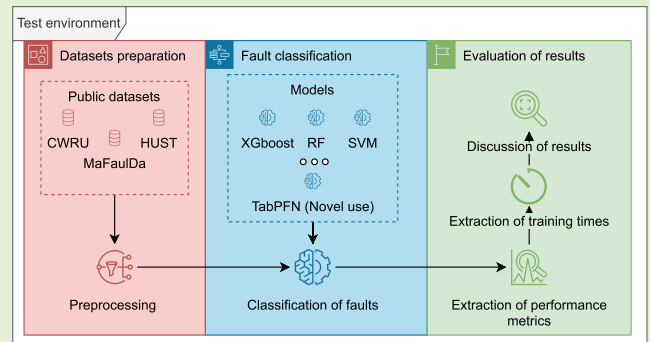


Early Fault Classification in Rotating Machinery With Limited Data Using TabPFN

L. Magadán¹, J. Roldán-Gómez¹, J. C. Granda¹, and F. J. Suárez¹

Abstract—Intelligent fault detection and classification is a cornerstone of prognostic and health management of rotating machinery (RM) research. Correctly classifying and predicting RM faults not only increases productivity in industrial plants but also reduces maintenance costs. The datasets from real facilities needed to train fault classifiers often have few samples due to the expense of provoking faults in real scenarios to obtain data. This article proposes the use of the tabular prior-data fit network (TabPFN) model for the classification of faults in RM. TabPFN is a model which has been pretrained with a large amount of synthetic data with many causal relationships. This allows the model to perform Bayesian inference on the data used for training. The advantages of this model are its ability to be trained with limited data without generating overfitting problems and its high speed (if a graphics processing unit (GPU) is available). To compare its performance with traditional algorithms for tabular classification such as XGboost and random forest, three public datasets were used. Results show that TabPFN performs more accurately than algorithms with limited data, so it is suitable to be deployed in real scenarios when the amount of data available from the monitored RM is limited.

Index Terms—Fault classification, IIoT, predictive maintenance, rotating machinery (RM), tabular prior-data fit network (TabPFN).



I. INTRODUCTION

THE fast growth of industrialization requires increasingly safe, accurate, and effective machinery. Approximately 40% of electro-mechanical and mechanical systems include rotating machinery (RM). These fundamental elements may trigger serious faults due to the conditions where they are installed and the long operating times [19]. Depending on where these faults occur, they may be minor or major, but they always entail a risk. Between 40% and 50% of RM faults are caused in bearings. The second most common faults are stator-related, accounting for 30%–35% of the total. Rotor faults account for approximately 10%. The remaining faults are caused principally by looseness or gear faults [7], [15].

Manuscript received 5 July 2023; revised 24 October 2023; accepted 4 November 2023. Date of publication 14 November 2023; date of current version 14 December 2023. This work was supported in part by the Spanish National Plan of Research, Development, and Innovation through the Project EDNA under Grant PID2021-124383OB-I00; and in part by the University of Oviedo. The work of L. Magadán was supported by the Severo Ochoa Program under Grant PA-22-BP21-120. The associate editor coordinating the review of this article and approving it for publication was Prof. Weihua Li. (Corresponding author: L. Magadán.)

The authors are with the Department of Computer Science and Engineering, University of Oviedo, 33204 Gijón, Spain (e-mail: magadanluis@uniovi.es; roldangiose@uniovi.es; jcgranda@uniovi.es; fjsuarez@uniovi.es).

Digital Object Identifier 10.1109/JSEN.2023.3331100

Maintenance accounts for 15%–60% of the total production costs [30]. Predictive maintenance is the best alternative in terms of both cost and performance, carrying out maintenance tasks only when a fault has been detected, preferably at an early stage [13]. Predictive maintenance can increase productivity by 20%, and the life of RM by up to 50%. Maintenance costs can be reduced by 50% [6]. Detecting when an RM is about to fail requires continuous monitoring, and it depends on many operational factors, such as the rotating speed, the load it supports, the type of bearings, and all the nearby machines [14]. Early detection of faults is vital, leading to a considerable increase in the availability and productivity of machinery.

Fault classification enables proactive maintenance strategies, such as condition monitoring and predictive maintenance, by providing insights into the health status of RM [24]. Accurate fault classification advances the field of machinery design and manufacturing, facilitating the development of more robust and reliable RM with improved fault tolerance and enhanced performance [12].

Data-driven models have emerged as a promising approach for fault classification in RM, analyzing large amounts of sensor data and identifying patterns for different fault types. Data-driven models need a large amount of data for training. In RM fault classification, this must contain RM faults, which

is difficult to collect in real environments. A further complication is that RM behavior is not always consistent as RM can work under different loads or rotating speed, and faults may manifest themselves differently under different operating conditions. The limited amount of data available is especially problematic with recently installed RM. Overfitting is another problem with data-driven models, especially when dealing with limited data, as models tend to memorize noise rather than learn true patterns [29]. Overcoming these challenges is crucial to ensure the reliability and effectiveness of data-driven fault classification models in real-world applications.

Although there are different data-driven approaches to fault classification in RM, one of the most widely used is deep learning [10], [11]. *Gradient-flow-based meta generative adversarial network (GAN) for data augmentation in fault diagnosis* by Wang et al. [25] addresses the scarcity of labeled data. The authors propose using a flow-based meta generative adversarial network (GFMGAN) and support this model with single natural image (SinGAN), which deploys generative models with a single natural image [23]. The distribution of data is captured and characterized as vibrations by concatenating training samples and converting them into an image. The generative model obtained is used to generate new samples during the data augmentation process. To improve the ability to generate the model, they introduce a gradient-flow-based meta-learning technique. Finally, they use a convolutional neural network (CNN) to perform the classification. The authors validate their proposal with two types of bearing faults: inner race and outer race. A minor limitation of the proposal is that if a new type of fault is included, it is necessary to repeat the whole training process including the data augmentation process with the new fault. The characterization of vibrations as images may limit the deployment of this system in real-time.

Another work that uses CNN for the classification of faults is *CFCNN: A novel convolutional fusion framework for collaborative fault identification of RM* by Xu et al. [28]. The authors extract multilevel features from the different vibrations and integrate a module to merge these features using the correlations between them. Then, a smoothing mechanism is used to reduce possible overfitting when training the model. Finally, the authors validate their model with a cylindrical rolling bearing dataset and a planetary gearbox dataset. This work integrates vibrations from different sources, although it does not solve the problem of the limited datasets in this field. The inclusion of a smoothing mechanism to avoid overfitting is a step in the right direction, but may be insufficient when classifying more types of faults.

An intelligent fault diagnosis method of small sample bearing based on an improved auxiliary classification GAN by Meng et al. [16] uses two different deep learning architectures to solve the problem of data scarcity. The authors use a GAN to generate data and perform the data augmentation process. This neural network introduces the Wasserstein distance in the cost function in a way that alleviates the vanishing gradient problem. In addition, an attention mechanism is introduced that focuses on the blocks obtained with the convolutions. The features obtained by the attention mechanism are merged with

the convolutions to classify faults. The method is tested with the public bearing dataset of Case Western Reserve University and the bearing simulation dataset of the Yanshan University Laboratory. Although the proposal is very interesting, it has certain drawbacks. First, GANs are very prone to overfitting, especially with limited data, which can cause them to generate very similar elements. They are also very sensitive to hyperparameter selection, because a bad balance between the generator and the discriminator can result in learning nothing or always generating similar samples. Finally, the authors use accuracy as a metric in imbalanced datasets, although other metrics usually fit better in this type of datasets.

There are also proposals that focus on the methods to characterize data, such as *a visual vibration characterization method for intelligent fault diagnosis of RM* by Peng et al. [20], which measures vibrations using images. These vibrations are obtained through the phase difference in the image. This means that the differences between images are obtained with their maximum frequencies. A video of the machinery while working is recorded and this technique is performed between different frames. This makes it possible to characterize the vibrations as an image directly, without transforming the vibrations into images, as in previous works. Once the vibrations are characterized as image differences, a CNN is trained to classify and detect faults. The novelty of this work lies in the elimination of the signal-to-image transformation process for the use of a subsequent CNN, but the authors do not propose a mechanism to solve the problem of limited training data.

It is common to use CNNs and characterize vibrations as images. However, there are other approaches to data characterization, such as the graphs used in *fault diagnosis of rolling bearing based on knowledge graph with data accumulation strategy* by Xiao et al. [27]. In this work, the nodes contain information about the different extracted features, and the edges contain the feature-fault correlation. This representation constantly updates the model with new data using weighted random forest, which classifies the faults and also updates the network information when training with more samples. The strength of this proposal is that it generates a network structure for the characterization, which enables incremental training. However, it does not offer solutions for datasets with few samples: the recurring problem in this field.

Different data-driven approaches to characterization have been tested, and different data augmentation methods have been developed to address the problem of the scarcity of labeled data and consequently the overfitting of most of the data-driven approaches. However, to the best of the authors' knowledge, there is no work that proposes using a tabular prior-data fit network (TabPFN) for fault classification. This article proposes an early fault classification method using TabPFN as a data-driven model to classify different RM operating conditions when very little data are available, such as in new installations or after changes in the RM operating configuration. The performance of the model has been analyzed using three different datasets and its response tested as the size of the training set decreases and the size of the test set increases. The results show that the proposed model performs

better than the traditional machine learning (ML) algorithms when the amount of available data is limited. Furthermore, the model makes better predictions without overfitting and adjusting hyperparameters, training in less than a second using graphics processing unit (GPU). This makes the proposed classification method suitable for implementation with newly installed or reconfigured RM.

The rest of this article is organized as follows. The key concepts of the research are outlined in Section II. Section III describes the environment used for testing. A description of the tests carried out and the results obtained by the proposed solution are shown and discussed in Section IV. Finally, the concluding remarks and future work are outlined in Section V.

II. BACKGROUND

In this section, the key concepts for the development of this work are described. In particular, the different RM faults to be classified are analyzed, and TabPFN and its advantages for the classification of these faults are described.

A. RM Faults

RM refers to any device that rotates, converting electrical energy into rotational kinetic energy. It has two main components: the stator, which is stationary, and the rotor, which moves. The bearing, which enables relative motion between the rotating and stationary parts, is essential to the operation of RM. To ensure consistent productivity, RM must operate continuously, with high reliability and no breakdowns. This work uses datasets in which the major contributors to RM faults are present: unbalance, misalignment, and bearing faults [5]. Misalignment can be vertical or horizontal and there are several types of bearing faults depending on where they occur. Each of these faults is explained in more detail below.

- 1) Unbalance occurs when there is an uneven distribution of mass within the rotating component of the machine. Unbalance is usually caused by defects in manufacturing, wear, or improper assembly. This leads to excessive centrifugal forces during operation, generating vibrations that affect performance and can be a risk for the entire system [4]. If they are not properly fixed, they can produce significant damage, reducing the remaining useful lifetime of RM and in the worst cases causing catastrophic damage. Vibrations can also have detrimental effects on the surrounding equipment and structures.
- 2) Horizontal misalignment happens when the centers of rotation of the driving and driven components are not perfectly aligned along the horizontal axis. It commonly appears due to installation errors, thermal expansion, or incorrect placement of the machine on its base [21]. If not properly addressed, it can cause excessive friction and accelerated wear on horizontal shafts and bearings, leading to reduced energy efficiency and increased power consumption. To rectify it, precise alignment techniques such as laser alignment or dial indicator measurements are used.
- 3) Vertical misalignment is similar to horizontal misalignment. It occurs when the centers of rotation of the

driving and driven components are not aligned, but in this case along the vertical axis. Both types of misalignment lead to similar issues: increased forces, vibrations, and strains [21]. However, vertical misalignment affects vertical components and can cause an increased stress on thrust bearings. Like horizontal misalignment, vertical misalignment can be rectified using precise alignment techniques.

- 4) Outer race bearing faults refer to the breakdown or malfunction of the outer race of bearings. The outer race is the outer ring of the bearing, which encloses the rotating elements that support and guide the rotating shaft [26]. These faults usually appear due to misalignment, improper lubrication, or excessive loads. They can lead to severe consequences in RM, causing excessive vibrations and noise, and reducing productivity and efficiency. When outer race bearing faults are not properly addressed, the fault can cause damage to other components such as the shaft or the housing of the RM. They can also lead to damage in external components or structures that are connected to the bearing. Regular maintenance, proper lubrication, and condition maintenance are usually used to prevent them.
- 5) Inner race bearing faults are similar to outer race bearing faults, but they appear in the inner ring of bearings [26]. The inner ring interacts directly with the rotating elements, facilitating smooth rotation. These faults are caused by excessive loads, misalignment, improper lubrication, or fatigue over time. They can cause excessive vibrations and noise, as well as damage to the shaft or housing of the RM, reducing RM performance.
- 6) Ball bearing faults are the breakdown or malfunction of the rotating elements of the bearing, which are metal balls or cylinders that roll between the inner and outer races of bearings [26]. They are used to reduce friction and enable smooth rotation of shafts. When these faults occur, they can lead to increased friction, excessive heat generation and vibrations, reducing the RM efficiency, and increasing energy consumption. They can also affect other parts such as the shaft.

B. Tabular Prior-Data Fit Network

TabPFN is a prior-data fit network (PFN) [17] designed to perform supervised classification tasks on tabular data [8]. This model proposes a radical paradigm shift with respect to other current models in the state of the art. TabPFN provides a transformer that has been trained to perform Bayesian inference. For this purpose, single causal models are generated and Bayesian neural networks are used. The objective is to train the PFN model to approximate the posterior predictive distribution of the data to be predicted as closely as possible. The goal is for the transformer to perform Bayesian inference for a wide range of causal relationships. When trained on a different dataset, such a model is able to approximate the posterior predictive distribution of new data in a single pass.

While classical approaches, such as neural networks, focus on finding the distribution of the data based on each of the

samples, TabPFN focuses on finding the best final distribution given the initial distribution of the training data. Since TabPFN has been pretrained with synthetic datasets that hold simple causal relationships, TabPFN searches for approximations to such causal relationships [18].

The experiments carried out by the authors seem to conclude that a key difference with respect to other state-of-the-art models is that TabPFN does not incur overfitting when the dataset is small. This is because it is pretrained with a large number of datasets that collect different types of causal relationships. Therefore, the model is able to achieve an approximation of the final distribution of the new training data, as long as the initial distribution of the new training data follows the patterns it extracts from the pretraining data. It is also capable of training without hyperparameter optimization, obtaining very good training times, especially if a GPU is used.

However, TabPFN has a number of limitations. It can deal with 1024 training samples, ten different families, and 100 features at most. These limitations are due to the fact that the execution times of the PFNs scale quadratically based on the training data and the need to keep model training times short. A softer limitation is that TabPFN degrades its performance with missing values or categorical features. One example of categorical feature is the lubrication level, which can be low, normal, or excessive. Another limitation of TabPFN is that it may fail to generalize when the data contain very complex causal relationships, since it has not been pretrained with data with such relationships.

These limitations mean that TabPFN has very specific requirements. When these requirements are met, it is a viable alternative that does not require hyperparameter optimization and allows training in less than 1 s.

III. TEST ENVIRONMENT

The test environment was configured with an Intel¹ Core² i7-9750H CPU (2.6 GHz), 16-GiB RAM, and an NVIDIA GeForce GTX 1650 (4 GiB) graphic card. The programming language used was Python 3.10 using the PyCharm framework. The proposed model was tested using three different datasets to evaluate its performance under different RM conditions and fault types. These tests were repeated with different traditional ML classification models: random forest (RF), support vector machine (SVM), gradient boosting (GB), multilayer perceptron (MLP), eXtreme gradient boosting (XGB), and label propagation (LP). The results obtained by TabPFN were compared with those of these traditional ML classification models.

The datasets were divided into the training and testing sets, varying the size of the training set. For each training set size, each ML model was trained with the same training set and tested with the same testing set to compare their performance. If the models are trained with only one training set and tested with only one testing set, the results obtained may be biased by the training set used. For this reason, cross-validation was carried out, repeating each test 50 times, varying the

TABLE I
EXTRACTED TIME-DOMAIN FEATURES

Time-domain feature	Equation
Mean	$\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$
Standard Deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}}$
Minimum	$\min_X = \min(X)$
Maximum	$\max_X = \max(X)$
Kurtosis	$Kurt = \frac{1}{N} \sum_{i=1}^N \left[\frac{(X_i - \bar{X})}{\sigma} \right]^4$
Skewness	$Skew = \frac{1}{N} \sum_{i=1}^N \left[\frac{(X_i - \bar{X})}{\sigma} \right]^3$
Root mean square	$RMS = \sqrt{\frac{\sum_{i=1}^n x_i^2}{N}}$

training and testing sets and computing the mean of each of the evaluation metrics. In addition, different experiments were performed where the number of samples of the training set was increased progressively, from 2% to 50% of the whole dataset, using the remaining data in the dataset as the corresponding testing set. The objective was to test how the different algorithms behave with small training datasets.

The first step is to extract a total of seven time-domain features from the raw vibration data of each vibration axis. These time-domain features and the formulas used to extract them are shown in Table I. Next, each of the datasets is divided into several training and testing sets, using different sizes of the training set. All the traditional classification methods were previously configured by manually adjusting their hyperparameters to obtain the best possible results. In contrast, TabPFN does not need hyperparameter optimization.

To measure the performance achieved by TabPFN and compare it with that of the traditional classification methods, each of them was analyzed as a multiclass classifier. True positives (TP) are obtained when the classifier correctly identifies an instance as belonging to a specific class, making the prediction correctly. False positives (FP) occur when the classifier identifies an instance as belonging to a particular class when it actually belongs to another, making an incorrect prediction. True negatives (TN) refer to instances that are correctly classified as not belonging to a specific class. Finally, false negatives (FN) appear when the classifier identifies an instance as belonging to a specific class, but the actual class label indicates that the prediction is wrong. The following metrics were used to evaluate the performance of the tests.

- 1) *Accuracy*: This metric measures the overall correctness of the classification model by calculating the proportion of correctly classified instances out of the total number of instances in the dataset. It provides an assessment of the ability of the model to correctly classify both the positive and negative instances. Accuracy can

¹Registered trademark.

²Trademarked.

be computed using the following equation:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (1)$$

- 2) *Precision*: This metric measures the ability of the classification model to avoid FP. Precision is computed with the following equation:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (2)$$

- 3) *Recall*: This metric measures the ability of the classification model to avoid FN. It is a very important metric as it tests the possibility allowing faults to appear in the RM without being detected, which could cause a reduction in productivity and even the complete breakdown of the RM. Recall is computed using the following equation:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (3)$$

- 4) *F1-Score*: This is used to provide a more global view of the performance of the classification model. It is the most balanced metric as it takes into account both precision and recall, working well in cases of imbalanced datasets [22]. The F1-score is computed with the following equation:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (4)$$

A brief description of each of the datasets is given below. These datasets have been selected because they are imbalanced and exhibit different types of failures in RM.

A. CWRU Dataset [1]

The CWRU dataset is one of the most commonly used datasets in bearing fault classification. The test rig used for this dataset consists of a motor, a token converter, an encoder, and a dynamometer. Three vibration sensors are placed at different locations (drive end, basement, and fan end) to collect vibration time-series data.

CWRU simulates the system working under different operating states: healthy, inner race bearing fault, ball bearing fault, and outer race bearing fault. Data were collected at different rotating speeds, and each of the faulty operating conditions was simulated with different levels of severity. Drive end bearing faults were collected at a frequency of 12 and 48 kHz, while fan end and basement bearing faults were collected only at 12 kHz. For healthy operating conditions, data were collected at 48 kHz.

During the tests, only healthy operating conditions and drive end bearing faults were analyzed. Data collected from the sensor were as segmented into 10 248 data points to generate samples from which to extract the time-domain features shown in Table I. After preprocessing this dataset, a total of 2208 samples composed of seven time-domain features were extracted, corresponding to healthy operating conditions, inner race bearing faults, outer race bearing faults, and ball bearing faults.

B. MaFaulDa Dataset [2]

The MaFaulDa dataset is another commonly used dataset in fault classification in RM. It comprises various types of faults at different rotating speeds and levels of severity. It consists of multivariate time-series data collected from a machinery fault simulator test rig. The machinery fault simulator simulates the system working under different states: healthy, unbalanced, horizontal misalignment, vertical misalignment, inner race bearing fault, ball bearing fault, and outer race bearing fault.

The dataset is composed of 1951 files including vibration data collected from two three-axis accelerometers (axial, radial, and tangential), the rotating speed, and the noise around the system. Each data sample was collected at a frequency of 50 kHz for 5 s.

The dataset was preprocessed to extract the time-domain features shown in Table I for each vibration axis. After preprocessing the dataset, a total of 1951 samples were extracted, each composed of 42 time-domain features (seven time-domain features per axis per sensor).

C. HUST Dataset [9]

This recently published dataset consists of vibration data gathered at 51.2 kHz from different bearings under different operating conditions. There are a total of 90 raw vibration data sample files, including four types of bearing states (healthy, inner race, outer race, and ball bearing faults) on five different bearings, under three different working conditions.

During the tests, all the bearings were taken into account. Data collected from the sensor were segmented into 18 000 data points to generate samples that were used to extract the time-domain features indicated in Table I. Once this dataset was preprocessed, a total of 2036 samples composed of seven time-domain features were extracted, corresponding to healthy operating conditions, inner race bearing faults, outer race bearing faults, ball bearing faults, and both inner race and outer race bearing faults of five different types of bearings.

All these datasets after feature extraction are summarized in Table II. First, the CWRU dataset was selected to test the performance of TabPFN while classifying bearing faults only. Then, the MaFaulDa dataset was used for the tests with misalignment and unbalance faults. Both the datasets exhibit a clear imbalance between healthy and fault states. Finally, the HUST bearing dataset was used for the tests with bearing faults for different bearing types while classifying states where inner race and outer race bearing faults occur simultaneously.

IV. RESULTS AND DISCUSSION

The performance of the proposed model using TabPFN and the performance obtained with traditional classification models were compared. As explained in Section III, three public RM fault diagnosis datasets were used.

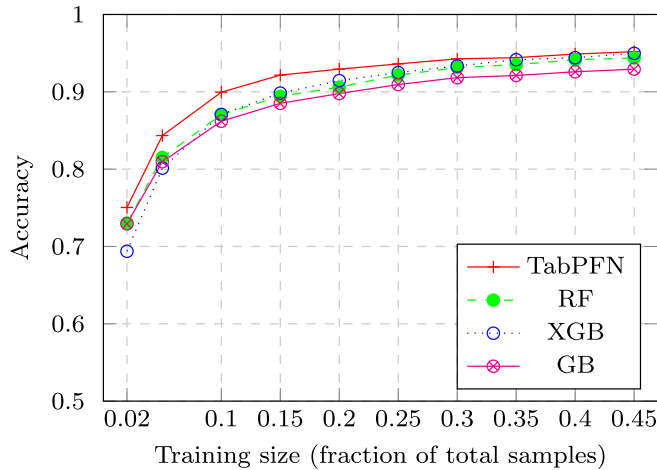
In the next subsections, the results obtained for each of the datasets are presented and discussed, comparing the TabPFN model with the rest of the classification models.

A. Results Obtained With CWRU Dataset

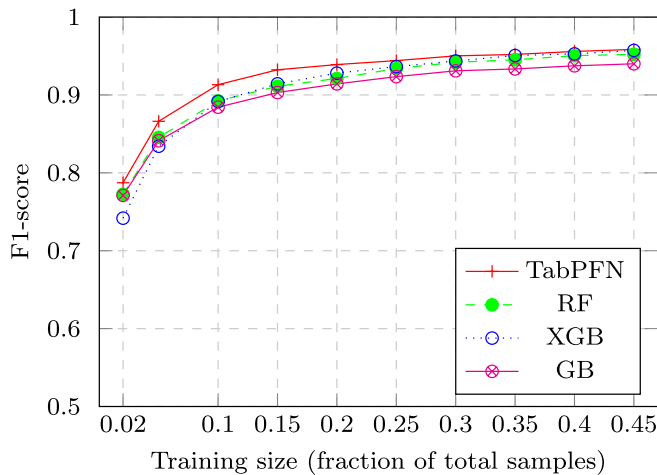
Once the CWRU dataset was preprocessed, a total of 2208 samples, each of them with seven time-domain features,

TABLE II
DESCRIPTION OF USED DATASETS

Dataset	Samples	Sampling rate	Bearings	State
CWRU	2208 x 7	48 kHz	1	Healthy / Inner / Outer / Ball
MaFaulDa	1951 x 42	50 kHz	1	Healthy / Unbalance / Horizontal misalignment / Vertical misalignment / Inner
HUST	2375 x 7	51.2 kHz	5	Healthy / Inner / Outer / Ball / Inner-Outer



(a)



(b)

Fig. 1. Results of the CWRU dataset. (a) Accuracy. (b) F1-score.

were extracted. Only 216 of the samples correspond to a healthy state of the RM. The rest of them contain bearing faults: 649 ball bearing faults, 717 inner race bearing faults, and 626 outer race bearing faults. The dataset was divided into training and testing using 2%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, and 45% for training, which means using between 45 and 994 samples for training. The number of training samples was always lower than the 1024 training sample limitation of TabPFN.

Table III includes all the performance results (accuracy, precision, recall, and F1-score) obtained by TabPFN and traditional ML models used for fault classification in the CWRU dataset while the training set size increases from

2% (45) to 45% (994) of the samples. The average training time for each of the tests is also included. As can be seen in Table III, TabPFN offers the best results in terms of accuracy, precision, recall, and F1-score in all the cases. As for the other ML models, XGB, GB, and RF offer good results, but always worse than TabPFN. TabPFN is always trained in less than 0.35 s. LP is the fastest trained in all the cases, but its performance metrics are significantly worse than those obtained with other models. The results of all the models improve as the size of the training set grows. A detailed comparison of the performance of TabPFN, XGB, GB, and RF is shown in Fig. 1(a) and (b), where accuracy and F1-score are analyzed, respectively.

B. Results Obtained With MaFaulDa Dataset

After preprocessing the MaFaulDa dataset, a total of 1951 samples were obtained, each of them with 42 time-domain features: 49 of the samples correspond to the healthy state and the rest to fault states. Of the fault state samples, 197 have horizontal misalignment, 301 vertical misalignment, 333 unbalance, 376 inner race bearing faults, 372 outer race bearing faults, and 323 ball bearing faults. The dataset was divided into training and testing using 2%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50% for training, which means using between 40 and 976 samples for training. The number of training samples was always lower than the 1024 training sample limitation of TabPFN.

The results obtained with TabPFN and the traditional ML models are shown in Table IV. TabPFN offers the best results in terms of accuracy, precision, recall, and F1-score, except when 20% of the dataset is used for the training set, when MLP has slightly higher recall. As for the other ML models, MLP, XGB, and RF have good results, but always worse than TabPFN. TabPFN is always trained in less than 0.60 s. In these tests, SVM and LP are trained faster than the rest of the ML models, but the performance metrics obtained are worse than the rest of the ML models. Again, the results of all the models improve as the size of the training set grows. To facilitate the comparison, Fig. 2(a) and (b) shows the accuracy and F1-score of each of these models, respectively, as the training set size increases.

C. Results Obtained With HUST Dataset

The HUST bearing dataset was preprocessed, and a total of 2036 samples, each of them with seven time-domain features, were extracted. From these samples, 435 correspond to healthy state while the rest correspond to bearing faults: 296 samples have ball bearing faults, 435 inner race bearing faults,

TABLE III
RESULTS USING CWRU DATASET

Tr. size	Model	Accuracy	Precision	Recall	F1-score	Tr. time(s)
2%	RF	0.7303	0.7728	0.7720	0.7724	0.0379
	SVM	0.6511	0.7113	0.7026	0.7069	0.0059
	GB	0.7293	0.7702	0.7718	0.7710	0.1585
	MLP	0.7284	0.7727	0.7686	0.7707	0.3288
	XGB	0.6938	0.7436	0.7401	0.7418	0.5684
	LP	0.5937	0.6538	0.6578	0.6558	0.0045
	TabPFN	0.7504	0.7892	0.7853	0.7873	0.3097
5%	RF	0.8151	0.8481	0.8430	0.8455	0.0370
	SVM	0.6883	0.7411	0.7316	0.7363	0.0072
	GB	0.8096	0.8437	0.8387	0.8412	0.1957
	MLP	0.7915	0.8259	0.8213	0.8236	0.4268
	XGB	0.8013	0.8366	0.8318	0.8342	0.5406
	LP	0.6978	0.7474	0.7457	0.7465	0.0053
	TabPFN	0.8434	0.8697	0.8628	0.8662	0.3195
10%	RF	0.8705	0.8953	0.8894	0.8924	0.0413
	SVM	0.7065	0.7615	0.7485	0.7549	0.0112
	GB	0.8619	0.8862	0.8824	0.8843	0.2388
	MLP	0.7734	0.8155	0.8080	0.8117	1.1628
	XGB	0.8707	0.8926	0.8903	0.8915	0.4256
	LP	0.7733	0.8120	0.8082	0.8101	0.0048
	TabPFN	0.8995	0.9144	0.9119	0.9132	0.2911
15%	RF	0.8943	0.9128	0.9095	0.9111	0.0461
	SVM	0.7090	0.7613	0.7537	0.7575	0.0167
	GB	0.8851	0.9046	0.9017	0.9031	0.2964
	MLP	0.8478	0.8719	0.8695	0.8707	4.7591
	XGB	0.8983	0.9149	0.9136	0.9143	0.5109
	LP	0.8150	0.8447	0.8427	0.8437	0.0055
	TabPFN	0.9216	0.9322	0.9323	0.9323	0.3476
20%	RF	0.9060	0.9230	0.9193	0.9212	0.0491
	SVM	0.7136	0.7636	0.7569	0.7602	0.0233
	GB	0.8977	0.9157	0.9127	0.9142	0.3426
	MLP	0.8352	0.8651	0.8592	0.8621	3.0139
	XGB	0.9145	0.9287	0.9273	0.9280	0.4864
	LP	0.8361	0.8632	0.8599	0.8616	0.0059
	TabPFN	0.9293	0.9393	0.9388	0.9390	0.3223
25%	RF	0.9217	0.9351	0.9327	0.9339	0.0537
	SVM	0.7193	0.7692	0.7613	0.7652	0.0291
	GB	0.9094	0.9243	0.9227	0.9235	0.3960
	MLP	0.8623	0.8850	0.8806	0.8828	5.0629
	XGB	0.9249	0.9367	0.9360	0.9364	0.5661
	LP	0.8430	0.8684	0.8651	0.8667	0.0064
	TabPFN	0.9361	0.9441	0.9442	0.9441	0.3370
30%	RF	0.9313	0.9429	0.9410	0.9419	0.0578
	SVM	0.7166	0.7700	0.7587	0.7643	0.0371
	GB	0.9183	0.9318	0.9302	0.9310	0.4506
	MLP	0.8578	0.8823	0.8779	0.8801	4.0489
	XGB	0.9333	0.9439	0.9432	0.9436	0.5834
	LP	0.8490	0.8729	0.8709	0.8719	0.0069
	TabPFN	0.9426	0.9502	0.9503	0.9502	0.3330
35%	RF	0.9351	0.9460	0.9444	0.9452	0.0627
	SVM	0.7153	0.7693	0.7584	0.7638	0.0452
	GB	0.9211	0.9343	0.9325	0.9334	0.5107
	MLP	0.8697	0.8920	0.8880	0.8900	6.2848
	XGB	0.9415	0.9509	0.9503	0.9506	0.6951
	LP	0.8533	0.8768	0.8748	0.8758	0.0078
	TabPFN	0.9442	0.9519	0.9520	0.9520	0.3425
40%	RF	0.9415	0.9511	0.9498	0.9504	0.0670
	SVM	0.7219	0.7733	0.7628	0.7680	0.0571
	GB	0.9259	0.9380	0.9367	0.9374	0.5536
	MLP	0.8689	0.8919	0.8871	0.8895	4.4364
	XGB	0.9441	0.9530	0.9523	0.9527	0.5988
	LP	0.8548	0.8769	0.8760	0.8764	0.0083
	TabPFN	0.9488	0.9560	0.9558	0.9559	0.3352
45%	RF	0.9437	0.9530	0.9515	0.9522	0.0707
	SVM	0.7241	0.7739	0.7658	0.7698	0.0686
	GB	0.9293	0.9407	0.9392	0.9400	0.6086
	MLP	0.8755	0.8974	0.8926	0.8950	6.0812
	XGB	0.9497	0.9575	0.9569	0.9572	0.6625
	LP	0.8591	0.8807	0.8795	0.8801	0.0092
	TabPFN	0.9519	0.9586	0.9585	0.9585	0.3410

TABLE IV
RESULTS USING MAFAULDA DATASET

Tr. size	Model	Accuracy	Precision	Recall	F1-score	Tr. time(s)
2%	RF	0.6540	0.6210	0.6490	0.6347	0.0353
	SVM	0.6366	0.6103	0.6351	0.6225	0.0068
	GB	0.6065	0.5793	0.6022	0.5905	0.4436
	MLP	0.6474	0.6323	0.6563	0.6441	0.2156
	XGB	0.6145	0.5831	0.6130	0.5977	0.4690
	LP	0.4924	0.5139	0.4920	0.5027	0.0311
	TabPFN	0.7104	0.6877	0.6887	0.6882	0.4100
5%	RF	0.7850	0.7570	0.7467	0.7518	0.0480
	SVM	0.7508	0.6973	0.7236	0.7102	0.0084
	GB	0.7720	0.7344	0.7335	0.7340	0.7272
	MLP	0.7748	0.7402	0.7696	0.7546	0.8098
	XGB	0.7740	0.7309	0.7475	0.7391	0.7680
	LP	0.6948	0.6454	0.6457	0.6456	0.0110
	TabPFN	0.8305	0.8125	0.7767	0.7942	0.4514
10%	RF	0.8527	0.8530	0.8063	0.8290	0.0540
	SVM	0.8225	0.7759	0.7984	0.7870	0.0177
	GB	0.8547	0.8376	0.8045	0.8207	1.3035
	MLP	0.8571	0.8271	0.8466	0.8367	2.5948
	XGB	0.8602	0.8327	0.8369	0.8348	1.1146
	LP	0.8005	0.7502	0.7512	0.7507	0.0248
	TabPFN	0.8853	0.8873	0.8385	0.8622	0.4161
15%	RF	0.8878	0.8871	0.8436	0.8648	0.0598
	SVM	0.8559	0.8119	0.8268	0.8193	0.0148
	GB	0.8859	0.8720	0.8380	0.8547	1.6181
	MLP	0.8884	0.8659	0.8743	0.8701	2.5787
	XGB	0.8930	0.8783	0.8655	0.8718	0.7460
	LP	0.8385	0.7970	0.7836	0.7903	0.0168
	TabPFN	0.9134	0.9109	0.8692	0.8895	0.5853
20%	RF	0.9066	0.9157	0.8721	0.8934	0.0716
	SVM	0.8796	0.8448	0.8566	0.8506	0.0206
	GB	0.9036	0.8980	0.8546	0.8758	2.0322
	MLP	0.9122	0.8958	0.9019	0.8988	3.6027
	XGB	0.9104	0.8983	0.8850	0.8916	0.5847
	LP	0.8675	0.8393	0.8258	0.8325	0.0162
	TabPFN	0.9309	0.9328	0.9018	0.9170	0.4682
25%	RF	0.9196	0.9263	0.8857	0.9055	0.0789
	SVM	0.8970	0.8652	0.8774	0.8713	0.0229
	GB	0.9166	0.9058	0.8673	0.8861	2.5403
	MLP	0.9245	0.9081	0.9147	0.9114	4.6142
	XGB	0.9299	0.9230	0.9086	0.9157	0.7300
	LP	0.8838	0.8672	0.8491	0.8581	0.0174
	TabPFN	0.9411	0.9395	0.9181	0.9286	0.4627
30%	RF	0.9291	0.9335	0.9024	0.9177	0.0864
	SVM	0.9122	0.8845	0.8927	0.8886	0.0273
	GB	0.9316	0.9277	0.8901	0.9085	2.9130
	MLP	0.9365	0.9228	0.9262	0.9245	4.1291
	XGB	0.9397	0.9345	0.9207	0.9276	0.6116
	LP	0.8977	0.8826	0.8650	0.8737	0.0188
	TabPFN	0.9504	0.9494	0.9330	0.9411	0.3522
35%	RF	0.9396	0.9442	0.9206	0.9322	0.1071
	SVM	0.9177	0.8935	0.9015	0.8975	0.0379
	GB	0.9385	0.9347	0.9024	0.9182	3.5247
	MLP	0.9418	0.9296	0.9342	0.9319	5.8059
	XGB	0.9472	0.9415	0.9345	0.9380	0.9148
	LP	0.9101	0.9006	0.8859	0.8932	0.0247
	TabPFN	0.9587	0.9574	0.9478	0.9526	0.3995
40%	RF	0.9421	0.9500	0.9264	0.9381	0.1149
	SVM	0.9254	0.9055	0.9120	0.9088	0.0442
	GB	0.9440	0.9411	0.9092	0.9248	3.8892
	MLP	0.9485	0.9364	0.9440	0.9402	5.0845
	XGB	0.9559	0.9518	0.9415	0.9466	0.7363
	LP	0.9151	0.9090	0.8878	0.8983	0.0256
	TabPFN	0.9599	0.9607	0.9490	0.9548	0.3880
45%	RF	0.9490	0.9548	0.9343	0.9444	0.1224
	SVM	0.9277	0.9068	0.9160	0.9114	0.0584
	GB	0.9486	0.9451	0.9201	0.9324	4.6471
	MLP	0.9529	0.9445	0.9487	0.9466	6.8747
	XGB	0.9577	0.9548	0.9456	0.9502	1.2203
	LP	0.9230	0.9130	0.9045	0.9087	0.0329
	TabPFN	0.9649	0.9610	0.9562	0.9586	0.4149
50%	RF	0.9522	0.9598	0.9392	0.9494	0.1354
	SVM	0.9325	0.9175	0.9213	0.9194	0.0718
	GB	0.9525	0.9519	0.9200	0.9356	5.3505
	MLP	0.9558	0.9499	0.9506	0.9502	8.7609
	XGB	0.9628	0.9614	0.9501	0.9557	1.2050
	LP	0.9304	0.9260	0.9118	0.9188	0.0711
	TabPFN	0.9684	0.9675	0.9620	0.9648	0.5720

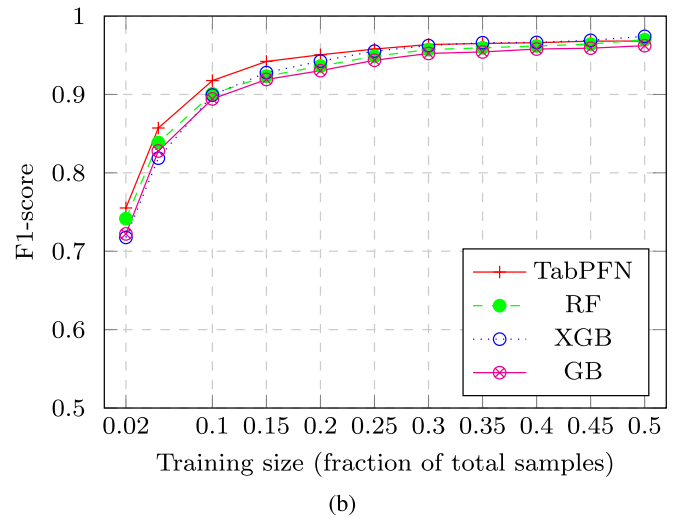
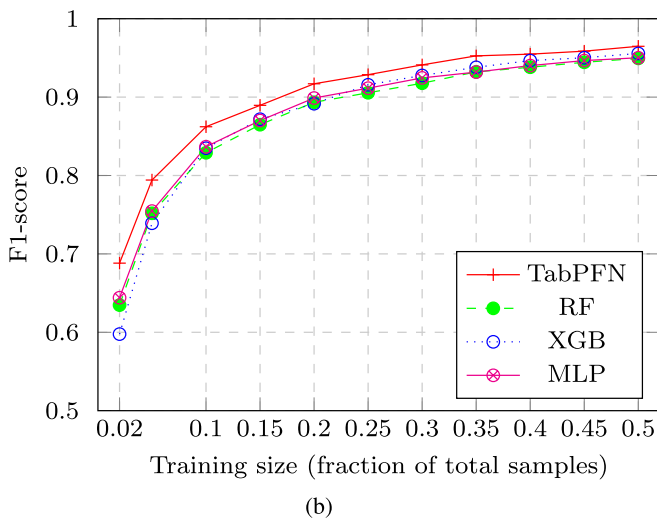
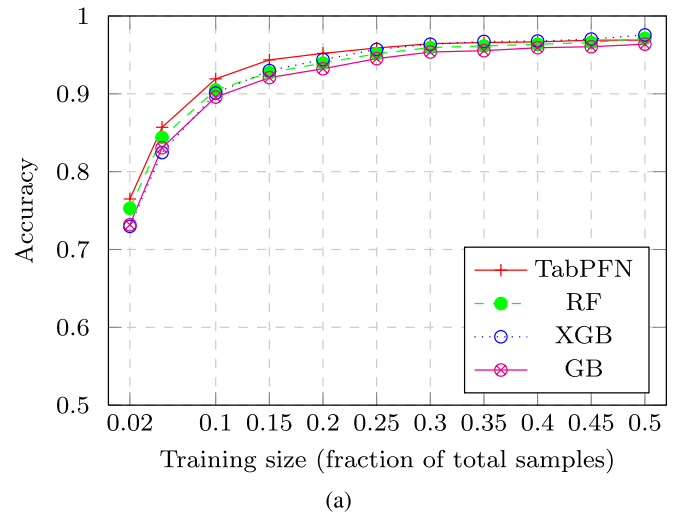
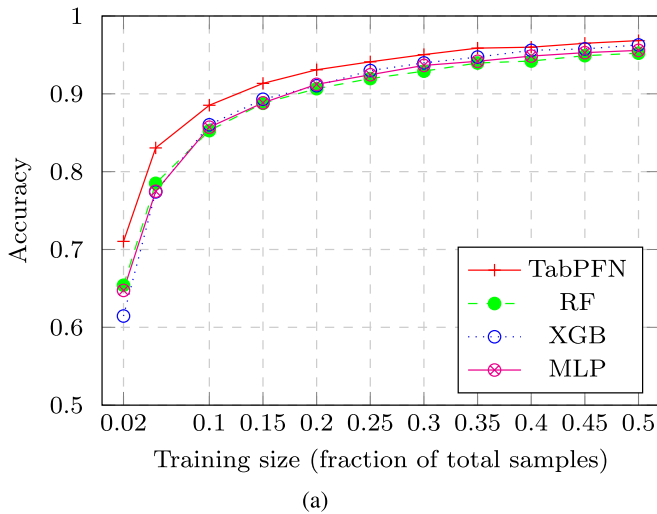


Fig. 2. Results of the MaFaulDa dataset. (a) Accuracy. (b) F1-score.

Fig. 3. Results of HUST dataset. (a) Accuracy. (b) F1-score.

435 outer race bearing faults, and the remaining 435 samples correspond to RM where there are both inner race and outer race bearing faults. This dataset was divided into training and testing using 2%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50% for training, which means using between 41 and 1018 samples for training. The number of training samples was always lower than the 1024 training sample limitation of TabPFN.

The results for each test are shown in Table V. TabPFN provides the best results until the training set size exceeds 30% of the total. After this, the results obtained by TabPFN are still good, but those of XGB are better. RF results are also better when the size of the training set is 50%. The training times of XGB and GB increase as the number of training samples increases, always lower than 1.20 s. The training time of TabPFN increases less than the other two ML models, always lower than 0.65 s. LP is again the fastest ML model to be trained, but the results are significantly worse than the results of the rest of ML models. Fig. 3(a) and (b) shows the accuracy and F1-score of the different ML models and TabPFN as the training set size increases.

D. Discussion

As can be seen in the results shown in Sections IV-A–IV-C, TabPFN provides significant improvements in the classification of RM faults. With the CWRU bearing dataset, the results obtained with TabPFN are significantly better than those obtained with the traditional ML models, especially when the number of samples used for training is very low. As the size of the training set increases, the performance obtained by the traditional ML models becomes more similar to that obtained by TabPFN. The results with the CWRU dataset show TabPFN to be a valuable tool when classifying bearing faults and healthy operation, in spite of the imbalance of the dataset.

In the second test, the MaFaulDa dataset was used, combining bearing faults, vertical and horizontal misalignment, and unbalance faults. In this case, TabPFN has significantly better results for all the sizes of the training set. Thus, not only is TabPFN valid for bearing faults' classification but also it provides better results when the dataset used is clearly imbalanced, containing only 49 samples of healthy behavior and approximately 300 samples for each type of fault. This confirms the initial assumption that TabPFN tolerates training with few samples without overfitting, even when the datasets

TABLE V
RESULTS USING HUST DATASET

Tr. size	Model	Accuracy	Precision	Recall	F1-score	Tr. time(s)
2%	RF	0.7528	0.7512	0.7323	0.7416	0.0360
	SVM	0.6813	0.6841	0.6712	0.6776	0.0057
	GB	0.7318	0.7320	0.7125	0.7221	0.2064
	MLP	0.7509	0.7635	0.7429	0.7531	0.3393
	XGB	0.7297	0.7261	0.7093	0.7176	0.7533
	LP	0.5183	0.5225	0.5071	0.5147	0.0044
	TabPFN	0.7649	0.7616	0.7489	0.7552	0.3318
5%	RF	0.8439	0.8511	0.8267	0.8387	0.0362
	SVM	0.7557	0.7689	0.7477	0.7581	0.0071
	GB	0.8307	0.8416	0.8147	0.8279	0.2214
	MLP	0.8481	0.8591	0.8418	0.8504	0.4041
	XGB	0.8245	0.8273	0.8104	0.8187	0.5367
	LP	0.7133	0.7377	0.7003	0.7185	0.0045
	TabPFN	0.8570	0.8682	0.8466	0.8572	0.2985
10%	RF	0.9049	0.9096	0.8922	0.9008	0.0390
	SVM	0.7687	0.7839	0.7604	0.7719	0.0110
	GB	0.8958	0.9034	0.8854	0.8943	0.2856
	MLP	0.5374	0.5198	0.5408	0.5301	0.3795
	XGB	0.9010	0.9060	0.8926	0.8992	0.7677
	LP	0.8354	0.8457	0.8258	0.8356	0.0053
	TabPFN	0.9192	0.9219	0.9136	0.9177	0.3135
15%	RF	0.9272	0.9293	0.9174	0.9233	0.0418
	SVM	0.7843	0.7948	0.7792	0.7870	0.0145
	GB	0.9208	0.9256	0.9127	0.9191	0.3461
	MLP	0.9275	0.9291	0.9240	0.9266	2.5690
	XGB	0.9299	0.9321	0.9238	0.9279	0.7001
	LP	0.8764	0.8829	0.8667	0.8747	0.0059
	TabPFN	0.9436	0.9438	0.9404	0.9421	0.3988
20%	RF	0.9395	0.9416	0.9308	0.9362	0.0462
	SVM	0.7871	0.7987	0.7801	0.7893	0.0193
	GB	0.9322	0.9363	0.9244	0.9303	0.3995
	MLP	0.7467	0.7614	0.7481	0.7547	1.3377
	XGB	0.9440	0.9467	0.9383	0.9425	0.7538
	LP	0.8922	0.8992	0.8818	0.8904	0.0067
	TabPFN	0.9520	0.9529	0.9483	0.9506	0.3834
25%	RF	0.9515	0.9530	0.9447	0.9488	0.0496
	SVM	0.7805	0.7881	0.7740	0.7810	0.0245
	GB	0.9452	0.9486	0.9387	0.9436	0.4642
	MLP	0.9411	0.9416	0.9386	0.9401	3.6958
	XGB	0.9571	0.9580	0.9524	0.9552	0.7381
	LP	0.9096	0.9139	0.9010	0.9074	0.0071
	TabPFN	0.9589	0.9593	0.9564	0.9579	0.4280
30%	RF	0.9594	0.9604	0.9540	0.9572	0.0527
	SVM	0.7882	0.7962	0.7811	0.7886	0.0296
	GB	0.9537	0.9567	0.9480	0.9523	0.5253
	MLP	0.8311	0.8398	0.8326	0.8362	2.3165
	XGB	0.9637	0.9644	0.9603	0.9623	0.7457
	LP	0.9166	0.9201	0.9087	0.9144	0.0076
	TabPFN	0.9644	0.9646	0.9628	0.9637	0.4329
35%	RF	0.9615	0.9617	0.9574	0.9595	0.0561
	SVM	0.7861	0.7937	0.7799	0.7867	0.0366
	GB	0.9554	0.9581	0.9503	0.9542	0.5827
	MLP	0.9426	0.9430	0.9401	0.9415	4.0072
	XGB	0.9673	0.9680	0.9640	0.9660	0.7820
	LP	0.9219	0.9251	0.9149	0.9200	0.0082
	TabPFN	0.9659	0.9657	0.9642	0.9650	0.4528
40%	RF	0.9636	0.9642	0.9594	0.9618	0.0592
	SVM	0.7845	0.7919	0.7775	0.7846	0.0435
	GB	0.9592	0.9617	0.9542	0.9579	0.6477
	MLP	0.8990	0.9041	0.8972	0.9006	3.0840
	XGB	0.9678	0.9683	0.9644	0.9664	0.8225
	LP	0.9250	0.9278	0.9176	0.9226	0.0090
	TabPFN	0.9667	0.9672	0.9649	0.9660	0.3874
45%	RF	0.9659	0.9666	0.9620	0.9643	0.0635
	SVM	0.7858	0.7922	0.7792	0.7857	0.0500
	GB	0.9605	0.9626	0.9555	0.9590	0.7020
	MLP	0.9467	0.9463	0.9447	0.9455	4.7160
	XGB	0.9703	0.9707	0.9674	0.9690	0.7795
	LP	0.9313	0.9333	0.9245	0.9289	0.0094
	TabPFN	0.9689	0.9688	0.9672	0.9680	0.4284
50%	RF	0.9713	0.9719	0.9680	0.9699	0.0728
	SVM	0.7892	0.7966	0.7835	0.7900	0.0761
	GB	0.9638	0.9653	0.9591	0.9622	1.0282
	MLP	0.9258	0.9278	0.9231	0.9254	5.5974
	XGB	0.9756	0.9754	0.9728	0.9741	1.1918
	LP	0.9335	0.9346	0.9271	0.9309	0.0127
	TabPFN	0.9690	0.9688	0.9673	0.9681	0.6399

are highly imbalanced. These results make the proposed solution suitable for real scenarios where the number of fault samples is low, such as new RM installations, changes in RM operating configuration, or when the installation is already in production but the number of faulty samples is limited.

Finally, TabPFN was tested with the HUST dataset, which includes faults in bearings of different types. It also includes samples with simultaneous inner race and outer race faults. In this case, the results obtained are better than those of the rest of the ML models until reaching 35% of the data as the training set. From this point on, the results obtained with TabPFN are still competitive but those of XGB are better. These results may be due to the bias that was arbitrarily introduced in TabPFN, with respect to causal relationships, based on the Occam's razor principle [3]. TabPFN gives more importance to simple causal relationships, because of the large amount of synthetic data with this type of relationship. In this dataset, there are more complex failures so the result is slightly worse, although it is still viable for classification in this dataset.

TabPFN's training times are consistently below 1 s, although they are not the fastest. This is because the training is performed with a single pass to adjust the weights of the neural network. This is possible because the preweights of this network were already defined in the pretraining with the synthetic data used to generalize a large number of causal relationships.

It is very important to predict healthy states correctly to prevent unnecessary downtime of RM and therefore a reduction in productivity and unnecessary maintenance costs. Therefore, each dataset was divided into training and testing using 10% for training and the remaining for testing to obtain the confusion matrices. It must be noted that this test was only done once, so the results shown in Fig. 4 correspond to the confusion matrices of a single test. The vertical labels correspond to the actual values and the horizontal labels to the predicted values classified by TabPFN, showing how the model classifies healthy states (H), ball bearing faults (B), inner race bearing faults (IR), outer race bearing faults (OR), horizontal misalignment (HM), vertical misalignment (VM), unbalance (U), and simultaneous IR-OR.

As can be seen in Fig. 4(a), the healthy states are all correctly predicted with the CWRU dataset. However, due to the imbalance of the CWRU dataset, eight faulty states are predicted as healthy. Something similar happens with the MaFaulDa dataset [see Fig. 4(b)], where only two healthy states are predicted as if they were faulty, but eight faulty states were predicted as corresponding to healthy operating conditions. Finally, as shown in Fig. 4(c), all the healthy and faulty states are correctly predicted using the HUST dataset.

Based on the results obtained, TabPFN classifies healthy states correctly although sometimes it classifies faulty states as healthy. This problem is due to the imbalance in the dataset of healthy and faulty states, where the number of faulty states is much higher than the number of healthy states. In real scenarios, the number of healthy states is much higher than the

H	194	0	0	0																
B	1	490	49	44																
IR	0	3	625	17																
OR	7	60	14	483																
	H	B	IR	OR																
	(a)																			

H	27	0	1	0	0	0	1													
B	0	291	0	0	0	0	0													
IR	1	13	237	74	9	0	4													
OR	0	0	22	313	0	0	0													
HM	2	1	26	1	112	35	0													
VM	1	3	5	0	4	257	1													
U	3	1	10	0	0	12	274													
	H	B	IR	OR	HM	VM	U													
	(b)																			

H	392	0	0	0	0															
B	0	239	19	6	1															
IR	0	28	340	14	10															
OR	0	16	5	369	2															
IR-OR	0	19	1	25	346															
	H	B	IR	OR	IR-OR															
	(c)																			

Fig. 4. Confusion matrices using 10% for training on each dataset. (a) CWRU dataset. (b) MaFaulDa dataset. (c) HUST dataset.

number of faulty states, making the proposed solution not only feasible for industry but also achieving better performance than the traditional ML models.

V. CONCLUSION AND FUTURE WORK

A novel early fault classification method for RM using TabPFN with limited data has been proposed. The method was tested using three different RM public datasets. Seven time-domain features from each vibration axis were extracted and then used to classify RM faults.

The results show that TabPFN works correctly with limited data for training, achieving better performance than the traditional classification algorithms without hyperparameter optimization, avoiding overfitting and completing the training process in less than 1 s in a GPU. Furthermore, its ability to correctly classify healthy and faulty states has been proved with both the balanced and imbalanced datasets. Thus, TabPFN is suitable for use in industrial plants where the monitored machinery has been recently installed or reconfigured and the amount of data for training is limited. To summarize, our achievements are as follows.

- 1) TabPFN is a reliable model for fault detection in RM.
- 2) The inherent characteristics of the datasets that generate RM faults are beneficial for TabPFN. Good results are achieved even with imbalanced datasets.
- 3) The TabPFN-based fault detector performs well in the three public datasets tested because the model can generalize even with different types of faults.
- 4) In this context, TabPFN achieves competitive results with other state-of-the-art algorithms, especially when there are very few samples to train on.

Future work will be geared toward applying TabPFN to current signature features in rotating machines that are not accessible for vibration measurement, such as submerged pumps. Moreover, its utility in other scenarios, such as gear faults or even combustion motor faults, will be explored.

REFERENCES

- [1] Case Western Reserve University. *Bearing Data Centre*. Accessed: Jun. 22, 2023. [Online]. Available: https://engineering.case.edu/bearing_datacenter
- [2] *MAFAULDA: Machinery Fault Database*. Accessed: Jun. 22, 2023. [Online]. Available: https://www02.smt.ufrj.br/~offshore/mfs/page_01.html
- [3] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, "Occam's Razor," *Inf. Process. Lett.*, vol. 24, no. 6, pp. 377–380, 1987.
- [4] P. Ewert, C. T. Kowalski, and M. Jaworski, "Comparison of the effectiveness of selected vibration signal analysis methods in the rotor unbalance detection of PMSM drive system," *Electronics*, vol. 11, no. 11, p. 1748, May 2022.
- [5] S. Gawde, S. Patil, S. Kumar, and K. Kotecha, "A scoping review on multi-fault diagnosis of industrial rotating machines using multi-sensor data fusion," *Artif. Intell. Rev.*, vol. 2022, pp. 1–54, Jan. 2022.
- [6] A. Gholaminejad, F. S. Bidgoli, J. Poshtan, and M. Poshtan, "A novel Kurtogram-based health index for induction motor fault diagnosis," in *Proc. Int. Aegean Conf. Electr. Mach. Power Electron. (ACEMP) Int. Conf. Optim. Electr. Electron. Equip. (OPTIM)*, Aug. 2019, pp. 85–92.
- [7] H. M. Hashemian, "Wireless sensors for predictive maintenance of rotating equipment in research reactors," *Ann. Nucl. Energy*, vol. 38, nos. 2–3, pp. 665–680, Feb. 2011.
- [8] N. Hollmann, S. Müller, K. Eggensperger, and F. Hutter, "TabPFN: A transformer that solves small tabular classification problems in a second," 2022, *arXiv:2207.01848*.
- [9] N. D. Thuan and H. S. Hong, "HUST bearing: A practical dataset for ball bearing fault diagnosis," 2023, *arXiv:2302.12533*.
- [10] R. Huang, J. Li, Y. Liao, J. Chen, Z. Wang, and W. Li, "Deep adversarial capsule network for compound fault diagnosis of machinery toward multidomain generalization task," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
- [11] R. Huang, J. Li, S. Wang, G. Li, and W. Li, "A robust weight-shared capsule network for intelligent machinery fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6466–6475, Oct. 2020.
- [12] A. Khan, H. Hwang, and H. S. Kim, "Synthetic data augmentation and deep learning for the fault diagnosis of rotating machines," *Mathematics*, vol. 9, no. 18, p. 2336, Sep. 2021.
- [13] Y. Li, Y. He, R. Liao, X. Zheng, and W. Dai, "Integrated predictive maintenance approach for multistate manufacturing system considering geometric and non-geometric defects of products," *Rel. Eng. Syst. Saf.*, vol. 228, Dec. 2022, Art. no. 108793.
- [14] L. Magadán, F. J. Suárez, J. C. Granda, and D. F. García, "Low-cost industrial IoT system for wireless monitoring of electric motors condition," *Mobile Netw. Appl.*, vol. 28, no. 1, pp. 1–10, Feb. 2022.
- [15] L. Magadán, F. J. Suárez, J. C. Granda, F. J. delaCalle, and D. F. García, "A robust health prognostics technique for failure diagnosis and the remaining useful lifetime predictions of bearings in electric motors," *Appl. Sci.*, vol. 13, no. 4, p. 2220, Feb. 2023.
- [16] Z. Meng, Q. Li, D. Sun, W. Cao, and F. Fan, "An intelligent fault diagnosis method of small sample bearing based on improved auxiliary classification generative adversarial network," *IEEE Sensors J.*, vol. 22, no. 20, pp. 19543–19555, Oct. 2022.
- [17] S. Müller, N. Hollmann, S. P. Arango, J. Grabocka, and F. Hutter, "Transformers can do Bayesian inference," 2021, *arXiv:2112.10510*.
- [18] T. Nagler, "Statistical foundations of prior-data fitted networks," 2023, *arXiv:2305.11097*.
- [19] A. G. Nath, S. S. Udmale, D. Raghuvanshi, and S. K. Singh, "Improved structural rotor fault diagnosis using multi-sensor fuzzy recurrence plots and classifier fusion," *IEEE Sensors J.*, vol. 21, no. 19, pp. 21705–21717, Oct. 2021.

- [20] C. Peng, H. Gao, X. Liu, and B. Liu, "A visual vibration characterization method for intelligent fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 192, Jun. 2023, Art. no. 110229.
- [21] N. Rezaadeh, A. De Luca, and D. Peretto, "Unbalanced, cracked, and misaligned rotating machines: A comparison between classification procedures throughout the steady-state operation," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 44, no. 10, p. 450, Oct. 2022.
- [22] J. Roldán-Gómez, J. Boubeta-Puig, J. Carrillo-Mondéjar, J. M. C. Gómez, and J. M. del Rincón, "An automatic complex event processing rules generation system for the recognition of real-time IoT attack patterns," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106344.
- [23] T. R. Shaham, T. Dekel, and T. Michaeli, "SinGAN: Learning a generative model from a single natural image," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4569–4579.
- [24] R. M. Souza, E. G. S. Nascimento, U. A. Miranda, W. J. D. Silva, and H. A. Lepikson, "Deep learning for diagnosis and classification of faults in industrial rotating machinery," *Comput. Ind. Eng.*, vol. 153, Mar. 2021, Art. no. 107060.
- [25] R. Wang, Z. Chen, and W. Li, "Gradient flow-based meta generative adversarial network for data augmentation in fault diagnosis," *Appl. Soft Comput.*, vol. 142, Jul. 2023, Art. no. 110313.
- [26] T. Williams, X. Ribadeneira, S. Billington, and T. Kurfess, "Rolling element bearing diagnostics in run-to-failure lifetime testing," *Mech. Syst. Signal Process.*, vol. 15, no. 5, pp. 979–993, Sep. 2001.
- [27] X. Xiao, C. Li, J. Huang, and T. Yu, "Fault diagnosis of rolling bearing based on knowledge graph with data accumulation strategy," *IEEE Sensors J.*, vol. 22, no. 19, pp. 18831–18840, Oct. 2022.
- [28] Y. Xu et al., "CFCNN: A novel convolutional fusion framework for collaborative fault identification of rotating machinery," *Inf. Fusion*, vol. 95, pp. 1–16, Jul. 2023.
- [29] X. Ying, "An overview of overfitting and its solutions," *J. Phys., Conf. Ser.*, vol. 1168, Feb. 2019, Art. no. 022022.
- [30] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, and G. P. Li, "Predictive maintenance in the industry 4.0: A systematic literature review," *Comput. Ind. Eng.*, vol. 150, Dec. 2020, Art. no. 106889.



L. Magadán received the B.S. degree in computer engineering in information technology and the M.S. degree in computer science from the University of Oviedo, Gijón, Spain, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree in computer architecture and technology.

He is currently working as an Interim Professor with the Department of Computer Science and Engineering, University of Oviedo. His main research interests include the Industrial Internet

of Things and artificial intelligence applied to fault detection and health prognostics of rotating machinery in industrial environments.



J. Roldán-Gómez received the B.S. and M.S. degrees in computer engineering and the Ph.D. degree in advanced computing technologies from the University of Castilla-La Mancha, Ciudad Real, Spain, in 2017, 2018, and 2023, respectively.

He is working as an Interim Professor with the Department of Computer Science and Engineering, University of Oviedo, Gijón, Spain. His main research interests include artificial intelligence applied to threat detection in IoT environments and automatic rule generation in CEP engines.



J. C. Granda received the M.S. and Ph.D. degrees in computer science from the University of Oviedo, Gijón, Spain, in 2004 and 2008, respectively.

He is an Associate Professor with the Department of Computer Science and Engineering, University of Oviedo. In recent years, he has been working on several projects related to wireless sensor networks applied and Industrial Internet of Things. He is also working on several projects related to multimedia systems and

multimedia networking. His research interests include quality inspection systems and predictive maintenance.



F. J. Suárez received the Ph.D. degree from the University of Oviedo, Gijón, Spain, in 1998.

He is a Full Professor with the Department of Computer Science and Engineering, University of Oviedo. His current research interests include wireless sensor networks and edge/fog/cloud architectures for IoT. In recent years, he has lead several projects in that field in collaboration with industrial companies as the Head of the IoT Systems and Services Research Team.