Bearing remaining useful life prediction using Support Vector Machine and Hybrid Degradation Tracking model

Group 14
ML Lab Project

Introduction

1. 2 new features - **RRMS, IRRMS**
2. SVM Based multi class classification.
3. Generalized Exponential Degradation Model for feature fitting.
4. Hybrid tracking model to fine tune RUL predictions.
5. Good performance, under gradual degradation process.

Previous Work in the field

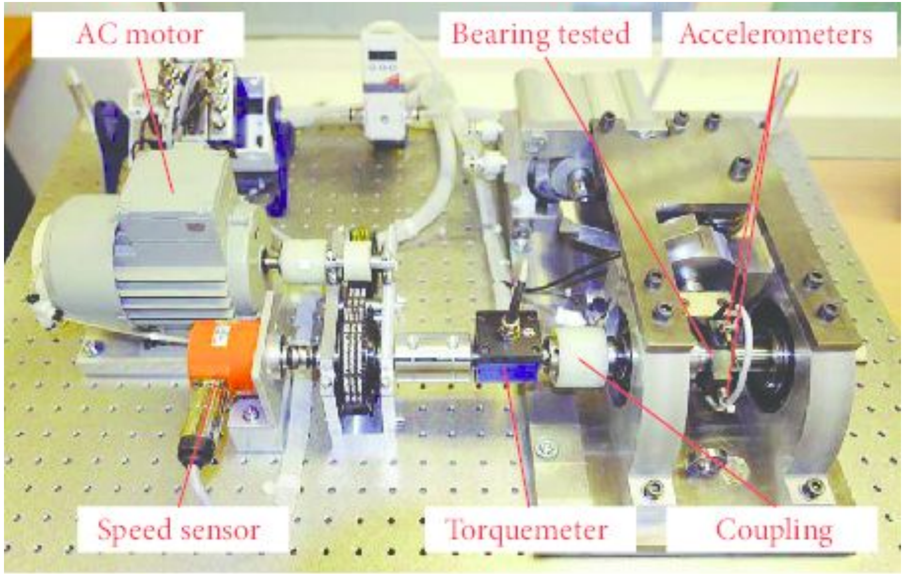
Author/ Publication	Work	Caveat
Ning et al.	RNN for bearing stage detection	RUL prediction results greatly affected by the fluctuation of the RNN results.
Wang et al.	Relevance Vector Machine (RVM) regressions to get different RVs	Cannot be generalized.

Contributions of the paper

1. Two new features are defined, which **enhances** the initial signs of degradation and **reduces** the random fluctuation of measurements. Both features are dimensionless, so this prognostic approach is **generalized** and can be applied under <u>different operating conditions</u>.
2. A new degradation model is proposed, which combines simplicity and flexibility.
3. Prognostic work is converted into **classification** work performed by SVM and **local prediction** work performed by hybrid degradation tracking model.
4. A hybrid degradation tracking model is proposed to improve the accuracy of the predicted RULs.

Introduction... to the dataset

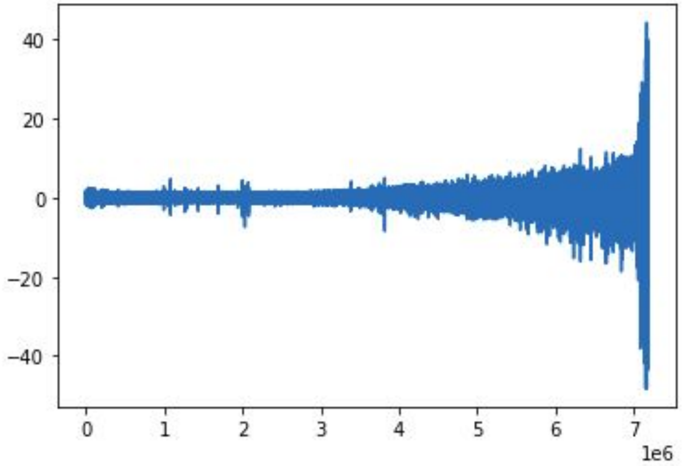
1. Provided by FEMTO-ST Institute (Besançon - France, http://www.femto-st.fr/).
2. Experiments carried out on a laboratory experimental platform (PRONOSTIA) that enables accelerated degradation of bearings under constant and/or variable operating conditions, while gathering online health monitoring data (rotating speed, load force, temperature, vibration).

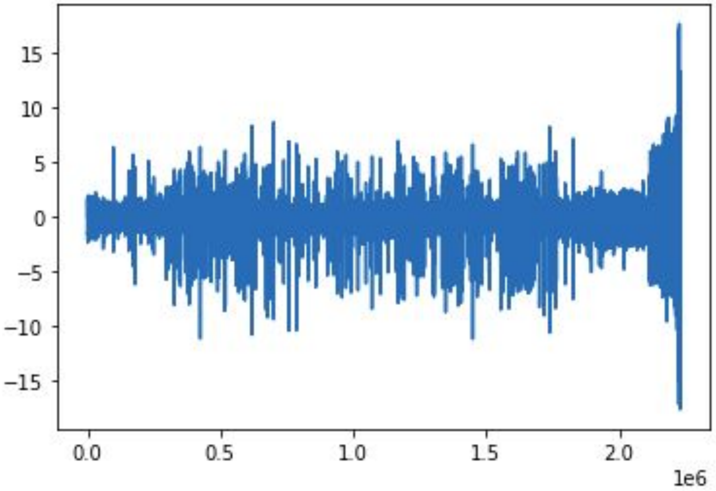


Introduction… to the dataset

1. Regarding the PHM Challenge, data representing 3 different loads were considered (rotating speed and load force).
2. The challenge datasets were characterized by a small amount of training data and a high variability in experiment durations (from 1h to 7h).
3. Theoretical framework (L10, BPFI, BPFE, etc.) **mismatches** the experimental observations.

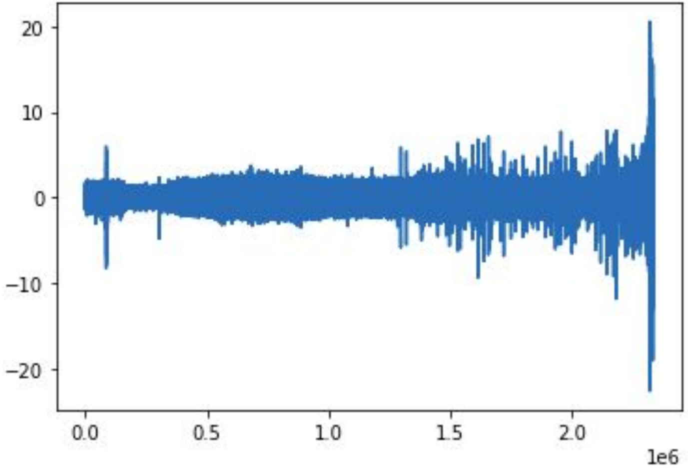
Visualizing the Dataset

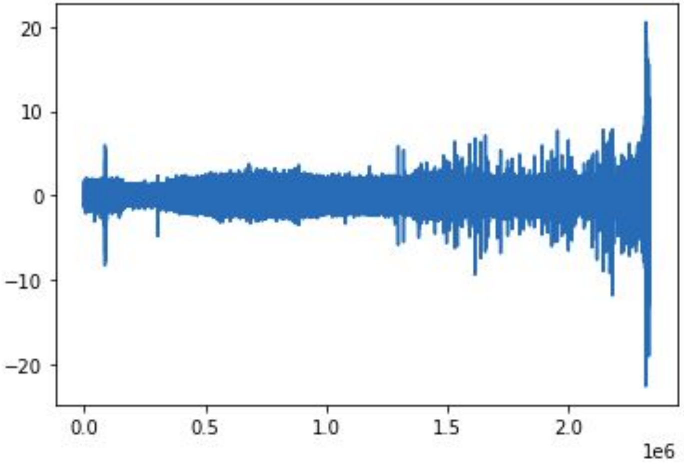




Run to failure acceleration for bearing1_1 and bearing 1_2

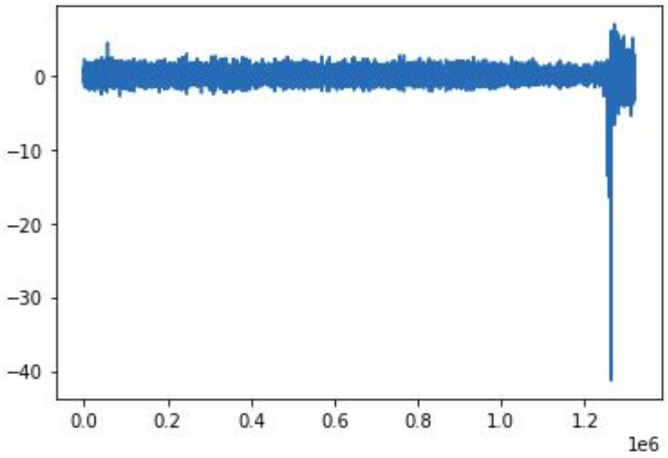
Visualizing the dataset...

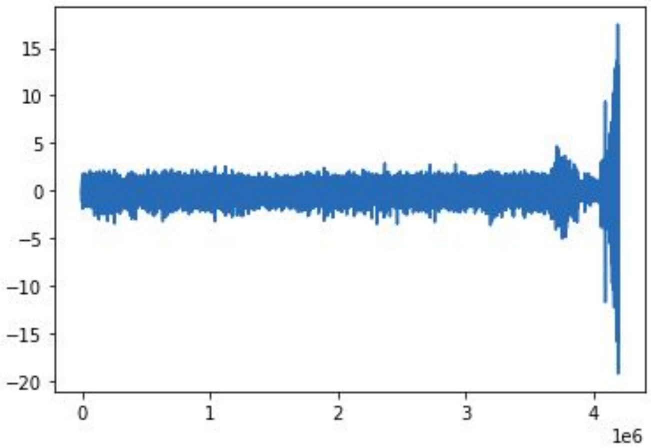




Run to failure acceleration for bearing2_1 and bearing 2_2

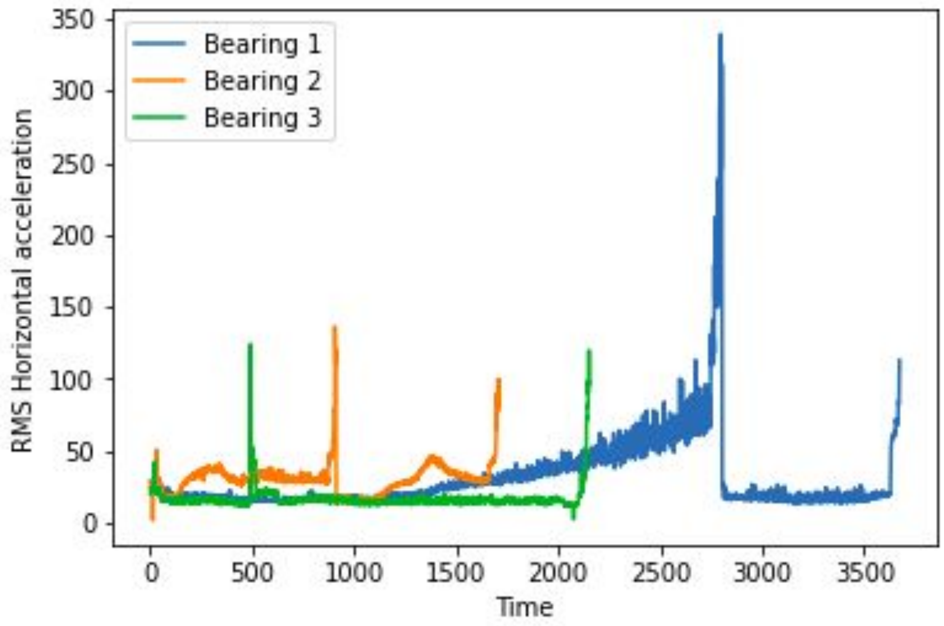
Visualizing the dataset...



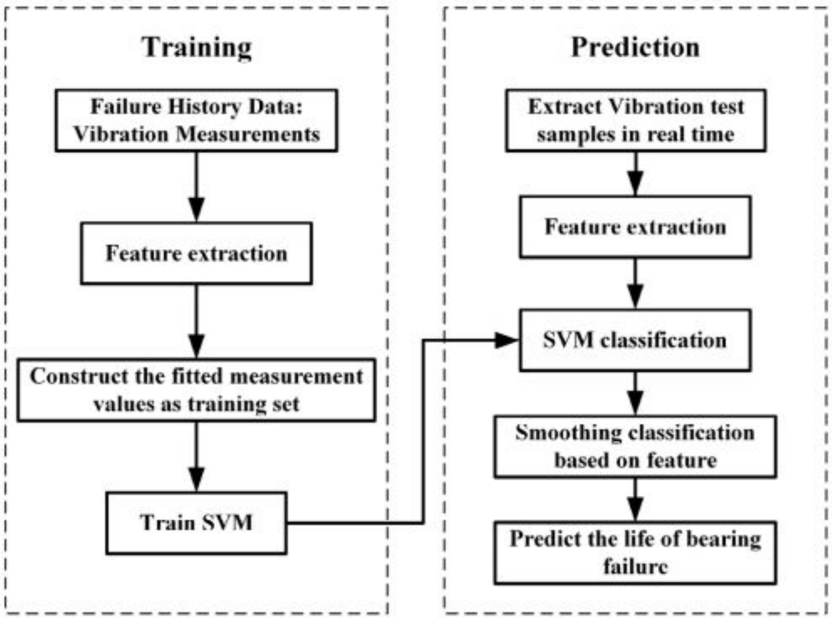


Run to failure acceleration for bearing3_1 and bearing3_2

Visualizing parameters (RMS_horizontal)



Flow chart of the proposed method



Features extraction

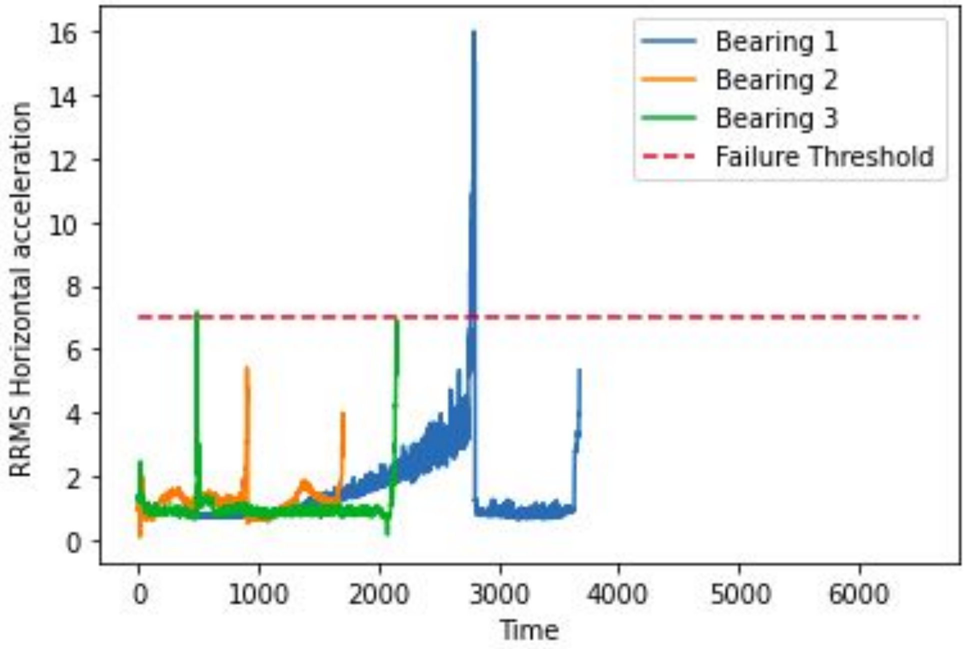
1. RMS?
    a. Positively correlated to Acceleration
    b. Spurious fluctuation due to damage propagation on race, sensor noise and vibration.
    c. New agnostic measurements required:
        i. **RRMS**
        ii. **IRRMS**

Features extraction - RRMS (Relative RMS)

$$RRMS(i) = \frac{RMS(i)}{RMS_{\mathrm{norm}}}, \tag{1}$$

$$RMS_{\mathrm{norm}} = {}^{1}/_{K} \sum_{i=1}^{K} RMS(i), \tag{2}$$

RRMS Visualization



Features extraction - IRRMS (Inertial Relative RMS)

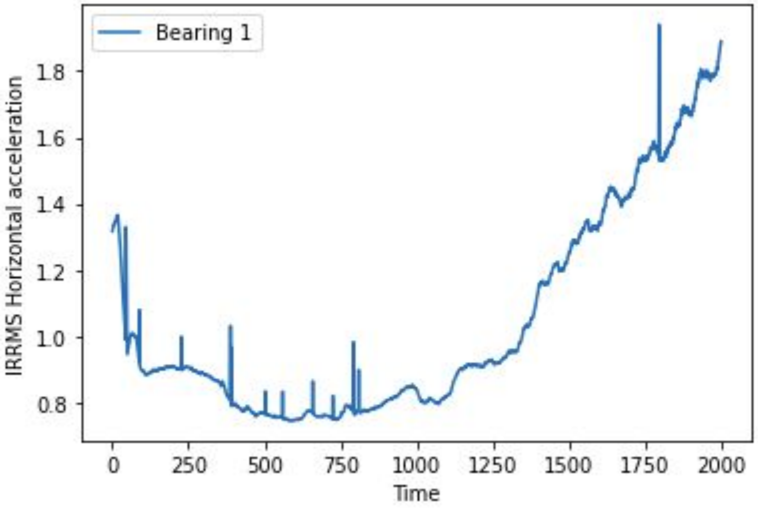
$$IRRMS(i) = \begin{cases} {}^1/_n \sum_{j=1}^{n} RRMS(j), & RRMS_d(\mathrm{i}) \leq \min\{RRMS_d(j)\} \\ k \cdot x_i + b, & \min\{RRMS_d(j)\} < RRMS_d(i) \\ & < \max\{RRMS_d(j)\} \\ k \cdot x_i + b \\ + \max\{RRMS_d(j)\}, & RRMS_d(i) \geq \max\{RRMS_d(j)\} \end{cases},$$

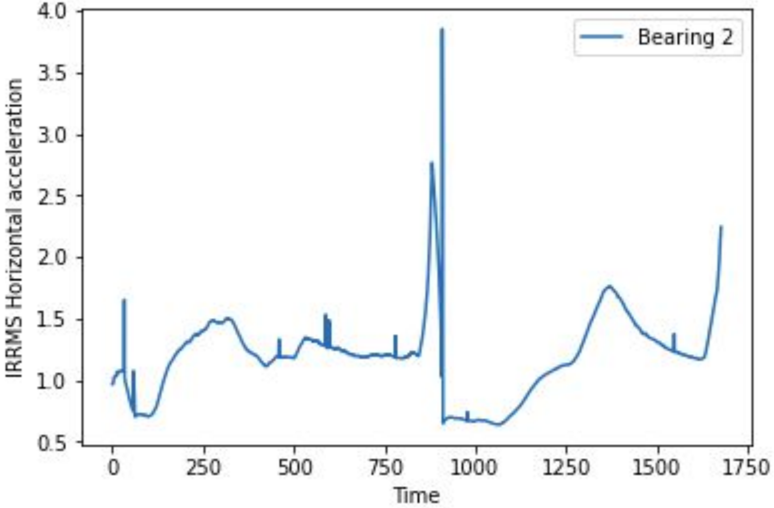
IRRMS...

$$k = \left[\sum_{j=1}^{n} x_j y_j - {}^1\!/_n \sum_{j=1}^{n} x_j \sum_{j=1}^{n} y_j \right] \Big/ \left[\sum_{j=1}^{n} x_j^2 - {}^1\!/_n \left(\sum_{j=1}^{n} x_j \right)^2 \right]$$

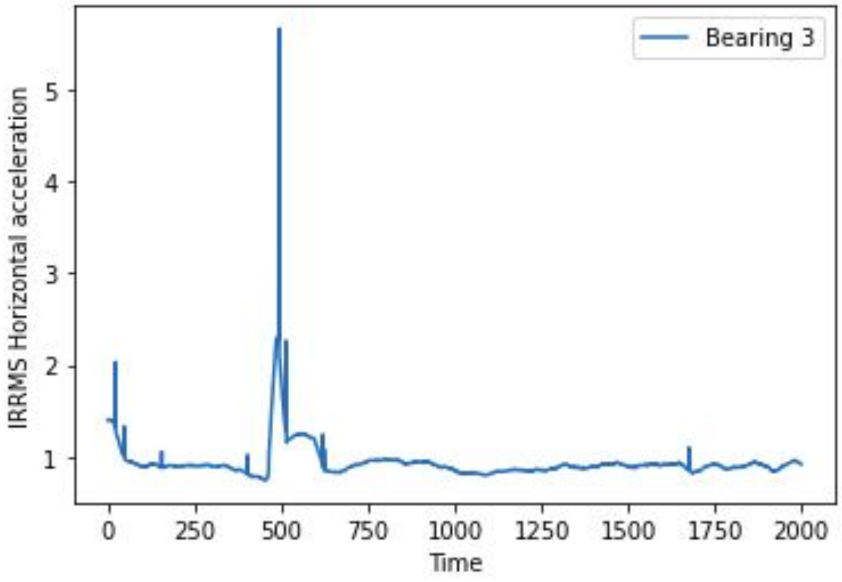
$$b = {}^1\!/_n \left[\sum_{j=1}^{n} y_j - k \sum_{j=1}^{n} x_j \right]$$

IRRMS Visualization





IRRMS Visualization...



Feature Fitting

A generalized degradation model used is, which combines simplicity and flexibility. The proposed model is:

$$\lambda(t) = Y + Mt\beta,$$

where Y is introduced to indicate the initial value, M is introduced to control the scale of the fitted values, and β (β > 0) is introduced to control the change rate of fitted values.

The actual fitting model which depicts more information of the degradation process of bearing is defined as follows:

here, k and b are the slope and intercept of the stable operation stage, respectively.

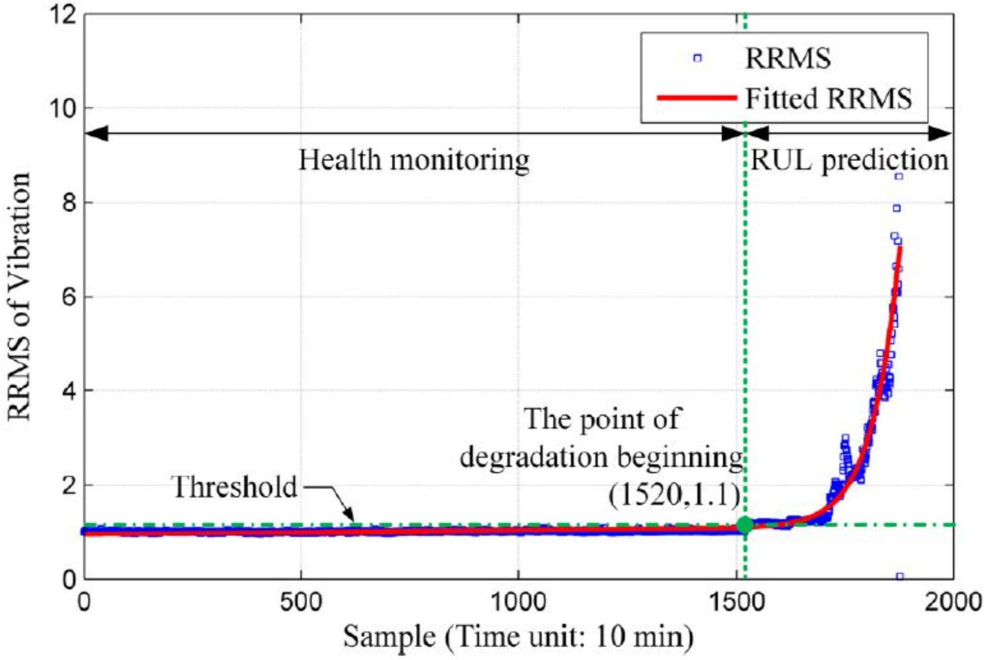
$$\begin{cases} y = kt + b, & y < 1.1 \\ y = Y + Mt^{\beta}, & y \geq 1.1 \end{cases},$$

Optimization Algorithm for Feature fitting

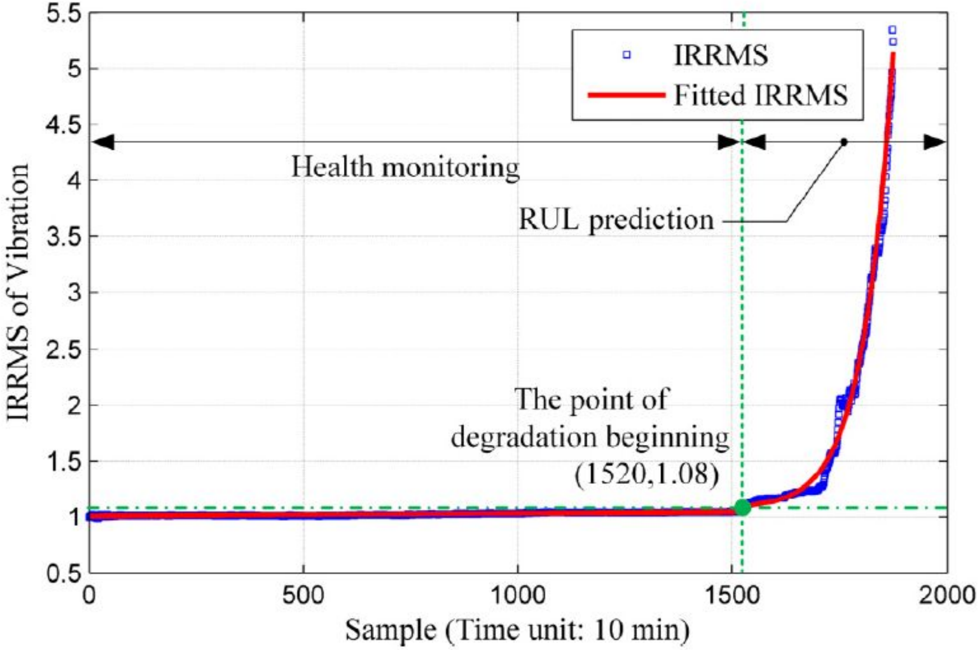
Particle Swarm Optimization Algorithm with Linearly Decreasing Inertia Weight is used.

Table 1
Optimal parameters for fitting model.

	k	b	Y	M	β
RRMS	2.31×10^{-5}	0.99	1.10	1.68×10^{-93}	28.58
IRRMS	2.41×10^{-5}	1.01	1.08	3.22×10^{-86}	26.30



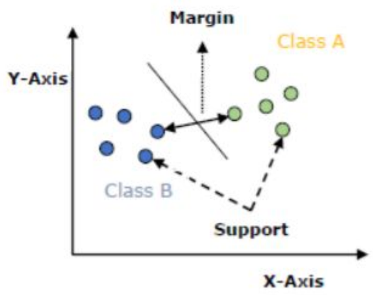
RRMS
Fitted RRMS
Health monitoring
RUL prediction
RRMS of Vibration
The point of
degradation beginning
(1520,1.1)
Threshold
Sample (Time unit: 10 min)
0
2
4
6
8
10
12
0
500
1000
1500
2000


IRRMS
Fitted IRRMS
Health monitoring
RUL prediction
The point of
degradation beginning
(1520,1.08)
IRRMS of Vibration
Sample (Time unit: 10 min)
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
500
1000
1500
2000

Classifier - SVM

Support vector machines (SVMs) are powerful yet flexible supervised machine learning algorithms which are used both for classification and regression.

An SVM model is a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH).



Kernel Functions

Kernel Function transforms the training set of data so that a non-linear decision surface is able to transformed to a linear equation in a higher number of dimension spaces.

Gaussian Radial Basis Function (RBF): It is one of the most preferred and used kernel functions in svm. It is usually chosen for non-linear data. It helps to make proper separation when there is no prior knowledge of data.

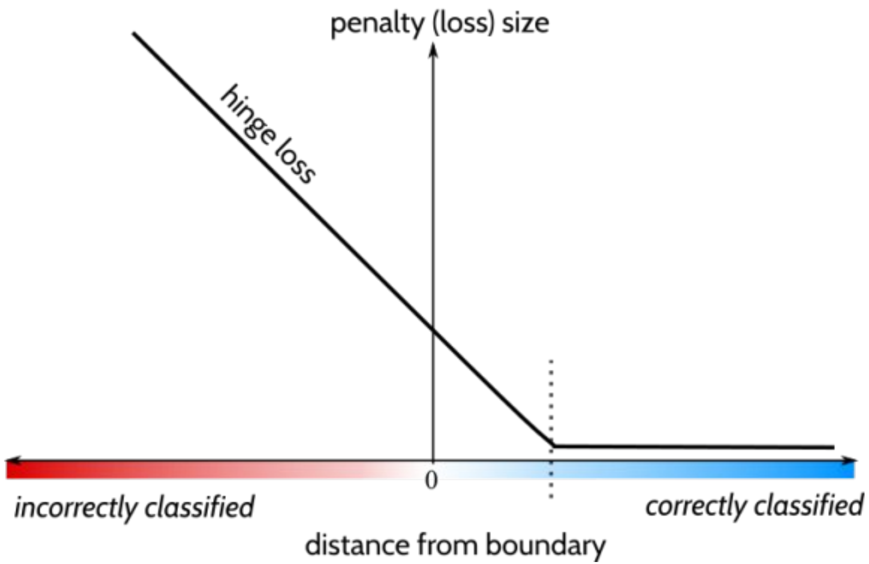
$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

Hinge Loss

$$L = \max(0, 1 - y * f(x))$$

The x-axis represents the distance from the boundary of any single instance, and the y-axis represents the loss size, or penalty, that the function will incur depending on its distance.

```python
def _cost(self):
    '''
    Hingle loss with l2 regularisation
    '''
    m = self.X.shape[0]
    hinge_loss = self.C * \
        (1 - np.multiply(self.Y, np.dot(self.X, self.W)))/m
    return -np.mean(1/2 * np.dot(self.W.T, self.W) + hinge_loss)
```



SGD and SMO Algorithms

The **advantages** of Stochastic Gradient Descent (SGD) are:

- Efficiency.
- Ease of implementation (lots of opportunities for code tuning).

The **disadvantages** of Stochastic Gradient Descent include:

- SGD requires a number of hyperparameters such as the regularization parameter and the number of iterations.
- SGD is sensitive to feature scaling.

The **advantages** of Sequential minimal optimization (SMO) are:

- Fast training of support vector machine
- Kernel trick is applied efficiently in SMO.

The **disadvantages** of Sequential minimal optimization include:

- Doesn't work well with large dataset.
- Hard to implement.

OVR vs OVO

OVR:

Also known as one-vs-all, this strategy consists in fitting one classifier per class. For each classifier, the class is fitted against all the other classes.

Advantages:

- Computational efficiency (only N classifiers are needed).
- Interpretability

OVO:

This strategy consists in fitting one classifier per class pair.

Advantages:

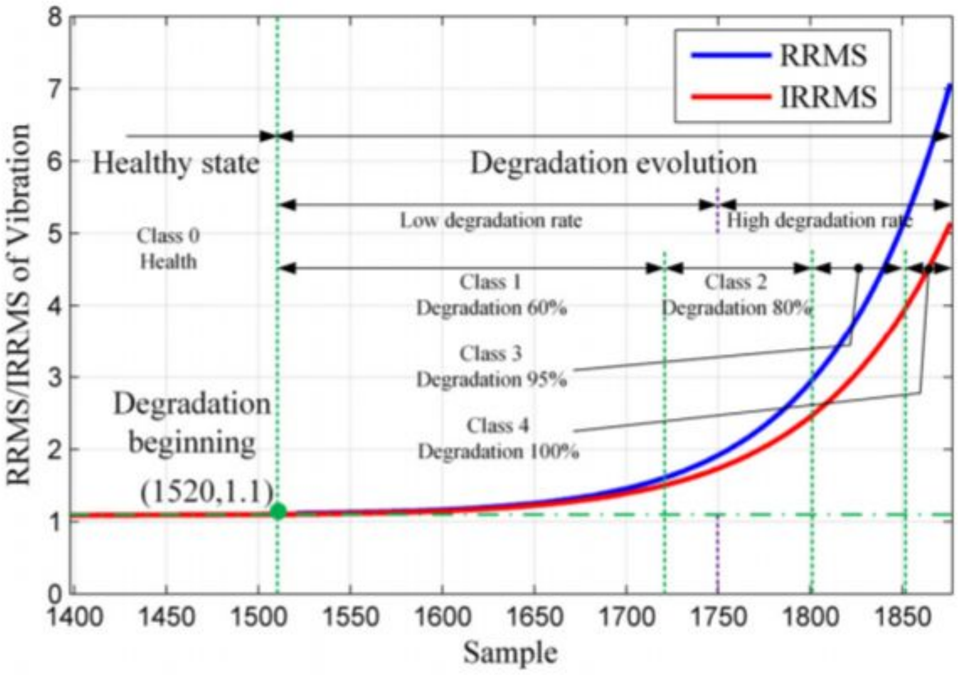
- Good for algorithms such as kernel algorithms which don't scale well with N samples.
- Each individual learning problem only involves a small subset of the data.

RUL Prediction

Use classes for different degradation states:

- Class 0: Healthy state
- Class 1: Degradation till 60%
- Class 2: Degradation 60-80%
- Class 3: Degradation 80-95%
- Class 4: Degradation 95-100%

Degradation starts when RRMS = 1.1



Hybrid Degradation tracking model

SVM classifier gives discrete values for degradation (ie, 60%, 80%, 95% or 100%) according to the classes we defined earlier.

In order to get more accurate RUL%, we need to get continuous values instead of discrete values.

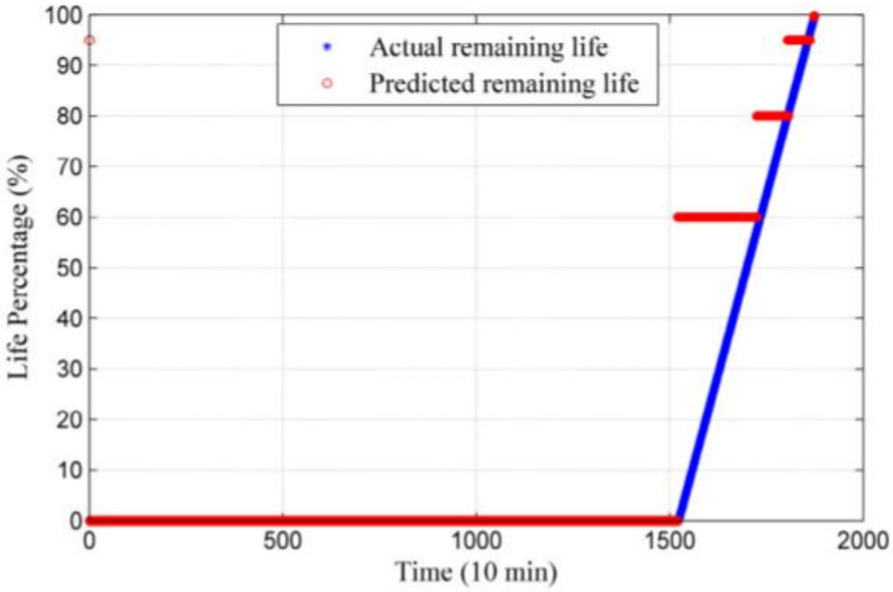


Fig. 18. Classification results of the real bearing degradation.

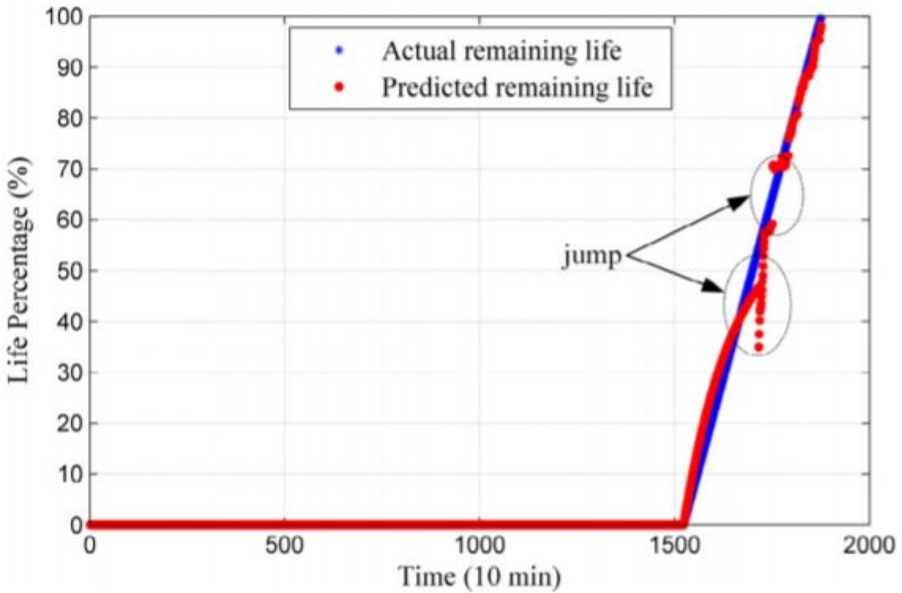
Hybrid Degradation Tracking Model (Contd..)

Algorithm of hybrid degradation tracking model.

Initialization
Set the number of averaging iterations si
Set the algorithm transformation threshold k_0
Set the initial value ($IRRMS_{initial}$) and terminated value ($IRRMS_{termination}$) for each class according to the fitted IRRMS measurement.
If $k \le k_0$(k is the slope of IRRMS curve, decided by Eq. (5))
Initialize the moving step $sm = 0$
Repeat
Calculate the predicted life percentage by $LP(sts) = [LP(sts) + LP(sts-1)]/2$, and sts is the class of sample.
Increasing the moving steps by $sm = sm + 1$
Until ($sm = si$)
End
If $k > k_0$
Calculate the smoothed life percentage by

$$LP(j) = \frac{(\ln IRRMS(j) - \ln IRRMS(j-1)) \cdot (LP(sts) - LP(sts-1))}{\ln IRRMS_{termination} - \ln IRRMS_{initial}} + LP(sts-1)$$

End
Choose the next new point to inspect it
Until the end of the time sequence



Future Works

- As future works, the raw vibration signals should be filtered by some suitable methods before extracting features.
- Besides, the initial value (IRRMSinitial) and terminated value (IRRMStermination) for each class based on degradation index need to be determined accurately by statistical analysis.
- To guarantee the accuracy and generalization of classification, the classification method employing on bearing degradation stage needs further research, such as XGboost, decision tree, extreme learning machine.