explained the model-agnostic methods for the black box model such as explaining individual predictions, feature importance and accumulated local effects.

End users normally do not trust the result of the ML predictions. Kulesza et al. [6] proposed an approach called Explanatory Debugging (ED). ED approach explains the reasons for any prediction to the end users. This explanation will help the users to build a mental model of ML while the process of explaining the correction back to the system. The work in Kulesza et al. [6] concluded that ED builds better mental models than the black box which is the traditional learning system.

Measuring and evaluating the effectiveness of an explanation is very important in predictive machine failure. Gunning et al. [3] mentioned that user satisfaction can be rated in terms of clarity and utility of an explanation through a subjective rating. This leads that the measures are subjective. Although Gunning et al. [3] and Samek et al. [7] ensured that task performance might be one of the measures that are objective for an explanation's effectiveness. The work in Gunning et al. [3] concluded that the measurement and the evaluation for XAI systems include common sense, evaluation frameworks, different thinking, and argumentation. The work in Matzka [1] used the bag of trees classifier as an evaluating measure on the used predictive maintenance dataset. The work in Matzka [1] and Molnar [2] concluded that the explanatory model has limitations in the quality of the explanation whereas the decision tree model has limitations in the number of the explanation.

A lot of challenges are there when an explainable model is used for predictive machine maintenance. The intersection of ML and explanation is not easy. Ribeiro et al. [8] ensured that it is difficult to act on human behalf using context-aware systems. The work in Tomsett et al. [9] and Ribeiro et al. [8] considered interpretability as a challenge because ML systems must provide a level of high explanation to confirm the justification needed. Bellotti et al. [10] ensured that it is difficult to act on human behalf using context-aware systems. Rather, systems needed to be able to defer to users in a non-obtrusive fashion and with high efficiency. Both Gunning et al. [3] and Bellotti and Edwards [10] proposed a design framework that can understand the human aspects of context. Four designed principles have been presented by the framework. Those principles support the intelligibility of system behavior, accountability of humans as well as several details of human context. Bellotti et al. [10] highlighted that the knowledge and background of each user make the interactive and the feedback is different from one user to another. Gunning et al. [3] talked about the balance between accuracy, interpretability, and tractability as another challenge. Gunning et al. [3] concluded another challenge in explaining competencies versus explaining decisions.

ML used the knowledge of understanding and monitoring the previous performance to predict and optimize the efficiency of the overall production. The work in Campos et al. [11] and Irrera and Vieira [12] used ML in the Online Failure Prediction (OFP) technique to avoid failures by prediction. Various ML in Campos et al. [11] had been studied on three datasets and concluded that using the same data, with different ML algorithms and techniques directly influence the prediction performance. Both Mattioli et al. [13] and Perico and Mattioli [14] explained that ML can improve the Total Production Maintenance (TPM) by predicting any future event or behaviors for the machine. Mattioli et al. [13] proposed ML model based on the current state of a system as well as the analysis of past data. Mattioli et al. [15] ensured that insufficient prior knowledge of critical components will increase the uncertainty in failure diagnoses when using the ML models, this will affect the accurate failure estimation of the machine Remaining Useful Life (RUL).

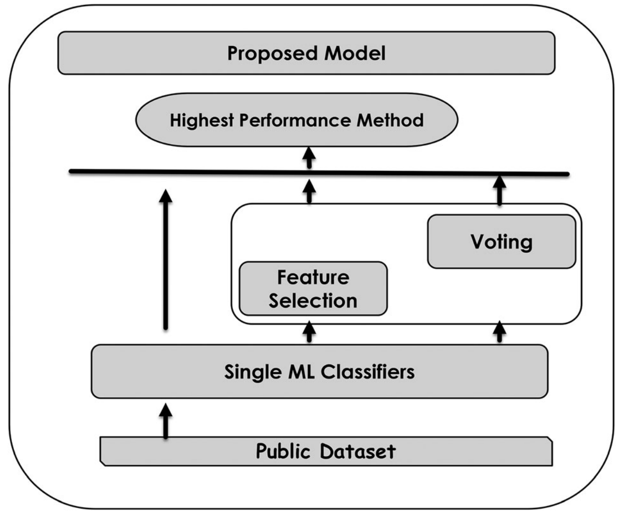


Figure 1. Block diagram of the proposed method using JRIP-CSE method.

In this paper, different techniques, procedures, and classifiers are used on the same dataset that is used in Matzka [1]. Five different ML classifiers are applied, as far as our knowledge those five classifiers have not before experimented on the same dataset. Feature selection technique is used with a 10-fold cross-validation procedure rather than fivefold cross-validation which was used in Matzka [1], Jain and Singh [16], Fathima et al. [17], Belete and Huchaiah [18], Talapula et al. [19] and Choudhary et al. [20]. The Voting technique is used using the average probability voting rule for the five classifiers, from our finding, this technique has not been applied before on the used dataset.

3. Proposed model

The proposed model used in this paper is presented as a block diagram in Figure 1. The model consists of three stages which are using single ML classifiers, applying ML techniques and finally selecting the highest performance method. A modification is done on the failure class type which is one of the dataset features. The change is needed because most of our classifiers are worked well on a nominal class attribute.

Mathematical model representation of the proposed work

Our proposed machine failure prediction model can be represented mathematically as follows:

Let X be the dataset containing n instances, denoted as $X = (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where x_i represents the feature vector for the ith instance and y_i is the corresponding target class label (machine failure or non-failure).

We assume the set of classifiers, $C = C_1, C_2, \ldots, C_m$, consists of five machine learning classifiers: Decision Table, Bayes Network, Logistic Regression, JRIP, and KStar. Each classifier C_i is trained on the dataset X to predict machine failures, and let F be the set of features in the dataset. We have dataset X containing instances and the corresponding target class labels Y. In Stage 1, we train each classifier c in C using the dataset X. Moving to Stage 2, we apply ML techniques for feature selection to identify the most relevant features from F. Proceeding to Stage 3, we apply a voting technique to aggregate the results of the trained classifiers in C. Finally, in Stage 4, we evaluate the performance of each method based on the f-measure values. From each stage, we select the method with the highest performance result, taking into account the selected features and the voting aggregation.

The following sections describe each stage in our model:

3.1. Applying ML technique

The second stage of the proposed model is applying the ML technique. The ML technique as illustrated in Figure 1 consists of two steps, which are:

3.1.1. Feature selection

Feature selection is a crucial machine learning technique employed to choose a subset of the most effective attributes from the dataset. Two primary methods, namely the filter method and wrapper method, are commonly used for feature selection. The wrapper method evaluates the features by incorporating them into an algorithm, while the filter method extracts relevant features based on the general characteristics of the dataset. In this paper, we employ three filter methods, namely Classifier Attribute Evaluation (CAE), Correlation Attribute Evaluation (COAE), and Infogain Subset Evaluation (ISE). Additionally, we utilize one wrapper method known as Classifier Subset Evaluation (CSE).

The feature method technique ranks all attributes from the highest relevance to the least. Thus features were selected based on the best performance result of the classifier with a smaller number of attributes. The experiment in this paper is done by excluding the lowest two attributes in the feature selection ranking for all the classifiers, this is done because the results did not improve when more than two attributes were excluded.

3.1.2. Applying voting technique

The last step of stage 2 is applying the technique of vote aggregating that merges the results of more than one classifier. All five classifiers are used as base classifiers with an average probability voting rule.

3.2. Highest performance method

The final stage of the proposed model is selecting the best method based on the best performance result. As shown in Figure 1 comparison is done between the results of the previous stages. The method that has the highest performance result from each stage has been selected as one of the best methods. Paired t-test evaluation measures on the f-measure values were considered to confirm the conclusion of the best result from each stage. Paired t-tests are performed by pairing up the method that shows the best result which is JRIP-CSE with other classification methods which are JRIP and vote. This is done to compute Probability Values (P-Values). The P-Value quantifies the probability of the paired distributions being nearby or not. More precisely, we assume that a p-value lower than 50% signifies reasonably different sample groups.

Table 1. Dataset attributes description.

Features	Description	Value
Prod Num	Serial Number depend on quality of the product	String
Air Heat	Temperature of the air	Integer
Process Heat	Walk heat process	Integer
Power speed	Machine power speed	Integer
Engine	Engine power	Pos Integer
Material	Depend on the tool that used	Time
Machine Failure	If the machine has failed or not in a specific time	Nominal

Table 2. Prod num statistics.

Label	Count	Weight
M	2997	2997
L	6000	6000
H	1003	1003

Table 3. Confusion matrix.

	True	
Predicted	Failure	Operation
Failure	TRP	FAP
Operation	FAN	TRN

3.3. Used dataset

The dataset that is used in this paper is AI4I 2020 predictive maintenance dataset which is available online thru ML repository [21]. The dataset reflects a real predictive maintenance dataset. The dataset has 10,000 cases for machine predictive maintenance.

Each predictive has seven features, which determine whether the machine fails or not. Table 1 presents a summary of the seven features. Prod num consists of the letter L, M, or H based on the product quality variants. Table 2 shows that 60% of all products are categorized as low, 30% are medium and for high 10%. A machine failure label is to indicate if there is a failure or not. The used data set has only 339 cases labeled as machine failure and 9661 cases labeled as operation. The failure rate is 3.39 and this would normally be a problem in the production environments.

4. Results and discussion

4.1. Evaluation criteria

The confusion matrix and the used evaluation measures are discussed in this section. Table 3 shows the confusion matrix, which is used as a classifier performance. The confusion matrix reviews the results of the testing algorithm. FAN, FAP, TRN and TRP, and present False Negative, False Positive, True Negative and True Positive, respectively.

The measures for our evaluation are Mean Absolute Error (MAE), accuracy, recall, precision, and f-measure. These measures are used to evaluate the performance of the classifiers. Used measures are described and computed as follows:

(1) Accuracy, which is the percentage of the correct predictions. If the result is 1, it means the best accuracy. If the result is 0, it means the worst. Accuracy is calculated using the following equation:

$$\text{Accuracy} = \frac{(TRP + TRN)}{(TRP + TRN + FAP + FAN)} \tag{1}$$

(2) Precision, which is the fraction of relevant instances among the retrieved instances. If the result is 1, it means the best precision.

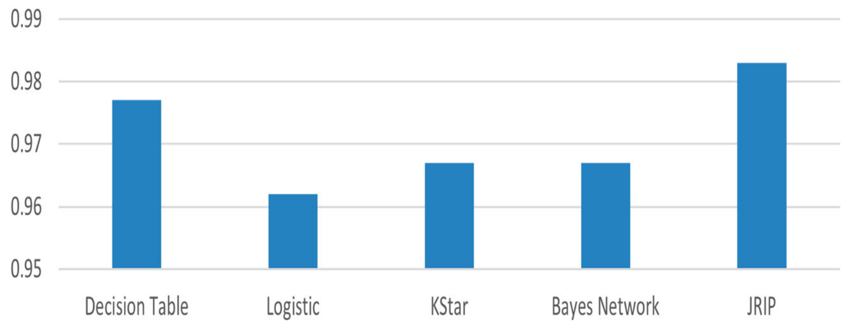


Figure 2. Result of *F*-measure for the five classifiers.

Table 4. Results of performance measure for the five classifiers.

Model	MAE	Accuracy	Recall	Precision	*F*-Measure
Decision Table	0.039	0.979	0.979	0.977	0.977
Logistic	0.049	0.969	0.970	0.963	0.962
KStar	0.034	0.972	0.972	0.967	0.967
Bayes	0.065	0.964	0.964	0.971	0.967
JRIP	0.026	0.984	0.984	0.984	0.983

If the result is 0, it means the worst. The following formula can estimate the precision value:

$$\text{Precision} = \text{confidence} = \frac{TRP}{TRP + FAP} \tag{2}$$

(3) Recall, which is how many of the true positive's instances are recalled. If the result is 1, it means the best recall. 1f the result is 0, it means the worst. It is computed using the following equation:

$$\text{Recall} = \frac{TRP}{TRP + FAN} \tag{3}$$

(4) *F*-measure, the combination of the precision and the recall result. The following formula to calculate the result:

$$F\text{—measure} = 2(\text{PrecisionRecall}/\text{Precision} + \text{Recall}) \tag{4}$$

(5) MAE, the total of true instances minus the predicted instances. The formula that calculates the value of MAE is

$$MAE = \text{Truevalues} - \text{Predictedvalues} \tag{5}$$

4.2. Experimental results

The evaluation process is implemented using the Weka tool [22]. Weka can be used to process the machine learning algorithms on any data. Cross-validation procedure using 10-fold is set for all used algorithms that measure the impact of the dataset in the prediction.

Table 4 shows the accuracy, recall, precision, MAE and *F*-measure results for the five classifiers prediction models. JRIP and decision table have the best prediction accuracy result which is 0.984 and 0.979 respectively. Moreover, results show that the Bayes network and logistic have the lowest accuracy value which is 0.964 and 0.969 respectively. The lowest MAE value is for the JRIP classifier which is 0.026. MAE value for other classifiers is close and higher than 0.03. This will lead to that JRIP having the best accuracy and MAE prediction. We believe this happens because JRIP focuses on nominal and binary data.

Another comparison is done between the five classifiers using the combination between the recall and the precision results. Thus f-measure is used for this comparison. Figure 2 shows the values of the f-measure for each classifier. As shown in Figure 2, the highest f-measure value is 0.983 and it is achieved by JRIP. This result again matches what is shown in Table 4 where JRIP achieves the lowest MAE and highest accuracy score.

Table 5. Performance evaluation results before and after using CSE.

Model	Filter	MAE	Accuracy	Recall	Precision
Decision Table	Before	0.039	0.979	0,979	0.977
	After	0.036	0.979	0.980	0.978
	Diff	0.003	0	−0.001	−0.001
Logistic	Before	0.049	0.969	0.970	0.963
	After	0.052	0.972	0.972	0.972
	Diff	−0.003	−0.003	−0.002	−0.009
KStar	Before	0.034	0.972	0.972	0.967
	After	0.045	0.974	0.974	0.973
	Diff	−0.011	−0.002	−0.002	−0.006
Bayes Network	Before	0.065	0.963	0.964	0.971
	After	0.061	0.975	0.976	0.973
	Diff	0.004	−0.012	−0.012	−0.002
JRIP	Before	0.026	0.984	0.984	0.984
	After	0.025	0.985	0.985	0.984
	Diff	0.001	−0.001	−0.001	0

Table 5 shows the results before and after applying CSE, and the difference between the results for each classifier. Accuracy has increased almost in all the classifiers except for decision table as shown in Table 5. Bayes network has experienced the most accuracy increase by 0.012 points. MAE has decreased in some models as shown in Table 5. Bayes network has the most reduction in MAE, which is about 0.003 points. Both decision table and JRIP have also decreased by 0.003 and 0.001 respectively.

Figure 3 shows the accuracy values before and after applying CSE. The result shows an improvement in the accuracy of the classifiers. There is a change in accuracy rate, especially logistic and Bayes network, where the accuracy rate has increased.

Figure 4 shows the MAE's value before and after applying feature selection. Three classifiers show an improvement. KStar and Logistic have no improvement. JRIP-CSE scored the sharpest decrease value which is 0.025.

Figure 5 shows a comparison between the three filter algorithms based on the accuracy increased points. The accuracy increased point is the difference in the result before and after applying CAE, COAE, and ISE filters. CAE has the best increase for all the classifiers. The highest increase in accuracy is 1.32 points shown in the decision table when using the CAE. Whereas Accuracy dropped to 0.07 points in the decision table when using ISE.

Figure 6 shows a comparison in the MAE result of the difference between the result before and after applying the three filter algorithms. Decision table scored the best MAE decrease which is 0.027 points when applying CAE.

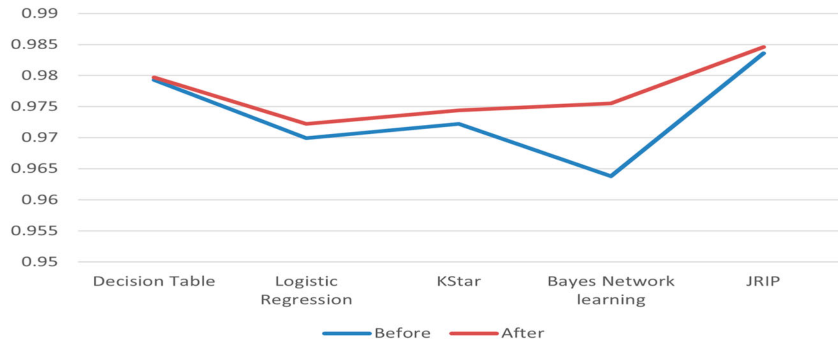


Figure 3. Accuracy values of classifiers before and after using CSE.

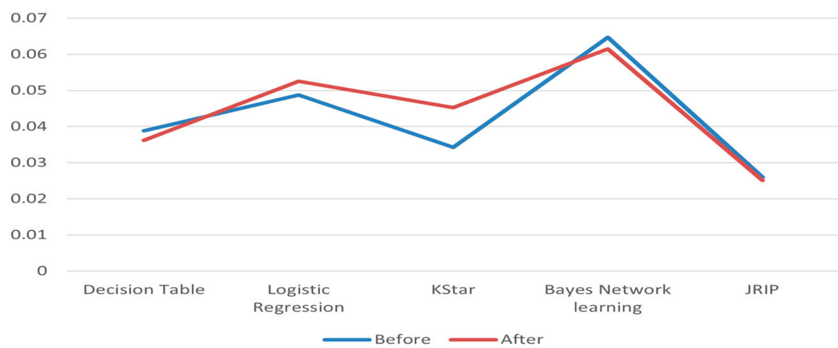


Figure 4. MAE values of classifiers before and after using CSE.

Although JRIP has the best second decrease in MAE and the best second increase in the accuracy using CAE filter, JRIP scored the highest accuracy result among the other classifiers after applying the CAE and the best MAE result which is 0.975 and 0.026 respectively as shown in Figures 7 and 8. The accuracy result for the JRIP before applying CAE is higher than the result after applying CAE and it is 0.983 as shown in Figure 7. Thus CAE shows the best result compared to the other used filter methods, but it is still having a lower result compared with the result before applying CAE.

This concludes that the best result from part 2 of our experiment is 0.985 which is shown in using the method JRIP-CSE in the dataset.

Table 6 shows the value of the vote when we combine the five classifiers. The f-measure result is 0.982 and the accuracy result is 0.983 and it is lower than the best result that scored when using the feature selection.

Table 7 shows the accuracy and MAE values for the vote, JRIP before applying CSE and JRIP-CSE. This concludes that applying the CSE technique on the JRIP classifier is giving the highest performance accuracy result which is 0.985 and the best MAE result which is 0.025.

Table 6. Performance results using vote.

	MAE	Accuracy	Recall	Precision	*F*-Measure
Vote	0.043	0.983	0.983	0.982	0.982

Table 7. Accuracy and MAE results for vote, JRIP before CSE and JRIP after CSE.

Method	MAE	Accuracy
Vote	0.043	0.983
JRIP	0.026	0.984
JRIP-CSE	0.025	0.985

Table 8 shows the value of f-measure and P-value for JRIP, vote with the combination of the five classifiers and JRIP-CSE. The highest f-measure is 0.984 and is shown in JRIP-CSE. The P-value result shows that JRIP and vote is less than 50%, which means that JRIP-CSE is statistically different than the other methods and it has the best performance result.

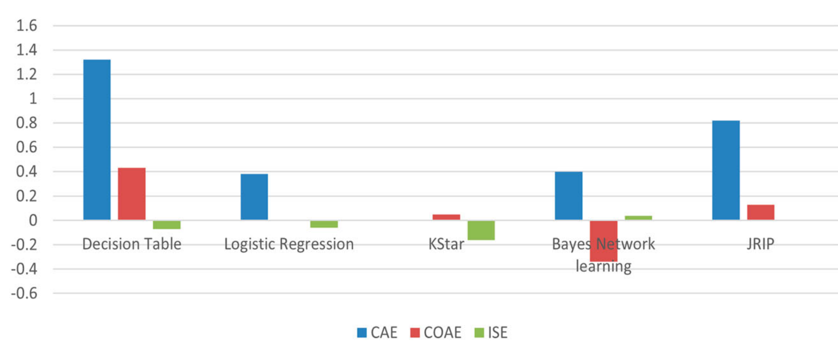


Figure 5. Accuracy increased points when applying CAE, COAE, and ISE on the five classifiers.

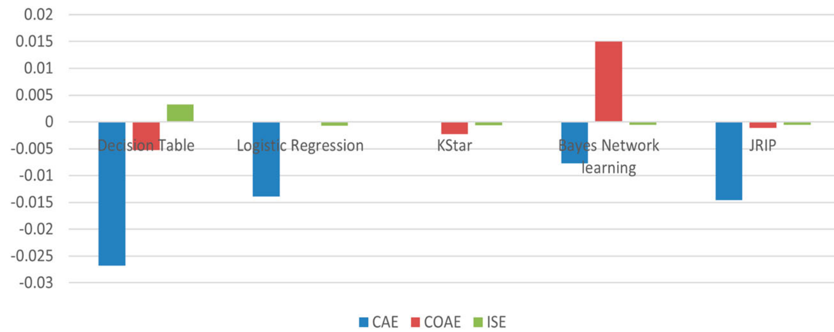


Figure 6. MAE decreased points when applying CAE, COAE, and ISE on the five classifiers.

Figure 7. Accuracy result comparison before and after using the CAE filter.

Table 8. *F*-measure and *P*-value results for vote, JRIP before CSE and JRIP-CSE.

Method	*F*-measure	*P*-value
Vote	0.982	38.9
JRIP	0.983	38.3
JRIP-CSE	0.984	–

Table 9 shows the comparison between the best accuracy result achieved in this paper and the best accuracy result that in Matzaka [1]. The best accuracy result for the work done in Matzaka [1] when applying bagged trees ensemble classifier using 5-flod cross validation and was 98.34%. The accuracy result achieved in this paper is higher by 0.120 points.

Table 10 shows a comparative performance analysis between JRIP in the previous study and our method. The table effectively demonstrates that our proposed method, JRIP-CSE, excels in comparison to the JRIP method in previous study across multiple datasets. It highlights the significance of the extension in predictive maintenance tasks and showcases the overall efficacy of JRIP-CSE as powerful classifiers for various applications. These results offer valuable insights into the performance and potential of our method and contribute to the advancement of classification techniques in different domains.

4.3. Analysis and discussion

The results obtained from our evaluation process provide valuable insights into the performance of different machine learning algorithms for machine failure prediction. In this discussion, we will analyze and interpret the results, focusing on the comparison of algorithm performance and highlighting the effectiveness of the feature selection technique. The accuracy results presented in Table 3 demonstrate the varying performance of the five classifiers. JRIP and decision table achieved the highest accuracy scores, indicating

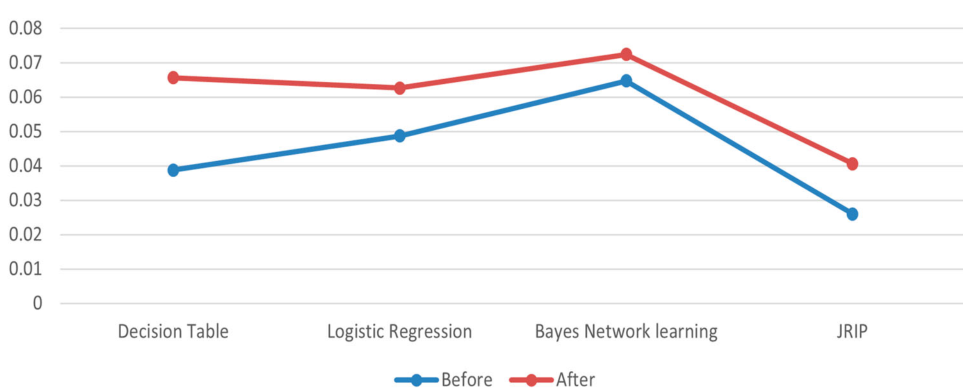


Figure 8. MAE result comparison before and after using the CAE filter.

Table 9. Comparison between the best accuracy result in previous work and best result in this paper.

Work	Classifier	Dataset	Technique	Procedure	Highest accuracy
Matzaka [1]	SVM	AI4I 2020	Feature selection	5-fold Cross-Validation	Bagged trees ensemble (98.34)
	ANN				
This paper	JRIP	AI4I 2020	1-Feature selection	10-fold Cross-Validation	JRIP-CSE (98.46)
	KStar		2-Vote		
	Bayes Network				
	Logistic				
	Decision Table				

Table 10. Comparison between the best result in previous work using JRIP and best result in this paper.

Method	Work	*F*-measure	Accuracy	Dataset
JRIP	This paper	0.983	0.984	I4I 2020 predictive maintenance
JRIP-CSE	This paper	0.984	0.985	I4I 2020 predictive maintenance
JRIP	[23]	0.667	–	cultural heritage
JRIP	[24]	0.88	93.28	Iris
JRIP	[25]	–	97.26	Heart Disease
JRIP	[26]	–	94.0	Iris

their effectiveness in predicting machine failures. On the other hand, Bayes network and logistic regression exhibited lower accuracy values. These findings suggest that the decision table and JRIP classifiers are particularly well-suited for this prediction task, outperforming the other algorithms. The MAE values, also shown in Table 3, provide insights into the precision of the classifiers. JRIP achieved the lowest MAE, indicating its ability to accurately predict machine failures with minimal errors. The other classifiers had slightly higher MAE values, indicating a relatively higher degree of error in their predictions. These results further support the superior performance of JRIP in terms of accuracy and precision. The *F*-measure, which combines recall and precision, was used to compare the classifiers' overall performance. As shown in Table 3, JRIP achieved the highest *F*-measure value, reinforcing its effectiveness in predicting machine failures. The consistent performance of JRIP across multiple evaluation metrics suggests its reliability and robustness as a machine learning algorithm for failure prediction. The next set of results, presented in Table 4 and Figure 4, focus on the impact of the Classifier Subset Evaluation (CSE) technique on the classifiers' performance. Notably, almost all classifiers showed improvements in accuracy after applying CSE, with the Bayes network demonstrating the most significant increase. This finding suggests that the CSE technique has a positive influence on the accuracy of machine failure prediction. Additionally, Figure 5 illustrates the improvements in MAE values for three classifiers after feature selection. JRIP-CSE exhibited the most substantial decrease in MAE, indicating its enhanced precision in predicting machine failures. The comparison of the three feature selection algorithms (CAE, COAE, and ISE) presented in Figure 6 highlights the effectiveness of CAE in increasing accuracy across all classifiers. The decision table, in particular, experienced a significant increase of 1.32 points in accuracy when CAE was applied. Conversely, the use of ISE led to a decrease in accuracy for the decision table. These results emphasize the importance of selecting an appropriate feature selection algorithm to improve the prediction performance of machine learning classifiers. Moreover, Figure 7 provides insights into the impact of feature selection on MAE values. The decision table exhibited the most substantial decrease in MAE when CAE was applied. While JRIP showed the second-best decrease in MAE and an increase in accuracy after applying CAE, it achieved the highest accuracy result among the classifiers. This finding suggests that feature selection, particularly with the CAE algorithm, can enhance the precision of machine failure prediction. The results obtained from applying the JRIP-CSE method, as presented in Table 6, demonstrate its superior performance compared to other techniques. JRIP-CSE achieved the highest accuracy result of 0.985, indicating its effectiveness in predicting machine failures. The MAE value of 0.025 further reinforces the precision and reliability of JRIP-CSE. Table 7 provides the f-measure and P-value results, confirming the superior performance of JRIP-CSE. It achieved the highest f-measure value of 0.984, indicating its balanced performance in terms of recall and precision. The statistical significance demonstrated by the P-value results further supports the superiority of JRIP-CSE compared to other methods. Finally, in Table 8, we compare our best accuracy result with the best accuracy result obtained in the work by Matzka [1]. Our accuracy result of 98.34% surpasses the previously achieved result by 0.120 points. This comparison highlights the effectiveness of our approach in improving machine failure prediction accuracy. In conclusion, our results demonstrate the effectiveness of machine learning algorithms, particularly JRIP, in predicting machine failures. The application of feature selection techniques, such as CAE, has been shown to significantly improve accuracy and reduce MAE values. The superior performance of JRIP-CSE, as supported by statistical analysis, further emphasizes the effectiveness of our proposed approach. These findings contribute to the development of more reliable and accurate machine failure prediction models, enabling proactive maintenance and minimizing the associated costs and downtime.

4.4. Comparison with state-of-the-art method

Our proposed machine failure prediction model, which incorporates various machine learning algorithms and techniques, has demonstrated promising results when compared to the state-of-the-art methods in the field. In this section, we will compare our model's performance with a notable existing approach, highlighting the advancements and improvements achieved. The state-of-the-art method chosen for comparison is the work by Matzka [1], which employed a bagged trees ensemble classifier using fivefold cross-validation. Their best accuracy result was reported as 98.34%. In contrast, our proposed model achieved an accuracy result of 98.46%, surpassing the state-of-the-art method by 0.120 points.

This comparison clearly indicates the superior performance of our machine failure prediction model. By combining multiple machine learning algorithms and applying feature selection techniques, we were able to enhance the accuracy of predictions. The use of the Joint Reserve Intelligence (JRIP) classifier, in particular, proved to be highly effective, outperforming the state-of-the-art method. Furthermore, our model employed classifier subset evaluation (CSE) with JRIP, resulting in an accuracy score of 98.50%. This represents a substantial improvement compared to the state-of-the-art method and demonstrates the importance of incorporating feature selection techniques tailored to the specific dataset and prediction task.

In addition to accuracy, our model also outperformed the state-of-the-art in terms of other evaluation measures, such as mean absolute error (MAE) and *F*-measure. The MAE values achieved by our model were consistently lower, indicating greater precision in predicting machine failures. Similarly, the *F*-measure values demonstrated the robustness and balanced performance of our model, particularly with JRIP-CSE, which achieved the highest score.

Overall, our comparison with the state-of-the-art method highlights the advancements made by our proposed machine failure prediction model. By leveraging multiple machine learning algorithms, employing feature selection techniques, and utilizing the strengths of the JRIP classifier, we achieved higher accuracy, improved precision, and enhanced overall performance. These findings contribute to the advancement of machine failure prediction and have practical implications for industries relying on accurate failure prediction to minimize costs and downtime while ensuring high-quality product delivery.

4.5. Threats to validity

This section explains a few issues that might affect the validity of our result. First, the experiment is done on one dataset. Using one dataset might affect the process of feature selection. The second threat is that applying one wrapper algorithm is not enough. Various algorithms might behave differently with the same dataset.

5. Conclusion and future work

Machine failure prediction is a critical issue that requires proactive measures to minimize the associated costs and damages. However, predicting when, why, and where a failure might occur is challenging due to various influencing factors. In this study, we proposed a machine failure prediction model based on machine learning (ML) techniques and applied it to a dataset. Our findings demonstrated promising results, with the JRIP classifier showing the best performance, achieving an accuracy value of 0.984 and a mean absolute error (MAE) of 0.026.

Furthermore, we explored the impact of feature selection by comparing results before and after excluding two attributes based on their performance ranking. The Bayes Network model exhibited significant improvement in accuracy after applying feature selection. Nevertheless, the JRIP classifier maintained its superior performance among all classifiers even after feature selection, demonstrating its robustness and effectiveness in this prediction task. Notably, the JRIP-CSE method further improved accuracy to 0.985, representing the highest accuracy result in the second part of our experiment.

In the third part of the experiment, we applied the voting technique, combining the predictions of all five classifiers. However, no significant improvement was observed in the results compared to using individual classifiers.

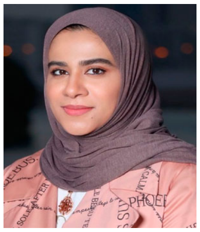
For future work, we suggest including multiple datasets in the experiment to validate the model's performance across different scenarios and enhance its generalizability. Additionally, since feature selection played a crucial role in improving our results, further investigation is warranted to explore different wrapper algorithms' effectiveness on the same dataset. By incorporating various wrapper algorithms, we can gain deeper insights into feature relevance and select the most suitable method for specific machine failure prediction tasks.

In conclusion, our proposed machine failure prediction model, particularly with the JRIP classifier and feature selection, has demonstrated promising results. This model holds significant potential in enhancing machine life cycle management, reducing costs associated with failures, and ensuring the production of high-quality goods. By addressing the challenges of machine failure prediction through advanced ML techniques, our approach contributes to improving the efficiency and reliability of predictive maintenance strategies in industries where machine failures have substantial consequences.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors



Amal Shaheen is a Ph.D. candidate in Computing and Information Science at the University of Bahrain. She received her Master's Degree in Information Technology from NYIT, Bahrain, in 2014. Her research interests include AI transformation, Machain Learning and AI modeling.

Mustafa Hammad is an Associate Professor in the Department of Software Engineering at Mutah University. He received his Ph.D. in Computer Science from New Mexico State University, USA in 2010. He received his Master's Degree in Computer Science from Al-Balqa Applied University, Jordan, in 2005 and his B.Sc. in Computer Science from The Hashemite University, Jordan, in 2002. His research interests include wireless sensor networks, machine learning, and software engineering, focusing on software analysis and evolution.

Wael Elmedany is currently an associate professor of Computer Engineering at University of Bahrain, Kingdom of Bahrain, and senior member of IEEE society. Elmedany holds a PhD degree in Electrical Engineering, Manchester University, UK, 1999; MSc degree in computer communications, Menoufia University, Egypt, 1991; BSc degree in Electronic Engineering, Menoufia University, Egypt 1987. Elmedany was assistant professor of computer communications, department of Electrical Engineering, Cairo University, Fayoum branch, 2000–2005, and assistant professor of electronics and computer communications, in Faculty of Electronic Engineering, Menoufia University, Egypt, 1999–2000. Elmedany was the research and development director of Evara R&D, Egypt, 2001–2003, and general director of Varia IC, Egypt, 2003–2005. He is the founder and managing editor of International Journal of Computing and Digital Systems (IJCDS). He is the founder and organizer of MobiApps, DPNoC, and WoTBD workshops and symposium series. He is also an editor of many journals special issues, member of editorial boards for many international journals, and member of technical program committees of many conferences. He is also reviewer in many international journals and conferences, and acts as chairperson in many conferences. El-Medany has over forty publications in a well-known international conferences and journals, and attended several national and international conferences and workshops. His research interests in ASIC design, FPGA, embedded systems, network on chip, remote monitoring systems, and reconfigurable computing.

Riadh Ksantini received the M.Sc. and Ph.D. degrees in Computer Science from the Université de Sherbrooke, Sherbrooke, QC, Canada, in 2003 and 2007, respectively. From 2001 to 2007, he was a graduate research associate with Bell Canada Laboratories and the research center MOIVRE (MOdElisationen Imagerie, Vision et REseaux de neurones). Presently, he is Associate Professor at the Department of Computer Science, College of IT, University of Bahrain, Adjunct Associate Professor at the School of Computer Science, within the Faculty of Science of the University of Windsor, Windsor, Ontario, Canada, and Adjunct Professor at the Department of Computer Science, Université du Québec à Montréal (UQAM). Prior to that, he was Postdoctoral Research Associate in the Ecole de Technologie Superieure, University du Quebec, Montreal, Canada. He has also served as Visiting Fellow Research Scientist at the Canadian Space agency, and Postdoctoral Research Associate in the School of Computer Science, within the Faculty of Science of the University of Windsor. In 2008, he was awarded a fellowship (of excellence) for postdoctoral research from the granting agency "Fonds quebecois de la recherche sur la nature et les technologies" (FQRNT). His PhD was evaluated and ranked third Ph.D. in Quebec for 2007 by the committee of Information Technology and Communications. His research interests include Artificial Intelligence, Machine/Deep Learning, Pattern Recognition and Computer Vision. His research work on Artificial Intelligence and Data Science has always involved collaboration between academia and industry. More than 100 Articles stemming from his research work have been published in several prestigious journals and conferences.

Dr. Saeed Sharif is the Leader of Intelligent Technologies Research Group, Course Leader for Msc Computer Science and Course Leader for Msc Computer Science with industrial Placement at the school of architecture, computing, and engineering. His research interests and expertise include Artificial Intelligence, Innovative Tele-Health, Medical Technology, Digital Health Care and Medical Assistive Technology, Medical Image Analysis and Visualisation, Intelligent Diagnosis Systems, Smart Biomedical Image, and Bio Signal Acquisition (MRI, PET, EEG), Nanotechnology, Medical Biotechnology, and Big Data Analysis. He is working closely with clinicians and policy makers nationally and internationally to improve the clinical settings and the healthcare systems. He has led the research development of different research projects associated with many industries and NHS trusts. He has developed his research with the goal of improving the effectiveness, efficiency of medical and health care systems.

Dr Saeed has published significant number of articles in highly reputable journals, international conferences, and book chapters, such as Elsevier Applied Soft Computing, Computer Methods and Programs in Biomedicine, IEEE Access etc. He is a member of the technical committee of several international conferences such as IEEE International Conference on Computer Systems and Applications and IEEE International Conference on Computer and Information Technology. He has participated in many national and international conferences e.g., ICIP. He has served as a reviewer for many journals e.g., IET IPJ, Elsevier CBMJ. He is Guest Editor at an international journal. He received many academic and research awards. He is a Fellow of The UK Higher Education Academy, and member of the British Computer Society.

ORCID

Wael Elmedany http://orcid.org/0000-0001-8827-6009

References

[1] Matzka S. Explainable artificial intelligence for predictive maintenance applications. In: 2020 Third International Conference on Artificial Intelligence for Industries (AI4I). IEEE; 2020. p. 69–74. Irvine, CA: USA
[2] Molnar C. Interpretable machine learning. Lulu.com; 2020.
[3] Gunning D, Stefik M, Choi J, et al. Xai – explainable artificial intelligence. Sci Robot. 2019;4(37):eaay7120. doi:10.1126/scirobotics.aay7120
[4] Benkedjouh T, Medjaher K, Zerhouni N, et al. Remaining useful life estimation based on nonlinear feature reduction and support vector regression. Eng Appl Artif Intell. 2013;26(7):1751–1760. doi:10.1016/j.engappai.2013.02.006
[5] Matzka S. Using process quality prediction to increase resource efficiency in manufacturing processes. In: 2018 First International Conference on Artificial Intelligence for Industries (AI4I). IEEE; 2018. p. 110–111. Laguna Hills, California, USA
[6] Kulesza T, Burnett M, Wong W-K, et al. Principles of explanatory debugging to personalize interactive machine learning. In: Proceedings of the 20th International Conference on Intelligent User Interfaces. 2015. p. 126–137. ACM IUI 2015, Atlanta, GA, USA.
[7] Samek W, Montavon G, Vedaldi A, et al. Explainable AI: interpreting, explaining and visualizing deep learning. 1 Vol. 11700. Springer Nature; 2019. p. 439.
[8] Ribeiro MT, Singh S, Guestrin C. 'Why should I trust you?' Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. p. 1135–1144. San Francisco California, USA
[9] Tomsett R, Widdicombe A, Xing T, et al. Why the failure? How adversarial examples can provide insights for interpretable machine learning. In: 2018 21st International Conference on Information Fusion (FUSION). IEEE; 2018. p. 838–845. Cambridge, United Kingdom.
[10] Bellotti V, Edwards K. Intelligibility and accountability: human considerations in context-aware systems. Hum Comput Interact. 2001;16(2–4): 193–212. doi:10.1207/S15327051HCI16234_05
[11] Campos JR, Vieira M, Costa E. Exploratory study of machine learning techniques for supporting failure prediction. In: 2018 14th European Dependable Computing Conference (EDCC). IEEE; 2018. p. 9–16. IaÈŹi, Romania
[12] Irrera I, Vieira M. Towards assessing representativeness of fault injection-generated failure data for online failure prediction. In: 2015 IEEE International Conference on Dependable Systems and Networks Workshops. IEEE; 2015. p. 75–80. Rio de Janeiro, Brazil
[13] Mattioli J, Perico P, Robic P-O. Improve total production maintenance with artificial intelligence. In: 2020 Third International Conference on Artificial Intelligence for Industries (AI4I). IEEE; 2020. p. 56–59. Irvine, CA, USA.
[14] Perico P, Mattioli J. Empowering process and control in lean 4.0 with artificial intelligence. In: 2020 Third International Conference on Artificial Intelligence for Industries (AI4I). IEEE; 2020. p. 6–9. Irvine, CA, USA.
[15] Mattioli J, Perico P, Robic P-O. Artificial intelligence based asset management. In: 2020 IEEE 15th International Conference of System of Systems Engineering (SoSE). IEEE; 2020. p. 151–156. Budapest, Hungary.
[16] Jain D, Singh V. A two-phase hybrid approach using feature selection and adaptive SVM for chronic disease classification. Int J Comput Appl. 2021;43(6):524–536.
[17] Fathima MD, Samuel SJ, Natchadalingam R, et al. Majority voting ensembled feature selection and customized deep neural network for the enhanced clinical decision support system. Int J Comput Appl. 2022;44(10):991–1001.
[18] Belete DM, Huchaiah MD. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. Int J Comput Appl. 2022;44(9):875–886.
[19] Talapula DK, Ravulakollu KK, Kumar M, et al. SAR-BSO meta-heuristic hybridization for feature selection and classification using DBNover stream data. Artif Intell Rev. 2023;56:1–39.
[20] Choudhary C, Singh I, Kumar M. Sarwas: deep ensemble learning techniques for sentiment based recommendation system. Expert Syst Appl. 2023;216:119420. doi: 10.1016/j.eswa.2022.119420
[21] Dataset-link: https://archive.ics.uci.edu/ml/datasets/ai4i+2020+predictive+maintenance+dataset.
[22] Weka-tool: https://www.cs.waikato.ac.nz/ml/weka/.
[23] Rehman IU, Ali Z, Jan Z, et al. Empirical performance analysis of classification methods on cultural heritage database. 2022.
[24] Veeralakshmi V, Ramyachitra D. Ripple down rule learner (RIDOR) classifier for IRIS dataset. Issues. 2015;1(1):79–85.
[25] Almustafa KM. Prediction of heart disease and classifiers' sensitivity analysis. BMC Bioinform. 2020;21(1):1–18. doi: 10.1186/s12859-020-03626-y
[26] Devasena CL, Sumathi T, Gomathi V, et al. Effectiveness evaluation of rule based classifiers for the classification of iris data set. Bonfring Int J Man Mach Interface. 2011;1:5.