Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi
Afyon Kocatepe University – Journal of Science and Engineering
https://dergipark.org.tr/tr/pub/akufemubid

Araştırma Makalesi / Research Article
e-ISSN: 2149-3367
DOI: https://doi.org/10.35414/akufemubid.1571811
AKÜ FEMÜBİD **25** (2025) 041304 (785-797)
AKU J. Sci. Eng. **25** (2025) 041304 (785-797)

Failure Prediction Using Ensemble Learning: A Comparative Study with Synthetic and Real-World Datasets

Makale Bilgisi / Article Info
Alındı/Received: 22.10.2024
Kabul/Accepted: 05.03.2025
Yayımlandı/Published: 04.08.2025

Topluluk Öğrenmesi Kullanılarak Arıza Tahmini: Sentetik ve Gerçek Dünya Veri Setleriyle Karşılaştırmalı Bir Çalışma

Asli Beyza CIFTPINAR[1], Pelin KANAR[2], Zeynep Idil ERZURUM CICEK[2*]

[1] *Bursa Uludağ University, Faculty of Engineering, Industrial Engineering Department, Bursa, Türkiye*
[2] *Eskisehir Technical University Faculty of Engineering, Industrial Engineering Department, Eskisehir, Türkiye*

Abstract
The ability to predict and prevent machine failures is a crucial task for businesses on a global scale at a time of increasing dependence on automation and technology. This paper primarily addressed a novel failure prediction model approach based on ensemble learning. Commonly used machine learning models including Decision Trees, K-Nearest Neighborhood, Support Vector Machines, and Logistic Regression and two different ensemble learning strategies were used: bagging and majority voting. The SZVAV real-life failure dataset provided by Lawrence Berkeley National Laboratory and the AI4I2020 Predictive Maintenance synthetic dataset were utilized to evaluate the performance of the proposed ensemble models. The preprocessing stage included the application of oversampling since there is an imbalance problem in both datasets. In this context, a comparison of three oversampling techniques was also presented for the datasets considered in the study. As a result of the tests, it was seen that the proposed models are superior to individual machine learning methods and Random Forest, which is an ensemble model itself, for the considered datasets. In addition, the proposed ensemble models were compared with the original failure prediction models previously presented in the literature on the AI4I2020 dataset, and it was reported that more successful results are obtained with the proposed approach.

Keywords *Failure prediction; Machine learning; Ensemble learning; Oversampling.*

Öz
Makine arızalarını tahmin etme ve önleme yeteneği, otomasyon ve teknolojiye olan bağımlılığın arttığı bir zamanda küresel ölçekte işletmeler için kritik bir görevdir. Bu çalışma öncelikle topluluk öğrenmeye dayalı özgün bir arıza tahmin modeli yaklaşımını ele almaktadır. Karar Ağaçları, K-En Yakın Komşuluk, Destek Vektör Makineleri ve Lojistik Regresyon dahil olmak üzere yaygın olarak kullanılan makine öğrenmesi modelleri ve iki farklı topluluk öğrenme stratejisi kullanılmıştır: torbalama ve çoğunluk oylaması. Lawrence Berkeley Ulusal Laboratuvarı tarafından sağlanan SZVAV gerçek yaşam arıza veri seti ve AI4I2020 Tahmini Bakım sentetik veri seti, önerilen topluluk modellerinin performansını değerlendirmek için kullanılmıştır. Her iki veri setinde de bir dengesizlik sorunu olduğu için ön işleme aşaması aşırı örnekleme uygulamasını içermektedir. Bu bağlamda, çalışmada ele alınan veri setleri için üç aşırı örnekleme tekniğinin bir karşılaştırması da sunulmuştur. Testler sonucunda, ele alınan veri setleri için önerilen modellerin bireysel makine öğrenmesi yöntemlerinden ve kendisi bir topluluk modeli olan Rastgele Orman'dan üstün olduğu görülmüştür. Ayrıca önerilen topluluk modelleri, AI4I2020 veri seti üzerinden literatürde daha önce sunulan orijinal hasar tahmin modelleri ile karşılaştırılmış ve önerilen yaklaşımla daha başarılı sonuçlar elde edildiği raporlanmıştır.

Anahtar Kelimeler *Arıza tahmini; Makine öğrenmesi; Topluluk öğrenmesi; Aşırı örnekleme.*

1. Introduction

Equipment performance and reliability are crucial for maintaining continuous production and profitability in the dynamic world of industrial operations. However, maintenance and operations teams continue to face difficulties due to the intrinsic complexity of contemporary industrial systems and the constant demands placed on them. Unexpected equipment failure can have devastating consequences, including unplanned downtime, compromised safety, and significant financial loss. Mainly, maintenance can be described as the prevention of equipment failure and making sure that the equipment performs effectively and failurelessly, at least for the duration of its useful life. Maintenance practices are an important part of the production systems. For this reason, it is of great importance to determine these maintenance times before failures occur and to eliminate the problem before it happens. Maintenance strategies

**Sorumlu Yazar/Corresponding Author: Zeynep Idil Erzurum Cicek* *e-mail: zierzurum@eskisehir.edu.tr*

are essential for ensuring the reliability, efficiency, and longevity of equipment, machinery, and infrastructure in various industries. These strategies encompass a range of proactive and reactive approaches aimed at preventing equipment failures, minimizing downtime, and optimizing operational performance. Proactive maintenance includes preventive measures such as regular inspections, scheduled maintenance routines, and predictive maintenance techniques like condition monitoring and predictive analytics. On the other hand, reactive maintenance involves addressing issues as they arise, often through corrective actions and emergency repairs. A well-balanced maintenance strategy combines these elements to maximize asset availability while minimizing costs and disruptions. Additionally, advancements in technology, such as the Internet of Things (IoT) and data analytics, are playing an increasingly significant role in modern maintenance strategies, enabling organizations to harness real-time data for predictive maintenance, ultimately driving greater efficiency and cost-effectiveness. The area of predictive maintenance has arisen as an innovative reaction to urgent concerns caused by unexpected equipment breakdowns, enabling industries to move past reactive tactics and adopt proactive strategies.

Predictive maintenance (PdM) is the prediction of failures that may occur in the future by using the data obtained during the use of the machines or equipment in production. Bousdekis *et al*. (2019) emphasized that PdM benefits significantly from technological advances, using real-time detection to predict future failures. PdM, after receiving and processing some physical data (vibration, temperature, etc.) from the machinery or equipment, helps to perform just-in-time maintenance by detecting when the failure may occur, with high accuracy, sufficiently in advance. Thus, it helps to get rid of cost and time loss by performing maintenance at the estimated failure times.

Various technologies are also used in PdM. Some of those; are thermal imaging tests with thermal cameras, oil and particle test, ultrasonic test, and vibration analysis. Sectors such as manufacturing, automotive, and aviation are among the sectors where predictive maintenance is frequently applied. PdM has become a popular research area since Industry 4.0 applications become widespread. Various topics are covered under PdM, such as the prediction of machine failures, and forecasting remaining useful life (RUL). For the machine failure prediction pillar of predictive maintenance, Zhu *et al.* (2019) emphasized that knowledge-based models, traditional machine learning models, and deep learning models are used. Zonta *et al.* (2020) also highlighted the use of data-driven solutions for PdM, due to the ample data collection opportunities and use of machine learning algorithms. Dundar *et al.* (2021) stated that individual machine learning models; various ensemble models and deep learning models are used for PdM.

PdM applications for identifying machine failures also use individual machine learning models, which are typically used for failure prediction issues. Individual models for PdM include Logistic Regression (Philips *et al.* 2015), Raza *et al*. (2010), Support Vector Machines (Baptista *et al.* 2018, Mei *et al.* 2022, Gohel *et al.*2020, Shamayleh and Awad 2020, Cakir *et al.* 2021, Lee *et al*. 2019, Arslan and Tiryaki 2020), Decision Trees (Kaparthi 2020, Mei *et al.* 2022, Bukhsh 2019), K-Nearest Neighbors (Cakir *et al*. 2021), Random Forests (Cakir *et al*. 2021, Janssens *et al*. 2019, Mei *et al.* 2022), Artificial Neural Networks (Gencer *et al.* 2021, Mei *et al.* 2022, Arslan and Tiryaki, 2020). In addition to traditional machine learning models, deep learning models such as Convolutional Neural Networks (Lee *et al.* 2019, Kaya *et al.* 2022), Recurrent Neural Networks, And Long Short-Term Memory (Wu *et al.* 2020, Patra *et al.* 2022) were also utilized.

In PdM applications, ensemble learning-based methods are used as well as individual machine learning models. Hung (2021) proposed an Adaptive Boosted Decision Tree approach using the output of a Decision Jungle model based on bagging and succeed to enhance the performance of the failure prediction model. In the prediction of wind turbines main bearing failures, Beretta *et al.* (2021) presented an anomaly detection model using an ensemble method, isolation forest. Mujib and Djatna (2020) developed an ensemble machine-learning model, using a voting strategy to predict the failures of a wafer stick machine. The model which increased the prediction accuracy includes Lazy-Locally-Weighted Learning (LWL), Zero-R, J48, and Random Forest classifiers. To diagnose the refrigerant charge failures, Zhang *et al.* (2023a) developed a stacking ensemble of 5 basic classifiers, namely SVM, Random Forests, Gradient Boosting Machines, Back Propagation Neural Network (BPNN), Multi-class Logistic Regression. Khan *et al.* presented an ensemble of XGBoost, Random Forest, and Extra Tree models to predict the failures of wind turbines. Zhang *et al.* (2023b) proposed a blending ensemble model comprised of a large number of machine learning models to predict hard disk failures. In a recent study conducted by Khalil and Rostam (2024), an ensemble model consists of SVM and Adaboost proposed for failure prediction in a rotating machinery and it was demonstrated that the bagging ensemble of the mentioned individual models enhanced the prediction performance. Similarly, ensemble learning-based models are being developed for the remaining useful life (RUL) (Gungor *et al*. 2022a, (Chen *et al.* 2021, Gungor *et al.* 2022b). Apart from the ensemble learning approach, various models have also been developed for PdM, which include a hybrid of different classifiers. (Andre *et al*. 2013, Fernandes *et al.,* 2020). To compare our study and the studies in the literature, the previous studies about failure prediction via ensemble learning are listed in Table 1.

Table 1. Comparison of this study with existing literature.

Study	Individual models	Ensemble learning strategy
Hung (2021)	Decision Jungle & Adaptive Boosted Decision Tree	Bagging & boosting
Beretta *et al.* (2021)	Artificial Neural Network & Isolation Forest	Combination through a rolling windowed sum
Mujib and Djatna (2020)	Locally weighted learning &0-R classifier& Decision Stump& Random Forest algorithm	Majority voting
Zhang *et al.* (2023a)	SVM, RF, Gradient Boosting Machine, Back Propagation Neural Network, Multi-class Logistic regression	Stacking
Khan *et al.* (2023)	XGBoost, Random Forest, and Extra Tree models	Stacking
Zhang *et al.* (2023b)	Logistic Regression, K-Nearest Neighbor, Support Vector Machine, Naïve Bayes, Random Forest, Gradient Boosting Decision Tree, Extreme Gradient Boosting, AdaBoost, Back-Propagation Neural Network, Long Short-Term Memory	Blending
Khalil and Rostam (2024)	Support Vector Machine & Adaboost	Bagging
This study	**Logistic Regression, Decision Trees, Support Vector Machines, K-Nearest Neighbor**	**Majority voting & Bagging**

Table 1 shows the position of our paper in the failure prediction literature. In our study, we aimed to develop an ensemble model structure different from the ensemble failure prediction models in the literature. For this purpose, we developed two different ensemble models using classical machine learning models and bagging and majority voting strategies. In addition, we added an oversampling process to our proposed approach to eliminate the class imbalance problem, which is a frequently encountered problem in real-life data. This increases the accuracy and robustness of the proposed failure prediction approach. This study contributes to the literature by demonstrating the effectiveness of ensemble strategies in fault prediction and provides a more reliable solution for PdM applications.

To sum up, in this study, to predict machine failures with high-performance, ensemble learning-based machine learning models are proposed. The main contributions of the study can be listed as follows:

- The class imbalance problem in the datasets was tackled with the oversampling step applied in the pre-processing step.
- Two tailored ensemble machine learning models comprised of 4 traditional machine learning algorithms – Logistic Regression, Support Vector Machines, K-Nearest Neighbor, and Decision Trees, and 2 ensemble learning strategies - majority voting and bagging, are proposed for machine failure prediction to enhance the prediction performance.
- The performance of the proposed models was compared with individual machine learning models and the Random Forest model, which is ensemble in nature.

The remaining of the study is organized as follows: In Section 2, the methodology of the study is explained, and the proposed approach is described. In Section 3, the experimental results are reported. The applicability of the proposed method and its comparison with existing studies in the literature are discussed in Section 4. The evaluation of the results obtained in the study and future work are expressed in Section 5.

2. Methodology

Since this study presents models based on ensemble learning, an overview of ensemble learning, and the considered machine learning methods is provided in this section.

2.1 Ensemble learning

Ensemble learning can be defined as an umbrella term for methods that combine multiple inducers to make a decision, typically in supervised machine learning tasks. Ensemble learning is a powerful technique in machine learning that leverages the combination of multiple individual models to improve predictive accuracy and robustness. By aggregating the knowledge of diverse models, ensemble methods can often outperform single models and mitigate the variance and the risk of overfitting (Saihood and Sonuc 2023). The three main ensemble learning methods are bagging, boosting, and stacking (Mienye and Sun 2022). Bootstrap aggregation, commonly referred to as bagging, is a widely used ensemble method that leverages the parallel combination of multiple instances of the same model to enhance model performance and stability (Saihood and Sonuc 2023). In the Bagging (Bootstrap Aggregating) method, each of the core learners are trained with different randomly selected subsets of the training set. After the data is separated as training and test, it is ’put in the bag’ of each learner by making a random selection from the data set allocated for training. The selected ones remain in the training set so that they can be selected again. In the final stage, the decisions made are combined with weighted voting. The purpose of choosing different training sets is to increase success by obtaining decision differences.

The selection of the technique used to merge the base learners is a crucial stage in the construction of ensemble classifiers. The combination mechanism is often determined by the type of ensemble learning approach being employed. Once the underlying models are trained, the combination rule can be used in bagging and boosting. Majority voting is the method most frequently used to aggregate ensemble base models. By aggregating the results from individual models’ predictions, a voting ensemble operates. When classifying data, each label’s predictions are summed together, and the label with the greatest number of votes is predicted.

2.2 Logistic regression

A statistical classification model for the estimate of a categorical variable in which the dependent variable is binary is known as logistic regression (LR). The fundamental idea behind LR is based on probability. As a cost function, LR uses the sigmoid function to constrain the output value to the range of 0 and 1. Regression assumptions are the same as classical regression in logistic regression except for the dependent variable is binary (Demir and Karaboga 2021).

2.2 Decision trees

A decision tree (DT), a non-parametric supervised learning method, builds classification or regression models using a tree-like topology. To generate a model that predicts the value of a target variable, the goal is to learn simple decision rules generated from the data attributes. Models of trees called classification trees allow the target variable to take a discrete value. The branches of these tree structures represent the features, and the leaves represent the class labels that are combined to make the class labels.

2.3 Support vector machines

One of the most popular machine learning methods for classification problems is Support Vector Machines (SVM) a supervised learning method developed by Vapnik (1995) based on statistical learning theory. The purpose of SVMs is to find a hyperplane that classifies data points separately in an N-dimensional feature space. Many hyperplanes can be chosen to classify data belonging to two classes. However, SVM finds the hyperplane with the maximum margin among these hyperplanes. In this optimization problem, the margin is defined as the shortest distance between the decision boundary, the hyperplane, and any sample in the data set. SVM may correctly classify data that cannot be separated linearly by employing kernel functions including linear, polynomial, and radial basis functions.

2.4 K-nearest neighbor

One of the simplest and most popular classification algorithms, K-Nearest Neighbor (KNN), works by calculating the minimal distance between a data point and each of its k-nearest neighbors. The number of nearest neighbors, or parameter k, can be hard to specify but is essential to classifier accuracy. The KNN method can be summed up in the following phases because it is relatively straightforward (Raschka 2015). First, the distance metric and the number k are determined. The sample that needs to be categorized then has k closest neighbors. The class label is finally chosen by a majority vote.

2.5 Random forest

Random Forest, first introduced by Breiman (2001), is an ensemble learning technique that builds a large number of decision trees during the training phase for classification and regression tasks. The class that the majority of the trees choose is the random forest’s output for classification tasks. Genuer *et al.* (2010) explain the principle of RF as to combine many binary decision trees built using several bootstrap samples coming from the learning sample L and choosing randomly at each node a subset of explanatory variables *X* (Erzurum Cicek and Kamisli Ozturk 2022).

3. Experimental Study

3.1 Dataset

One of the simplest and most popular classification algorithms, K-Nearest Neighbor (KNN), works by in the study, two failure detection datasets are used to test the proposed approach. The first one is AI4I 2020 Predictive Maintenance Dataset (UCI Machine Learning Repository 2020). The AI4I 2020 Predictive Maintenance Dataset is a synthetic dataset that mimics actual PdM data seen in the manufacturing sector. The dataset includes 10000 data points and comprises 14 columns, including one column indicating the presence or absence of failure and five columns representing different failure types. The identification variables for the records and products are denoted as unique identifier (UID) and product ID, respectively. The remaining six columns—type, air temperature, process temperature, rotational speed, torque, and tool wear—represent the dataset's features. Table 2 presents the descriptive statistics of these features.

Table 2. Descriptive statistics for AI4I2020 Predictive Maintenance dataset.

Numerical features					
Feature	**Unit**	**Min**	**Max**	**Mean**	**St. Dev.**
Air temperature	K	295.3	304.5	300.005	2.000
Process temperature	K	305.7	313.8	310.006	1.484
Rotational speed	rpm	1168	2886	1538.776	179.284
Torque	Nm	3.8	76	39.987	9.969
Tool wear	min	0	253	107.951	63.654
Categorical features					
Feature	**Value range**		**Frequency**	**Percent (%)**	
Type	L: Low		6000	60	
	M: Medium		2997	29.97	
	H: High		1003	10.03	

Table 3. Descriptive statistics for SZVAV dataset.

Feature	**Unit**	**Min**	**Max**	**Mean**	**St.Dev.**
Supply Air Temperature	^oF	49.55	104.6	71.427	11.155
Supply Air Temperature Heating Set	^oF	53.04	72.5	61.810	9.612
Supply Air Temperature Cooling Set	^oF	55.04	72.5	65.479	8.263
Outdoor Air Temperature	^oF				
Mixed Air Temperature	^oF	58.99	89.1	70.533	5.279
Return Air Temperature	^oF	66.33	90.28	75.127	4.774
Supply Air Fan Status	-	1	1	1	0
Supply Air Fan Speed Control Signa	-	0.1	0.5	0.277	0.179
Outdoor Air Damper Control Signal	-	0	1	0.540	0.465
Return Air Damper Control Signal	-	0	1	0.480	0.485
Exhaust Air Damper Control Signal	-	-0.04	0.9	0.471	0.434
Cooling Coil Valve Control Signal	-	0	1	0.297	0.411
Heating Coil Valve Control Signal	-	0	1	0.0699	0.109
Occupancy Mode Indicator	-			0.500	0.500

The other dataset, which was generated by Lawrence Berkeley National Laboratory (LBNL) for an air handling unit (AHU) and a single zone variable air volume (SZVAV) AHU in LBNL's FLEXLAB test facility (Granderson 2019). This dataset is one of the automated failure detection and diagnostics (AFDD) testing data sets created by LBNL, PNNL, NREL/ORNL, and Drexel University. There are 15840 samples and 14 features in SZVAV dataset. The features in the dataset are supply air temperature, supply air temperature heating set point, supply air temperature cooling set point, outdoor air temperature, mixed air temperature, return air temperature, supply air fan status, supply air fan speed control signal, outdoor air damper control signal, return air damper control signal, exhaust air damper control signal, cooling coil valve control signal, heating coil valve control signal, and occupancy mode indicator. Table 3 presents the descriptive statistics of these features.

Figure 1 displays a summary of samples from both datasets that are failure and non-failure in numbers. In the figures, 1 corresponds to states with a failure, 0 for cases without a failure. As can be seen in Figure 1, the distribution of samples with and without failures in both datasets is imbalanced.

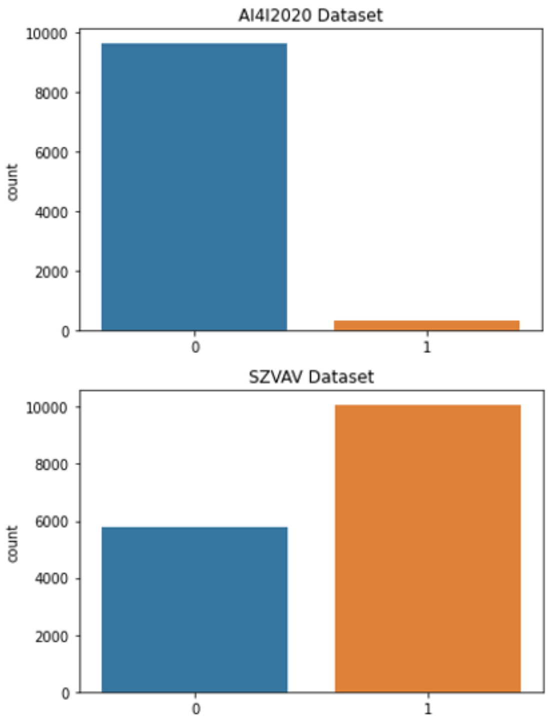

Figure 1. The counts of class labels (failure:1 and non-failure:0) in the considered datasets

3.2 Data pre-processing

A preprocessing step was applied to bot of the datasets before testing the proposed ensemble learning models. In the first stage, columns that were thought to be irrelevant in the datasets were deleted. Secondly, the missing data in both datasets were checked, and it was concluded that there was no missing data problem in AI4I2020 Data, while missing data was detected in two features - supply air temperature heating set point and supply air temperature cooling set point in SZVAV dataset. The missing data were imputed with the mean values of the relevant feature columns. After these preprocessing steps, min-max normalization was applied to the datasets to enhance model performance. For each feature in both datasets, min-max-normalization formulated in Equation 1 was applied to scale the data between 0 and 1:

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{1}$$

where x_i' is scaled value of x_i from data *x*. Table 1 provides the minimum and maximum values for each feature that was taken into consideration.

As mentioned in the previous section, the number of samples of class labels with and without failures is imbalanced in both datasets. In an imbalanced dataset, the data belonging to one class is significantly more in number than the data belonging to the other class, causing the class with less data to not be learned sufficiently by machine learning methods (Ay and Yolacan, 2022). The imbalance problem is an important problem that needs to be eliminated because it will not show the performance of the ensemble models to be proposed in the study and the individual models to be used for comparison. For this reason, this imbalance problem should be eliminated before the tests (Akgul et al., 2020).

A way to deal with imbalanced data is to stabilize the data with resampling methods (Aydin, 2022). As resampling techniques, undersampling or oversampling can be applied to balance the number of class labels in the dataset. In this study, it was decided to apply the oversampling procedure to overcome the imbalance problem.

Oversampling can be defined as increasing the amount of minority class instances or samples by producing new instances or repeating some instances (Mohammed et al., 2020). Numerous oversampling techniques, including SMOTE, ADASYN, random oversampling, and SVMSMOTE, are described in the literature. Random oversampling, SMOTE and ADASYN, which are the most used oversampling methods, were used in this study.

3.3 Proposed ensemble models

Two ensemble models are proposed in this study. Both proposed models include SVM, KNN, LR and DT classifiers. The difference between the models is the ensemble type. The first ensemble is based on the majority voting, the other is based on the bagging technique. Random subsamples are generated for the bagging ensemble using the original training dataset, while the entire training set is used to train each classifier in the voting ensemble model. Ensemble1 is referred to as the majority voting ensemble, while Ensemble2 is referred to as the bagging ensemble. The structure of the proposed models is visualized in Figure 2.

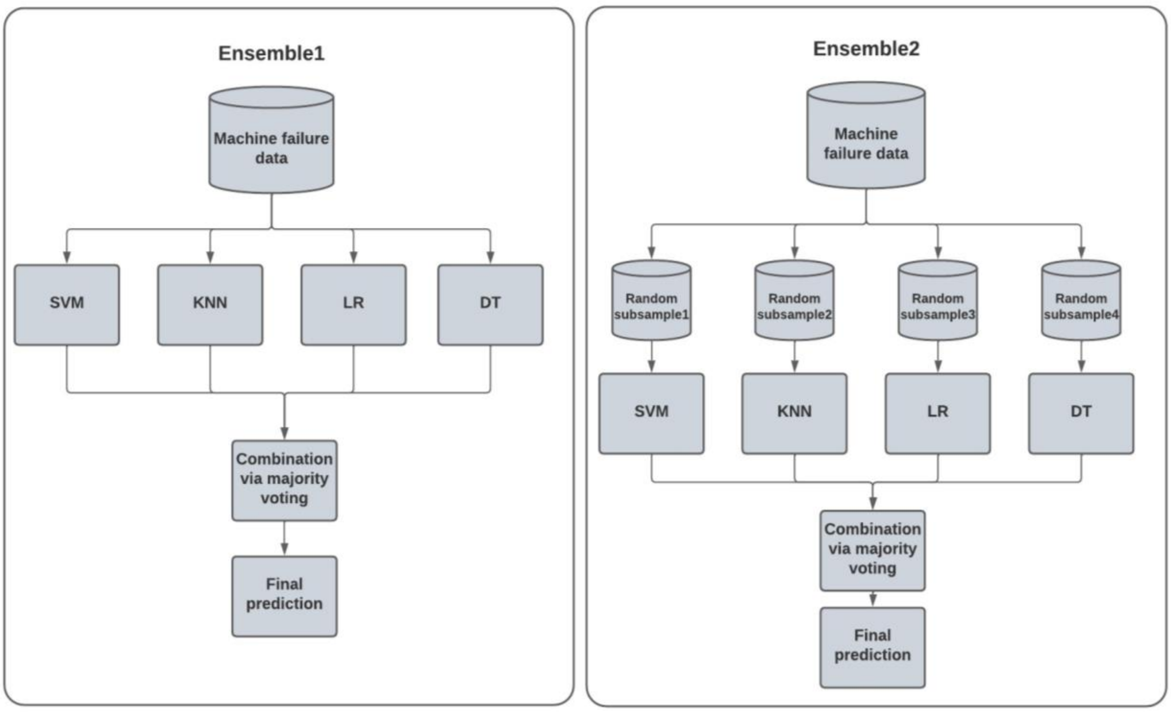

Figure 2. The structure of the proposed ensemble models

3.3 Computational experiments and results

Before testing the mentioned individual and proposed ensemble machine learning models, grid search method was used to determine algorithm parameters. For each classifier, training and tests were carried out by choosing the best combination of parameters among the parameter sets. Table 4 provides a summary of the parameter sets that were evaluated by grid search: Accuracy and F-score metrics were chosen to compare test performances. The formulations of the accuracy and F-score metrics are given in Equations 2-5, respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TN + FP} \quad (3)$$

$$Recall = \frac{TP}{TN + FN} \quad (4)$$

$$F - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

The tests were first started with individual machine-learning models. At this stage, first, the imbalance problem in the data was resolved, and then the training and testing processes were carried out by using the best parameter combinations obtained with the individual models.

The average accuracy and F-score values obtained as a result of the training and testing processes applied with the 10-fold cross-validation method are reported on the basis of the oversampling method in Table 5 and F-scores are visualized in Figure 3.

According to the F-score values obtained, it can be said that the RF model performs better than other models for AI4I 2020 dataset. For SZVAV dataset, SVM model mostly outperformed other individual classifiers.

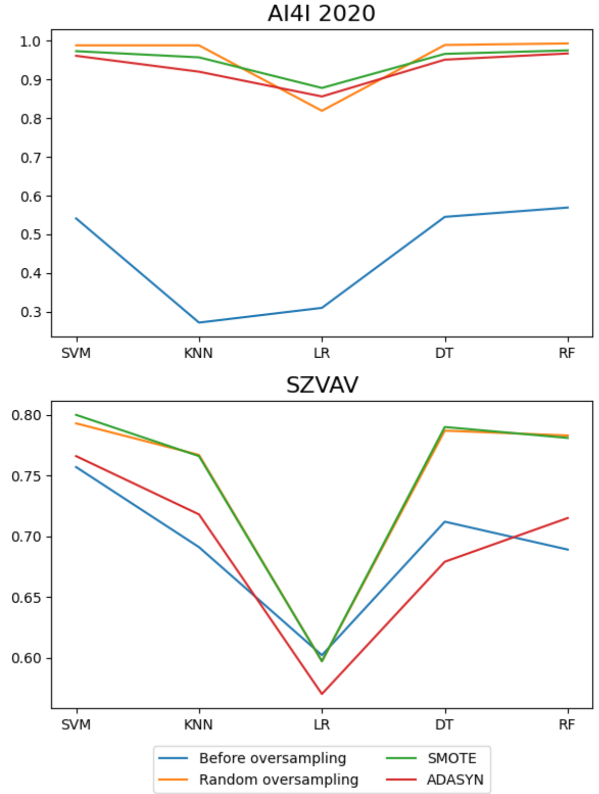

Figure 3. The test results for the individual models

Table 5. The test results for individual models.

Dataset	Oversampling method	Performance metric	SVM	KNN	LR	DT	RF
AI4I 2020	Before oversampling	Accuracy	0.974	0.971	0.971	0.969	0.977
		F-score	0.541	0.272	0.310	0.545	0.569
	Random oversampling	Accuracy	0.988	0.987	0.814	0.990	0.993
		F-score	0.988	0.988	0.819	0.990	0.993
	SMOTE	Accuracy	0.973	0.956	0.876	0.965	0.975
		F-score	0.973	0.957	0.878	0.966	0.976
	ADASYN	Accuracy	0.961	0.922	0.857	0.952	0.967
		F-score	0.961	0.920	0.856	0,951	0.967
SZVAV	Before oversampling	Accuracy	0.748	0.673	0.651	0.684	0.660
		F-score	0.757	0.691	0.602	0.712	0.689
	Random oversampling	Accuracy	0.848	0.837	0.722	0.844	0.850
		F-score	0.793	0.767	0.597	0.787	0.783
	SMOTE	Accuracy	0.851	0.836	0.721	0.843	0.849
		F-score	0.8	0.766	0.597	0.79	0.781
	ADASYN	Accuracy	0.814	0.771	0.677	0.721	0.767
		F-score	0.766	0.718	0.570	0.678	0.715

In addition, tests were performed without oversampling, so the effect of oversampling was demonstrated. As can be seen from the results, although the accuracy values are high before oversampling, the F-score values are quite low. Regardless of the method, when oversampling was applied, F-score values increased noticeably, and accuracy values improved. It has been demonstrated once again that the accuracy metric does not reflect the prediction performance fairly in imbalanced datasets and these results once again revealed the fact that working with imbalanced data can be misleading.

In the Ensemble2 model, each subsample is randomly generated. Therefore, for a reliable and fair comparison, 10 runs were made for both proposed ensemble models. It is ensured that the training and test sets in each run are generated randomly. The averages of accuracy and F-score values obtained as a result of the tests are reported in Table 6. The results clearly show that the proposed ensemble models increase prediction success in accuracy and F-score metrics.

The comparison of the performance of individual models with the proposed ensemble models is visualized in Figure 4. In addition, the results of different oversampling methods on the prediction performance can be observed in Figure 4. It is clear that the proposed ensemble models that combine individual models with both majority voting-only and bagging has improved prediction performance. Again, it is seen that the proposed models increase the prediction performance compared to RF.

It is concluded that ensemble models are clearly superior to individual models, especially in the SZVAV dataset containing real-life data. This indicates that the prediction success will be quite high in a PdM application where the proposed models will be included. Lastly, if we evaluate the prediction performance in terms of oversampling methods, it is seen that it does not significantly affect the performance of the ensemble models proposed in the study. However, applying oversampling increased the performance of all models in terms of F-score value for both datasets.

Table 6. The test results for the best individual and the proposed ensemble model.

		SVM	KNN	LR	DT	RF	Ensemble1	Ensemble2
AI4I2020	Accuracy	0.988	0.987	0.876	0.990	0.993	0.996	0.994
	F-score	0.988	0.988	0.878	0.990	0.993	0.996	0.994
SZVAV	Accuracy	0.848	0.837	0.602	0.843	0.850	0.976	0.976
	F-score	0.793	0.767	0.722	0.790	0.783	0.975	0.975

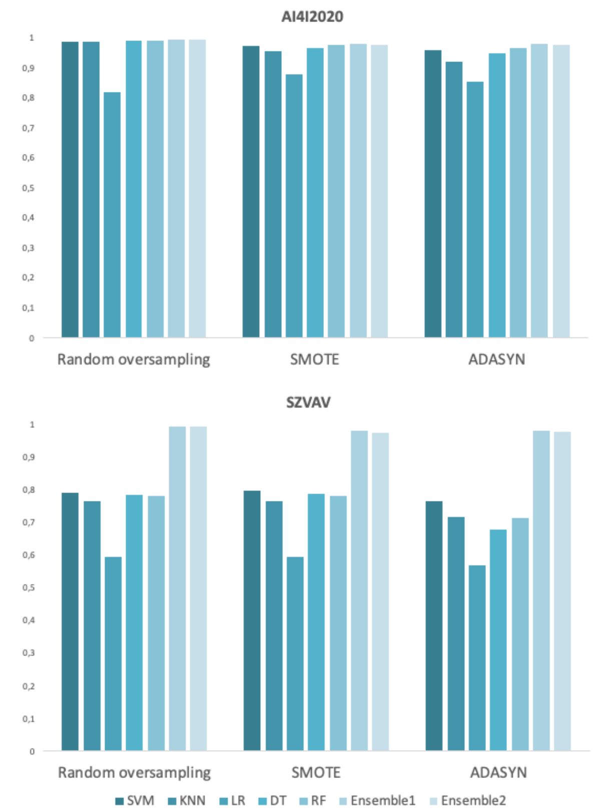

Figure 4. The test results for individual and proposed ensemble models based on different oversampling methods

To examine the usability of the proposed models in PdM applications, the tested models were evaluated in terms of training and testing times. The CPU times of the ensemble models and individual machine learning models proposed in the study are given in Table 7.

Table 7. CPU times of or individual and proposed ensemble models (sec).

	AI4I2020		SZVAV	
	Training	Test	Training	Test
SVM	2.028	0.065	16.499	0.163
KNN	0.018	0.064	0.145	0.594
LR	0.613	0.003	0.500	0.003
DT	0.044	0.002	0.054	0.194
RF	6.693	0.209	10.486	0.346
Ensemble1	28.260	0.589	8.882	0.563
Ensemble2	18.643	0.573	10.945	0.566

Although the training times of ensemble models are higher than individual models, they average around 16 seconds. This is due to the processes of generating random subsamples and combining or selecting the results of the individual models included in them. When the test times are examined, it is seen that the test times increase in the ensemble models due to the reasons set forth for the training periods. However, test times are around 0.5 seconds for ensemble models, which is a very short time for early detection of the failure.

4. Discussion

In addition to the CPU time comparison, we analyzed the possible true and false failure alarms that would occur assuming that the proposed ensemble models are implemented in a PdM application. Moreover, with this analysis, the failure rates that the developed models could not detect the failures were also revealed. The calculated ratios are visualized in Figure 5.

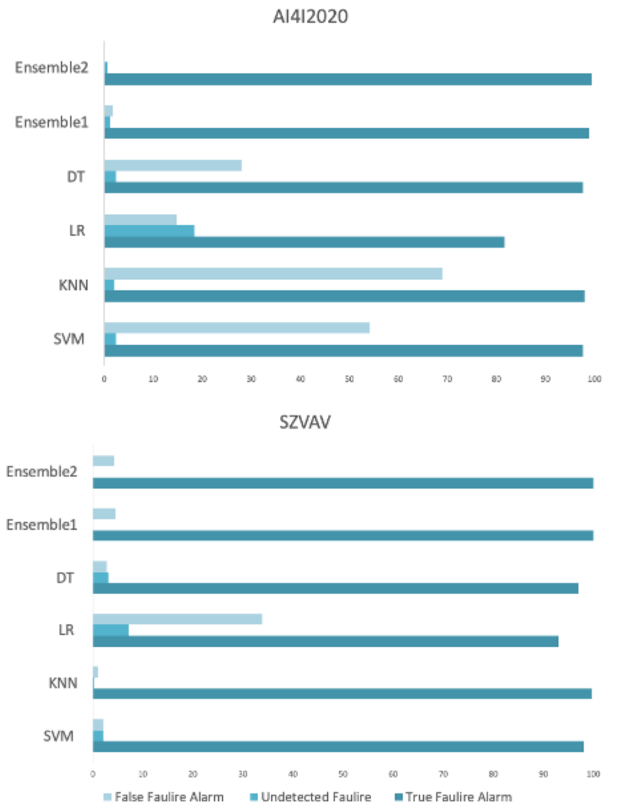

Figure 5. The test results for individual and proposed ensemble models based on different oversampling methods.

In the AI4I 2020 dataset, while reaching the best alarm levels with the proposed ensemble models; the undetected failure rate is very low. The false failure alarm rate has been greatly reduced for the AI4I2020 dataset. While 100% correct alarm rate was achieved for the SZVAV dataset; The undetected failure rate is 0%. False failure rates for the SZVAV dataset are already quite low in models other than LR, and although ensemble models seem to increase the false failure alarm rate, this rate is only 4%.

We reviewed previous studies using the same datasets in the literature to show the effectiveness of the proposed models performed. While there are more studies in the literature on the AI4I2020 dataset, studies on the SZVAV dataset are limited. In this regard, we compared the results obtained in the study with the results of the ensemble learning models we proposed through accuracy and F-score metrics. In a very recent study that we reached for this purpose, Liao *et al*. (2025) proposed a stacking model, Advanced Community Trees (AET), and compared the performance of this algorithm with the results of previous studies. In addition to this study, we reached one of the latest research conducted by Shaheen *et al*. (2023). In this paper, a novel machine failure prediction models, namely JRIP and JRIP-CSE, were proposed and tested with the AI4I 2020 predictive maintenance dataset. The comparison made based on the table in Liao *et al.* (2025) for the AI4I 2020 dataset is given in Table 8.

Table 8. The results of the comparison with a previous studies utilizing AI4I 2020 dataset.

Study	Model	Accuracy	F-score
This study	**Ensemble1**	**0.996**	**0.996**
	Ensemble2	**0.994**	**0.994**
Shaheen *et al.* (2023)	JRIP	0.984	0.983
	JRIP-CSE	0.985	0.984
Chen *et al*. (2022)	CatBoost	0.9867	-
	SmoteNC + CatBoost	0.9670	-
	ctGAN + Catboost	0.9082	-
	SmoteNC + ctGAN + Catboost	0.8712	-
Vuttipittayamongkol, *et al*. (2022)	SVM		0.6522
	DT		0.7766
	KNN		0.4348
	RF		0.7045
	NN		0.3359
Mota *et al*. (2022)	Gradient Boosting	0.9455	0.49
	SVM	0.9100	0.38
Iantovics *et al.* (2022)	Binary Logistic Regression	0.9710	0.4407
Torcianti *et al.* (2021)	RusBoost Trees	0.9274	0.4590
Matzka (2020)	Bagged Decision Trees	0.9834	0.9234
Ghasemkhani *et al.* (2023)	Balanced K-Star	0.9875	0.9875
Liao *et al*. (2025)	AET	0.9881	0.9881

The comparison results demonstrate that the proposed ensemble learning approach outperforms existing methods in failure prediction literature on the AI4I2020 Predictive Maintenance dataset.

As previously mentioned, studies utilizing the SZVAV dataset in the literature are limited. Since existing studies report only accuracy values, comparisons are based on this metric. In this context, Table 9 presents a comparison with two recent studies using this dataset.

Table 9. The results of the comparison with a previous studies utilizing SZVAV dataset.

Study	Model	Accuracy
This study	**Ensemble1**	**0.996**
	Ensemble2	**0.994**
Fan *et al.* (2024)	Parallel semi-supervised learning with active learning	0.7581
Fan *et al.* (2023)	Fully Connected Neural Network	0.75
	Graph Neural Network	0.66

Likewise, the proposed approach surpassed previous studies in the literature on the SZVAV dataset based on the accuracy metric.

One of the main advantages of the proposed approach in this study is addressing the class imbalanced problem. Unlike some previous studies where accuracy and F-score showed inconsistency due to class imbalance, our approach showed a more balanced performance between accuracy and f-score metrics for the considered datasets. These findings emphasize the usability of the proposed approach in real-world applications where the possibility of missing data problems is high. False failure alarms can cause many negative consequences such as unnecessary production interruptions, increased maintenance costs, and decreased system efficiency. The proposed ensemble model approach produces low false failure predictions with high performance. This makes it more suitable for real-world PdM applications.

Despite its strong performance, the tests show that the proposed ensemble models require higher computational power than individual machine learning models for training and testing.

5. Conclusion

In recent years, with increasing digitalization and Industry 4.0 applications, predictive maintenance applications have become widespread. Predicting failures in advance using data collected from machines via sensors prevents production losses.

In this study, individual and ensemble machine learning models were developed to predict machine failures in advance, considering various production characteristics. Performance evaluations revealed that the ensemble model, proposed using the bagging strategy, outperformed the individual models. Additionally, we assessed the proposed approach in terms of computational efficiency and compared it with existing studies in the literature, reporting the results. Our findings indicate that the proposed model demonstrates strong predictive performance.

In addition, an evaluation of the oversampling methods applied to solve the imbalance problem experienced in datasets containing events such as failures and accidents has been made, and the importance of balancing the dataset has been demonstrated once again.

The most significant limitation of the study is that the proposed approach is tested on one real-life and one synthetic dataset. To generalize the performance of our proposed ensemble models, we plan to test it on additional failure datasets. Additionally, we aim to enhance the accuracy and interpretability of the proposed ensemble models by applying a feature selection step as a preprocessing step. Future research can also include optimizing the running times of the proposed approach to be implemented in real-time PdM applications.

Ethical Statement

This study was developed as an extension of the undergraduate thesis titled "Ensemble Learning-Based Machine Learning Models for Predictive Maintenance Applications". Preliminary results obtained using synthetic datasets were presented as an abstract at the 5th International Conference on Sustainable Science and Technology (ICSuSaT-2022). The authors declare that this research was conducted in accordance with ethical principles and academic integrity standards.

Credit Authorship Contribution Statement

Author-1: Investigation, Methodology, Software, Validation, Visualization, Funding acquisition.
Author-2: Investigation, Methodology, Software, Validation, Visualization.
Author-3: Conceptualization, Methodology, Software, Validation, Visualization, Writing - original draft, Funding acquisition.

Declaration of Competing Interest

The authors have no conflicts of interest to declare regarding the content of this article.

Data Availability Statement

The data that support the findings of this study are openly available in references (U. M. L. Repository, 2020), (Granderson, 2019).

Acknowledgements

This study is supported by Eskisehir Technical University, Turkey Scientific Research Projects Committee (22LOP171) and TUBITAK 2209-A - Research Project Support Programme for Undergraduate Students.

References

Akgül, G., Çelik, A. A., Ergül Aydın, Z. and Kamışlı Öztürk, Z., 2020, Hipotiroidi hastalığı teşhisinde sınıflandırma algoritmalarının kullanımı. *Bilişim Teknolojileri Dergisi,* **13,** 255 – 268.

https://doi.org/10.17671/gazibtd.710728.

Andre, A. B., Beltrame, E. and Wainer, J., 2013. A combination of support vector machine and k-nearest neighbors for machine fault detection. *Applied Artificial Intelligence,* **27,** 36–49. https://doi.org/10.1080/08839514.2013.747370

Arslan, B. and Tiryaki, H., 2020. Prediction of railway switch point failures by artificial intelligence methods. *Turkish Journal of Electrical Engineering and Computer Sciences,* **28,** 1044–1058. https://doi.org/10.3906/elk-1906-66

Ay, A.K. and Yolacan, E. N., 2022. Yeniden Örnekleme Metotlarının Kredi Kartı Sahtecilik Tespiti için Topluluk Öğrenmesine Kapsamlı Analizi. *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi,* **22**, 1005-1015. https://doi.org/10.35414/akufemubid.1066453

Aydın, M. A., 2022. Müşteri kaybı tahmininde sınıf dengesizliği problemi, *Politeknik Dergisi*, **25,** 351 – 360. https://doi.org/10.2339/politeknik.734916

Baptista, M., Sankararaman, S., Medeiros, I. P. de, Nascimento, C., Prendinger, H. and Henriques, E. M., 2018. Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. *Computers Industrial Engineering,* **115**, 41–53. https://doi.org/10.1016/j.cie.2017.10.033

Beretta, M., Vidal, Y., Sepulveda, J., Porro, O. and Cusidó, J., 2020. Improved Ensemble Learning for Wind Turbine Main Bearing Fault Diagnosis. *Applied Sciences*, **11(16),** 7523. https://doi.org/10.3390/app11167523

Bousdekis, A., Lepenioti, K., Apostolou, D. and Mentzas G. 2019, Decision making in predictive maintenance: Literature review and research agenda for industry 4.0. *IFAC-PapersOnLine*, **52,** 607–612. https://doi.org/10.1016/j.ifacol.2019.11.226

Breiman, L., 2001. Random forests. *Machine Learning* **45,** 5–32. https://doi.org/10.1023/A:1010933404324

Bukhsh, Z.A., Saeed, A., Stipanovic, A. and Doree, A. G., 2019. Predictive maintenance using tree-based classification techniques: A case of railway switches. *Transportation Research Part C: Emerging Technologies*, **101**, 35–54. https://doi.org/10.1016/j.trc.2019.02.001

Cakir, M., Guvenc, M. A. and Mistikoglu, S., 2021. The experimental application of popular machine learning algorithms on predictive maintenance and the design of iot based condition monitoring system. *Computers Industrial Engineering*, **15,** 106948. https://doi.org/10.1016/j.cie.2020.106948

Chen, C. H., Tsung, C. K. and Yu, S. S., 2022. Designing a hybrid equipment-failure diagnosis mechanism under mixed-type data with limited failure samples. *Appl Sci*, **12**, 9286. https://doi.org/10.3390/app12189286

Chen, C., Zhu, Z. H., Shi, J., Lu, N. and Jiang, B., 2021. Dynamic predictive maintenance scheduling using deep learning ensemble for system health prognostics. *IEEE Sensors Journal,* **21**, 26878–26891. https://doi.org/10.1109/JSEN.2021.3119553.

Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine Learning,* **20 (3)**, 273–297. https://doi.org/10.1007/BF00994018

Demir, I. and Karaboga, H.A., 2021. Modeling mathematics achievement with deep learning methods. *Sigma J Eng Nat Sci,* **39**, 33–410. https://doi.org/10.14744/sigma.2021.00039

Dundar, D. R. , Saricicek, I., Cinar, E. and Yazici, A., 2021. Kestirimci bakimda makine ogrenmesi: Literatür arastirmasi. *ESOGÜ Müh Mim Fak Derg*, **29,** 256–276. https://doi.org/10.31796/ogummf.873963

Erzurum Cicek, Z. I. and Kamisli Ozturk, Z, 2022. Prediction of fatal traffic accidents using one-class SVMs: a case study in Eskisehir, Turkey*, International Journal of Crashworthiness,* **27 (5)**, 1433–1443. https://doi.org/10.1080/13588265.2021.1959168

Fan. C., Lin, Y., Piscitelli, M. S., Chiosa, R., Wang, H., Capozzoli, A. and Ma, Y., 2023. Leveraging graph convolutional networks for semi-supervised fault diagnosis of HVAC systems in data-scarce contexts. *Building Simulation*, **16**, 1499-1517. https://doi.org/10.1007/s12273-023-1041-1

Fan, C., Wu, Q., Zhao, Y. and Mo. L., 2024. Integrating active learning and semi-supervised learning for improved data-driven HVAC fault diagnosis performance. *Applied Energy*, **356**, 122356. https://doi.org/10.1016/j.apenergy.2023.122356

Fernandes, S., Antunes, M., Santiago, A. R., Barraca, J. P., Gomes, D., Aguiar, R. L., 2020. Forecasting appliances failures: A machine-learning approach to predictive maintenance. *Information,* **11(4)**, 208. https://doi.org/10.3390/info11040208.

Gencer, A., Yumusak, R., Ozcan, E. and Eren, T., 2021. An artificial neural network model for maintenance planning of metro trains. *Journal of Polytechnic,* **24,** 811–820. https://doi.org/10.2339/politeknik.693223

Genuer, R., Poggi, J.-M. and Tuleau-Malot, C., 2010. Variable selection using random forests, *Pattern Recognition Letters,* **31,** 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014

Ghasemkhani, B., Aktas, O. and Birant, D., 2023. Balanced k-star: An explainable machine learning method for internet-of-things- enabled predictive maintenance in manufacturing. *Machines*, **11**, 322. https://doi.org/10.3390/machines11030322

Gohel, H. A., Upadhyay, H., Lagos, L., Cooper, K. and Sanzetenea, A., 2020. Predictive maintenance architecture development for nuclear infrastructure using machine learning, *Nuclear Engineering and Technology,* **52**, 1436–1442. https://doi.org/10.1016/j.net.2019.12.029.

Gungor, O., Rosing, T. and Aksanli, B., 2022. Stewart: Sacking ensemble for white-box adversarial attacks towards more resilient data-driven predictive maintenance. *Computers in Industry,* **140**, 103660. https://doi.org/10.1016/j.compind.2022.103660.

Gungor, O., Rosing, T.S. and Aksanli, B., 2022. Dowell: Diversity-induced optimally weighted ensemble learner for predictive maintenance of industrial internet of things devices. *IEEE Internet of Things Journal*, **9,** 3125–3134. https://doi.org/10.1109/JIOT.2021.3097269.

Hung, Y.-H., 2021. Improved ensemble-learning algorithm for predictive maintenance in the manufacturing process, *Applied Sciences*, **11(15)**, 6832. https://doi.org/10.3390/app11156832.

Iantovics, L.B. and Enachescu, C., 2022. Method for data quality assessment of synthetic industrial data. *Sensors*, **22,** 1608. https://doi.org/10.3390/s22041608.

Janssens, O., Loccufier, M. and Van Hoecke, S., 2019. Thermal imaging and vibration-based multisensor fault detection for rotating machinery. *IEEE Transactions on Industrial Informatics*, **15**, 434–444. https://doi.org/10.1109/TII.2018.

Kaya, Y., Kuncan, F. and Ertunc, H. M., 2022. A new automatic bearing fault size diagnosis using time-frequency images of cwt and deep transfer learning methods. *Turkish Journal of Electrical Engineering and Computer Sciences*, **30,** 1851–1867. https://doi.org/10.55730/1300-0632.3909.

Khalil, A. F. and Rostam, S., 2024. Machine Learning-based Predictive Maintenance for Fault Detection in Rotating Machinery: A Case Study. *Engineering, Technology & Applied Science Research,* **14**, 13181-13189. https://doi.org/10.48084/etasr.6813

Khan, P. W., Yeun, C. Y. and Byun, Y. C., 2023. Fault detection of wind turbines using SCADA data and genetic algorithm-based ensemble learning. *Engineering Failure Analysis*, **148,** 107209. https://doi.org/10.1016/j.engfailanal.2023.107209

Lee, W. J., Wu, H., Yun, H., Kim, H., Jun, M.B., and Sutherland, J. W., 2019. Predictive maintenance of machine tool systems using artificial intelligence techniques applied to machine condition data. *Procedia CIRP*, **80**, 506-511.

https://doi.org/10.1016/j.procir.2018.12.019

Liao, Y., Li, M., Sun, Q. and Li P., 2025. Advanced stacking models for machine fault diagnosis with ensemble trees and SVM. *Applied Intelligence*, **55**, 251.

https://doi.org/10.1007/s10489-024-06206-2.

Matzka, S., 2020. *Explainable artificial intelligence for predictive maintenance applications*. 2020 Third international conference on artificial intelligence for industries (ai4i). 69–74. https://doi.org/10.1109/AI4I49448.2020.00023.

Mei, Y., Sun, Y., Li, F., Xu, X., Zhang, A. and Shen, J., 2022. Probabilistic prediction model of steel to concrete bond failure under high temperature by machine learning. *Engineering Failure Analysis,* **142,** 106786.

https://doi.org/10.1016/j.engfailanal.2022.106786.

Mienye, I.D. and Sun, Y., 2022. A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access,* **10,** 99129–99149.

https://doi.org/10.1109/ACCESS.2022.3207287.

Mohammed, R., Rawashdeh, J. and Abdullah, M., 2020. *Machine learning with oversampling and undersampling techniques: Overview study and experimental results*. 2020 11th International Conference on Information and Communication Systems (ICICS). 243–248. https://doi.org/10.1109/ICICS49469.2020.239556.

Mota, B., Faria, P. and Ramos, C., 2022. *Predictive maintenance for maintenance-effective manufacturing using machine learning approaches.* International workshop on soft computing models in industrial and environmental applications. 13– 22. https://doi.org/10.1007/978-3-031-18050-7_2

Mujib, A. and Djatna, T., 2020 *Ensemble learning for predictive maintenance on wafer stick machine using IoT sensor data*. 2020 International Conference on Computer Science and Its Application in Agriculture (ICOSICA). 1-5. https://doi.org/10.1109/ICOSICA49951.2020.9243180.

Patra, K., Sethi, R. N. and Behera, D. K., 2022. Anomaly detection in rotating machinery using autoencoders based on bidirectional LSTM and GRU Neural Networks. *Turkish Journal of Electrical Engineering and Computer Sciences,* **30,** 1637–1653. https://doi.org/10.55730/1300-0632.3870

Phillips, J., Cripps, E., Lau, J. W. and Hodkiewicz, M., 2015. Classifying machinery condition using oil samples and binary logistic regression. *Mechanical Systems and Signal Processing,* **60-61**, 316–325. https://doi.org/10.1016/j.ymssp.2014.12.020

Raschka, S., 2015. Python Machine Learning, Packt Publishing Ltd.

Raza, J., Liyanage, J. P., Al Atat, H. and Lee, J., 2010. A comparative study of maintenance data classification based on neural networks, logistic regression and support vector machines. *Journal of Quality in Maintenance Engineering*, **16,** 303–318. https://doi.org/10.1108/13552511011072934.

Saihood, Q. and Sonuc E., 2023. A practical framework for early detection of diabetes using ensemble machine learning models. *Turkish Journal of Electrical Engineering and Computer Sciences,* **31**, 722–738. https://doi.org/10.55730/1300-0632.4013.

Shamayleh, A. J. F. and Awad, M., 2020. Iot based predictive maintenance management of medical equipment. *J Med Syst.*, **44**, 72. https://doi.org/10.1007/s10916-020-1534-8.

Shaheen, A., Hammad, M., Elmedany, W., Ksantini, R. and Sharif, S., 2023. Machine failure prediction using joint reserve intelligence with feature selection technique. *International Journal of Computers and Applications,* **45**, 638–646. https://doi.org/10.1080/1206212X.2023.2260619.

Shashidhar Kaparthi, D. B., 2020. Designing predictive maintenance systems using decision tree-based machine learning techniques. *International Journal of Quality & Reliability Management,* **37, 4,** 659-686. https://doi.org/10.1108/IJQRM-04-2019-0131.

Torcianti, A. and Matzka, S., 2021. *Explainable artificial intelligence for predictive maintenance applications using a local surrogate model.* 2021 4th International conference on artificial intelligence for industries (ai4i), Laguna Hills, CA, USA, 86–88. https://doi.org/10.1109/AI4I51902.2021.00029.

Vuttipittayamongkol, P. and Arreeras, T., 2022. *IEEE Data-driven industrial machine failure detection in imbalanced environments*. 2022 IEEE international conference on industrial engineering and engineering management (IEEM), Kuala Lumpur, Malaysia, 1224–1227. https://doi.org/10.1109/IEEM55944.2022.9989673

Wu, H., Huang, A. and Sutherland, J. W., 2020. Avoiding environmental consequences of equipment failure via an LSTM-based model for predictive maintenance. *Procedia Manufacturing*, **43,** 666–673. https://doi.org/10.1016/j.promfg.2020.02.131

Zhang, L., Cheng, Y., Zhang, J., Chen, H., Cheng, H., and Gou, W., 2023. Refrigerant charge fault diagnosis strategy for VRF systems based on stacking ensemble learning. *Building and Environment*, **234**, 10209. https://doi.org/10.1016/j.buildenv.2023.110209.

Zhang, M., Ge, W., Tang, R. and Liu, P., 2023. Hard Disk Failure Prediction Based on Blending Ensemble Learning. *Applied Sciences,* **13**, 3288. https://doi.org/10.3390/app13053288.

Zhu, T., Ran, Y. and Wen, Y., 2019. A Survey of Predictive Maintenance: Systems, Purposes and Approaches-Arxiv.org. https://doi.org/10.48550/arXiv.1912.07383

Zonta, T. , Costa, C. A. da, Rosa Righi, R. da, Lima, M. J. de, Trindade, E. S. da and Li, G. P., 2020. Predictive maintenance in the industry 4.0: A systematic literature review. *Computers Industrial Engineering,* **150,** 106889. https://doi.org/10.1016/j.cie.2020.106889.

Internet References

AI4I 2020 Predictive Maintenance Dataset [Dataset]. (2020). UCI Machine Learning Repository. https://doi.org/10.24432/C5HS5C.

Granderson, J. G. L., 2019. Inventory of datasets for afdd evaluation, https://data.openei.org/files/910/lbnldatasynthesisinventory.pdf