Federico Gargiulo

Tutor: Pasquale Arpaia

XXIX Cycle - III year presentation

Failure Predictions and Anomaly Detections Based on Artificial Intelligence Techniques Exploiting Heterogeneous Measures

Università degli Studi di Napoli Federico II

Summary

- Personal Background
- Research Activity & Context
- Diagnostics in Screw Compressors
- Failure Prediction in Hard Disks
- Fault Detection and Failure Prediction in Injection Motors
- Conclusions

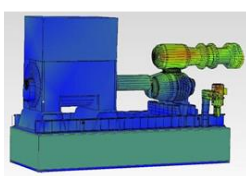
Personal Background

- Master Degree in Computer Engineering with a curriculum in Embedded Systems
- Computer Engineer in Control Section for Cryogenics at CERN from Feb 2018 until Gen 2019
- Data Scientist in Analysis Section for Data Center at CERN from May 2019 to April 2021
- Research Fellow at CESMA Unina from May 2021 to Present
- PhD Thesis's Title: *Failure Predictions and Anomaly Detections by artificial intelligence with heterogeneous measures*

Research Activity

A

Diagnostics in Screw Compressors

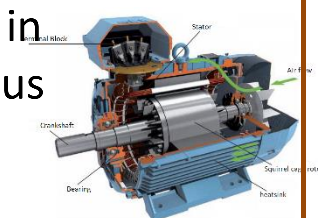


B

Failure Prediction in Hard Disks

C

Fault Detection and Failure Prediction in Asyncronous Motors



Research Activity

Failure Predictions and Anomaly Detections by artificial intelligence with heterogeneous measures

Data Acquisition ⟩ Preprocessing ⟩ Modelling ⟩ Predictive Maintenance

Data Acquisition
- Acquisition Sensor Setup
- Label Data
- Store Data

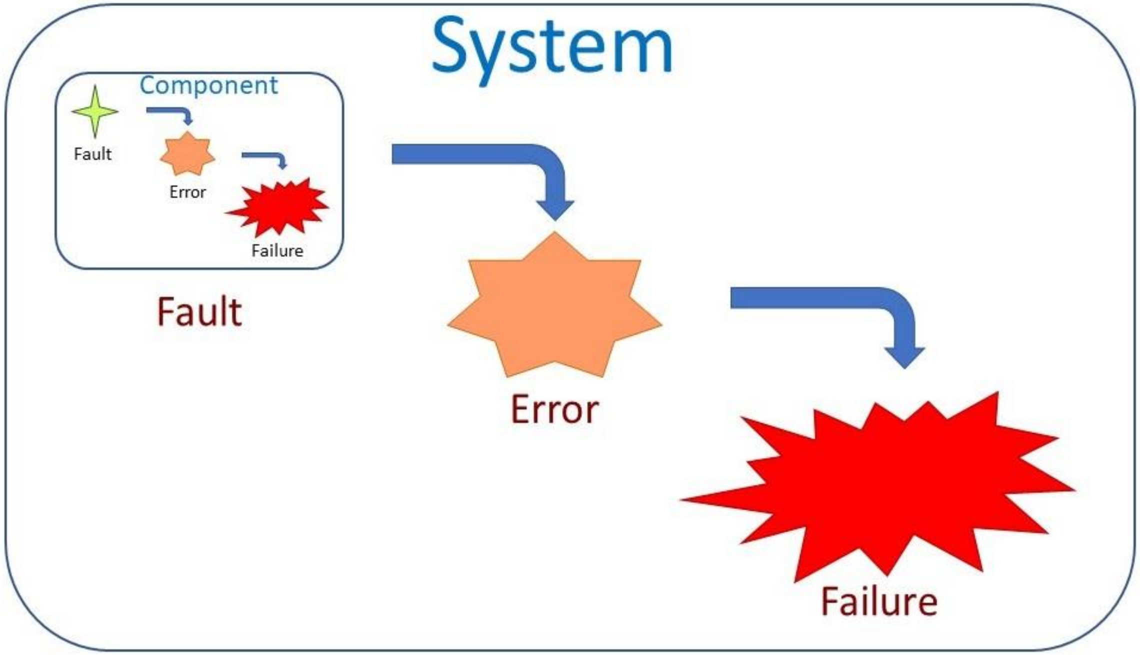
Preprocessing
- Data Cleaning
- Data Visualization
- Feauture Extraction

Modelling
- Supervised/Unsupervised Machine Learning
- Hyperparameters Optimization
- Performance Assesment

Predictive Maintenance
- Model Deployment
- System Integration
- Efficiency Optimization

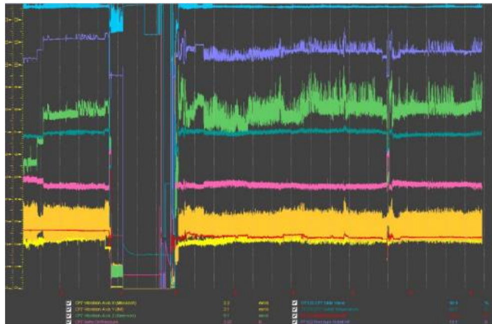
Fault-Error-Failure Definitions

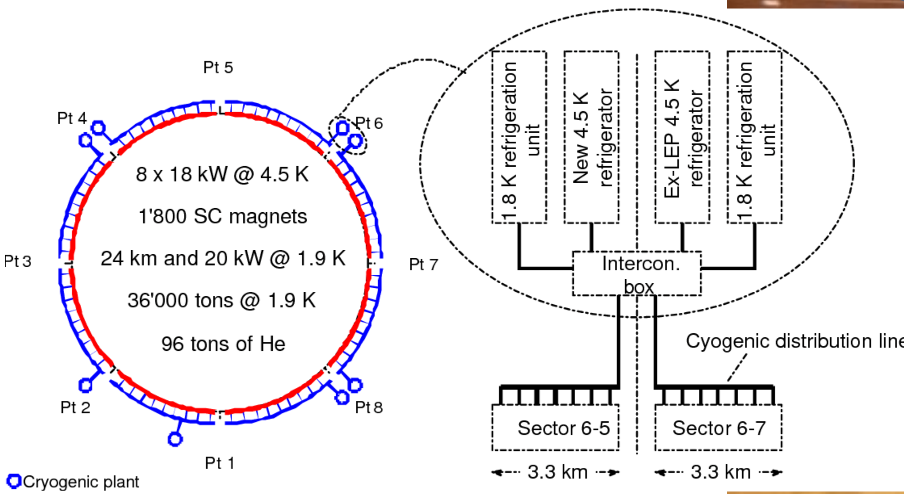
- A failure is a deviation from the specifications of a system;
- The presence of the fault (or even the error) does not produce necessarily a failure;
- Failure'cause is a fault whose manifestation is called error;
- Faults can be internal or external;
- A fault can be a system's component failure[1].



[1] Eric Alata, JoÃ£o Antunes, Mohamed Kaaniche, Nuno Neves, VincentNicomette, and Paulo VerÃssimo. Critical utility infrastructural resilienceproject acronym: Crutial. 01 2022.

Diagnostics in Screw Compressor



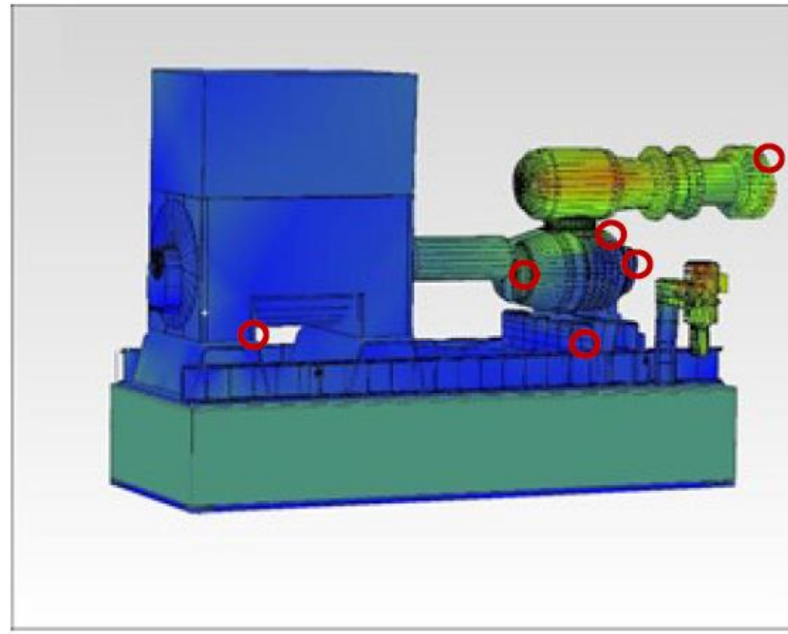


Diagnostics in Screw Compressor

Fluid machines are composed of hundreds of components and each can present completely different problems, with completely different effects on the final system.

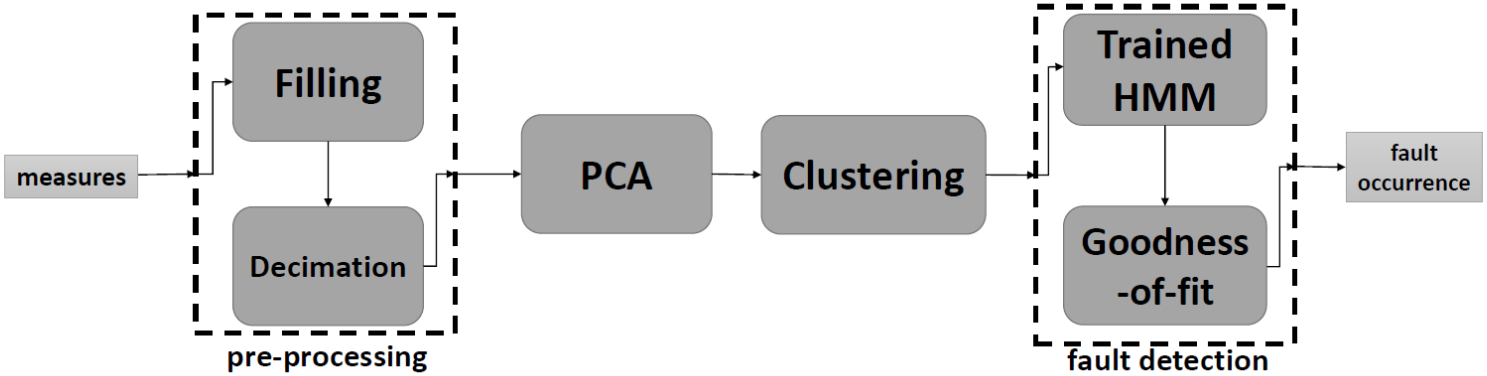
Until now in the literature there was no fault detection methodology valid for the case study (threshold, missing values etc).

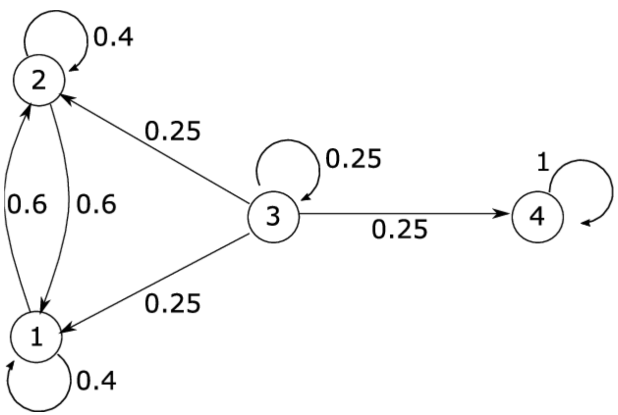
We addressed the problem using machine learning techniques based on ***Hidden Markov Models***. Results below.

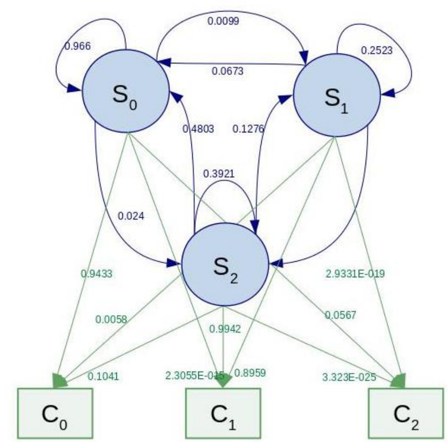


			
IMI 643A00 Piezoelectric vibration sensor	**PT100** Temperature transmitter Platinum Resistance	**Rosemount 1151** Pressure Transmitter	**SINEAX 1538** Transducer for AC current

Proposed Method for Diagnostics in Screw Compressors

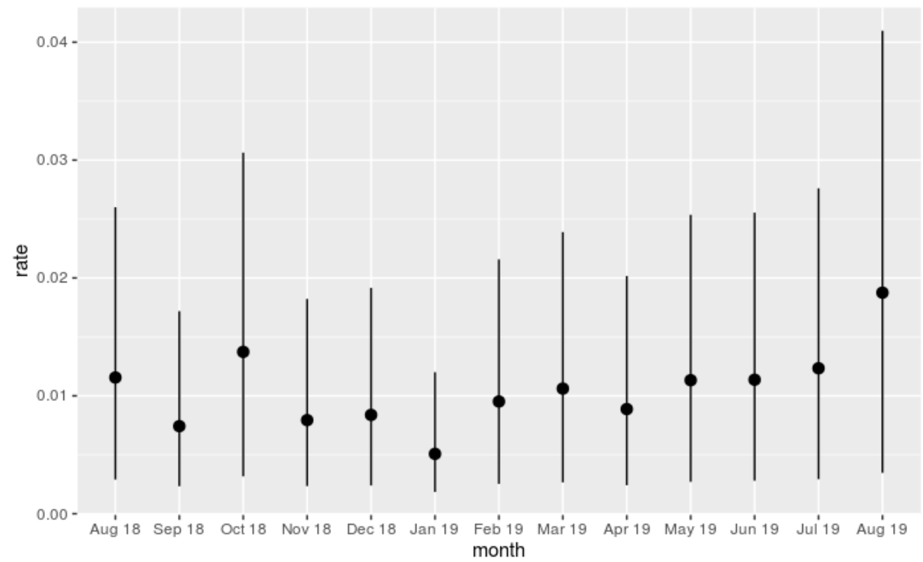






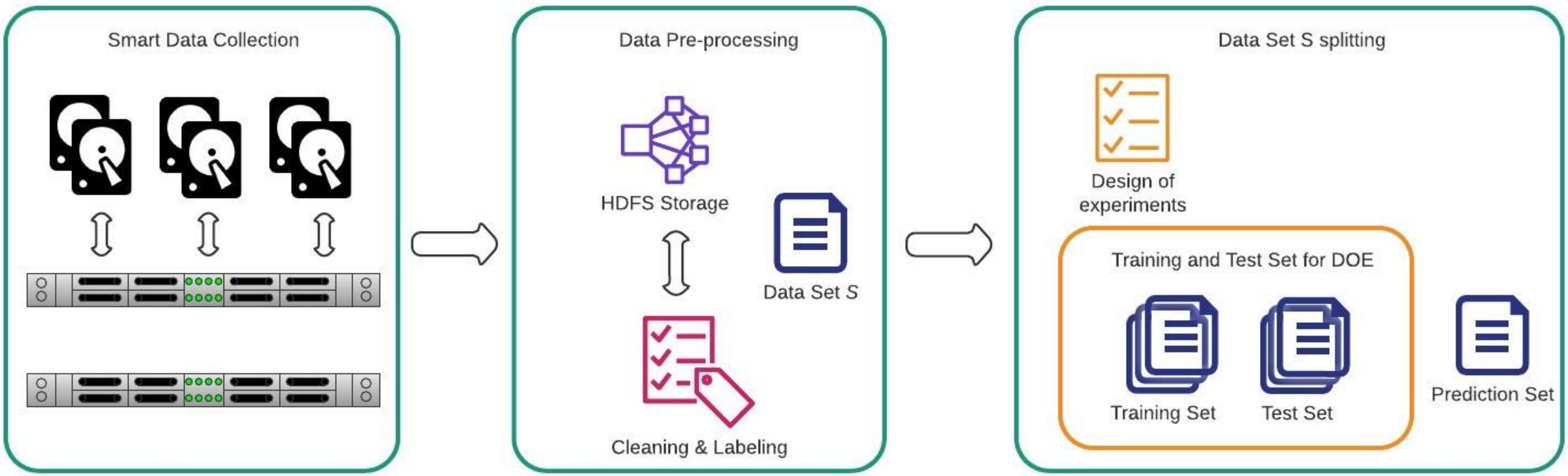
Failure Prediction in Hard Disks

The world of data centers is highly heterogeneous. Due to the rapid technological progress. So far, fault prediction methods do not provide good results because of several reasons. (early replacing, new models, fault in sockets etc).

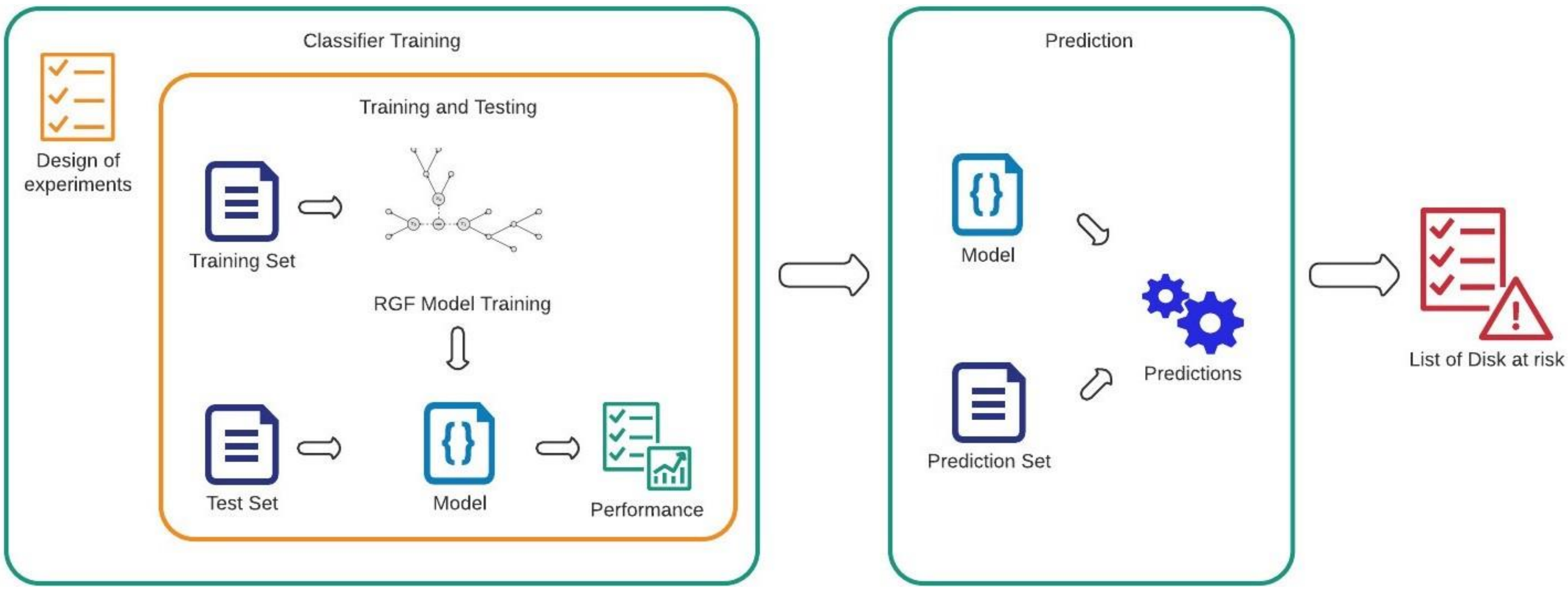


So far, there are few novelties in the new hard disk models and the technological progress in electromagnetic hard disks is slowed down. Thus, the number of general failures observed is increased as no more hard drives are replaced due to progress but due to failure.

Proposed Method for Failure Prediction in Hard Disks (1/2)



Proposed Method for Failure Prediction in Hard Disks (2/2)



Raw Data and Processing

- Raw Data need to be pre-processed and cleaned because of:

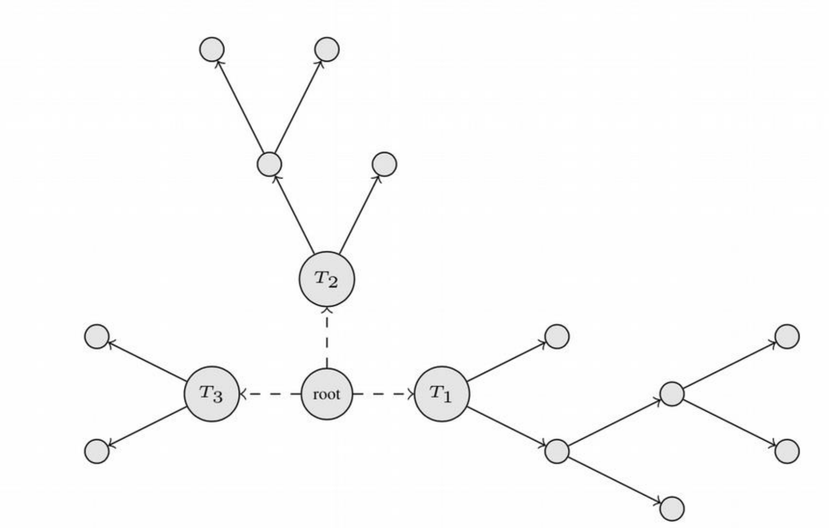
1. Measure Gaps
2. Missing Attributes
3. SMART sensors are proprietary technologies dependent
4. Errors in Measures
5. Disks moving (between hosts, Wigner etc.)

- A new labeling euristics proposed: *A disk is classified as broken if it has been removed and all other hard disk belonging to the same host machine continue to operate nominally.*

#	Attribute name [3]		Attribute name [3]
01	Read Error Rate	12	Power Cycle Count
03	Spin-Up Time	192	Unsafe Shutdown Count
04	Start/Stop Count	193	Load Cycle Count
05	Reallocated Sectors Count	194	Temperature
07	Seek Error Rate	197	Current Pending Sector Count
09	Power-On Hours	198	Uncorrectable Sector Count
10	Spin Retry Count	199	UltraDMA CRC Error Count

[3] https://en.wikipedia.org/wiki/S.M.A.R.T.

Regularized Greedy Forest



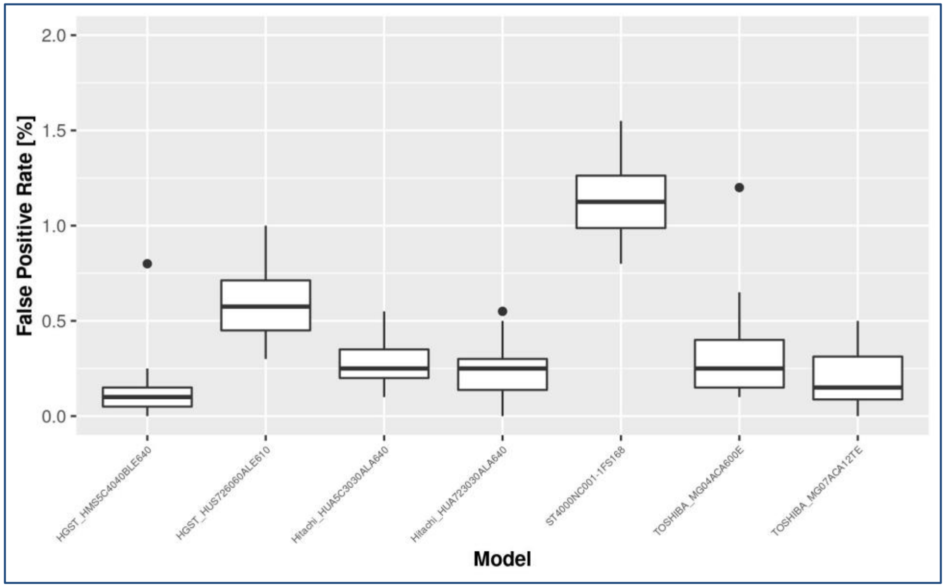
Regularized Greedy Forest (RGF) is a decision forest learning algorithm performed to have a tree-based structure for nonlinear decision. It was introduced in 2014 by R.Jhonson and T.Zhang and has been implemented in libraries for python, R, C++ etc.

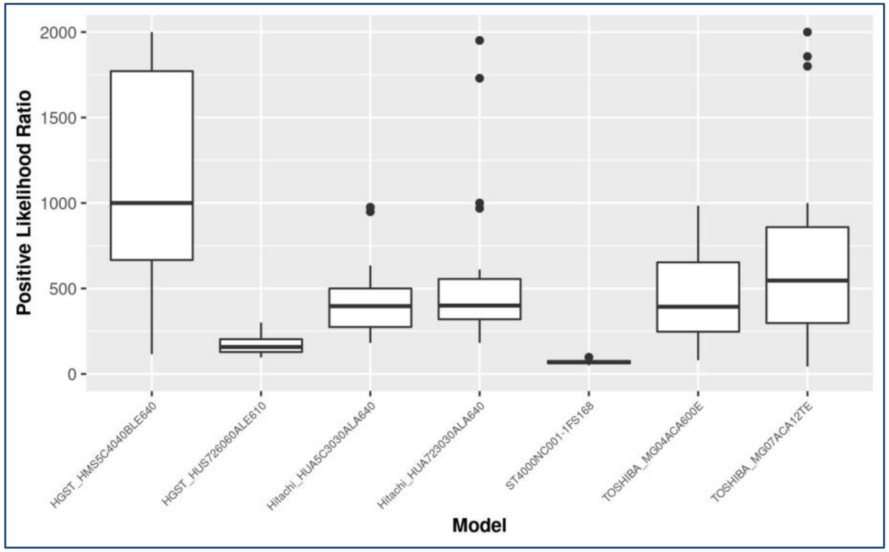
RGF performs 2 steps:

- Finds the one step structural change to the current forest to obtain the new forest that minimises the loss function (e.g. Least squares or logloss)
- Adjusts the leaf weights for the entire forest to minimize the loss function

False Positive Rate & Positive Likelihood Ratio

The False Positive Rate (FPR) is the probability of false alarm.





The Positive Likelihood Ratio (usually abbreviated as LR+) is the ratio between the sensitivity, which takes into account the rate of true positives and the false positive rate.

Results

Model	Total drives	Healthy-Broken	Recall*	FPR*	PLR*
HGST_HMS5C4040BLE640	12043	11962-81	98.4%	0.2%	659
HGST_HUS726060ALE610	11398	11238-160	92.2	0.8	115
Hitachi_HUA5C3030ALA640	9190	9172-18	95.8	0.6	160
ST4000NC001-1FS168	7968	7948-20	99.5	0.4	284
TOSHIBA_MG04ACA600E	5073	4695-378	78.1	0.9	86

$$Accuracy = \frac{\sum True\ Positive + \sum False\ Negative}{Total\ Population}$$

$$Recall = \frac{\sum True\ Positive}{\sum True\ Positive + \sum False\ Negative}$$

$$False\ Positive\ Rate = \frac{\sum False\ Positive}{\sum False\ Positive + \sum True\ Negative}$$

$$Positive\ Likelihood\ Ratio = \frac{\sum True\ Positive}{\sum True\ Positive + \sum False\ Negative} \frac{\sum False\ Positive + \sum True\ Negative}{\sum False\ Positive}$$

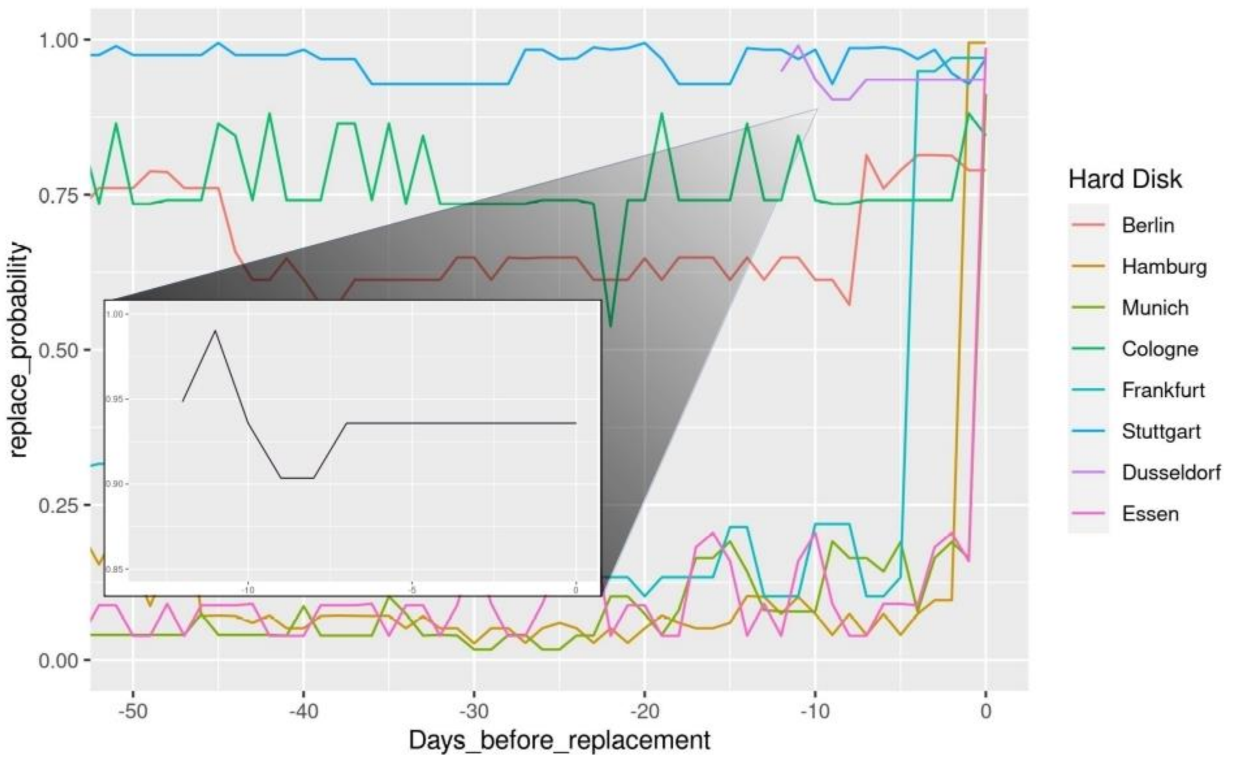
*The values in the table represent the medians of each index calculated on a sample of 30 experiments

Validation

A small fraction of the total drives' volume has been saved to validate the method.

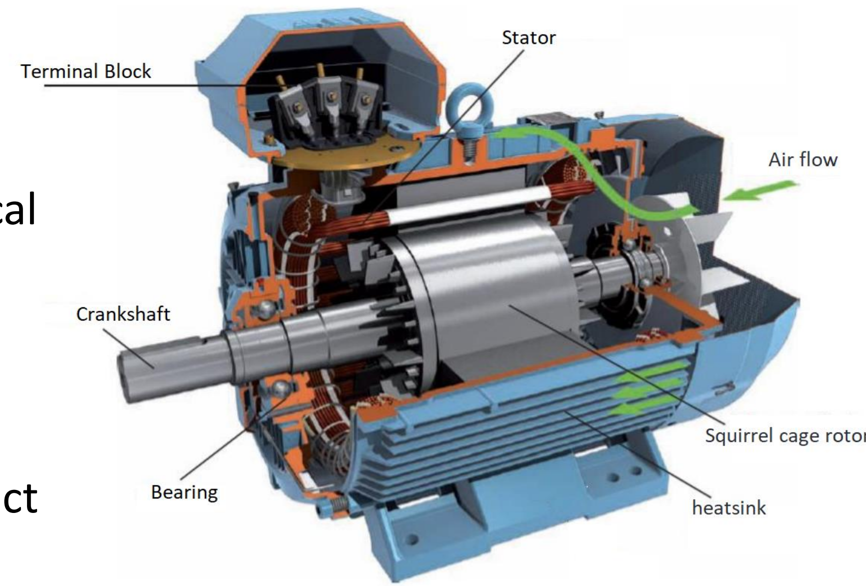
The drives in the figure have all been removed for maintenance and are tested according to the proposed method.

The method proved to be able to recognize them as close to bankruptcy many days in advance.

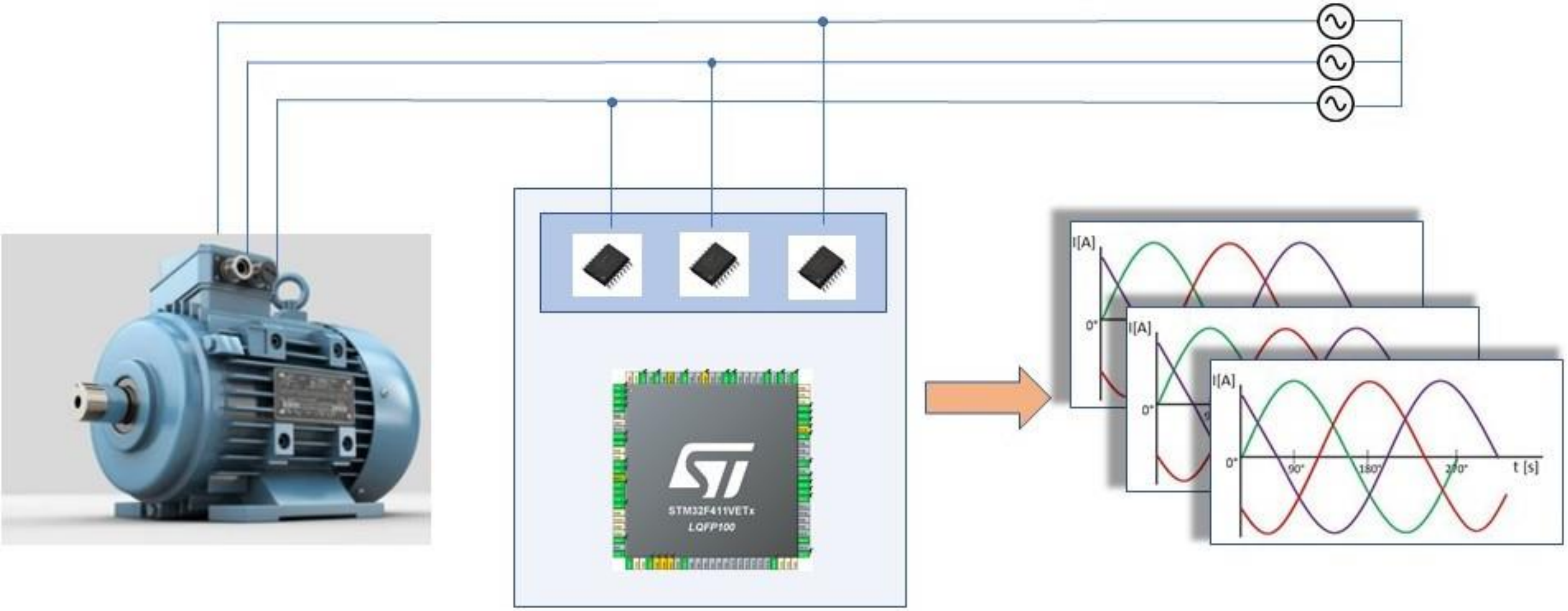


Fault Detection and Failure Prediction in Injection Motors

- Over 80% of electrical motors used in manufacturing companies are three-phase squirrel cage asynchronous;
- An engine is made up of several components and can be subject to electrical or mechanical failures;
- An induction motor's failure can have critical consequences for the system in which it is installed and an economic impact due to system downtime.

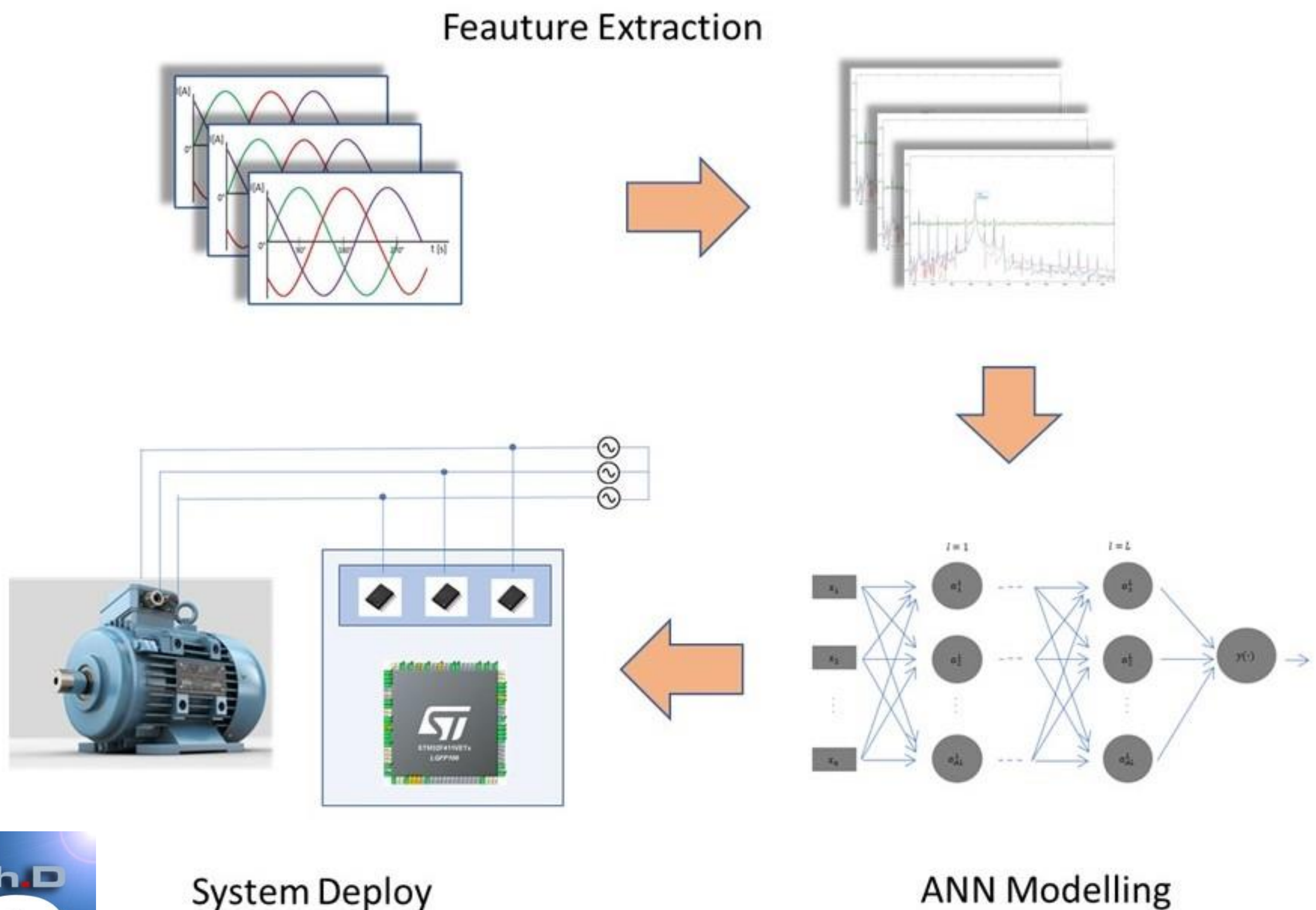


Proposed Method in Fault Detection and Failure Prediction in Asyncronous Motors



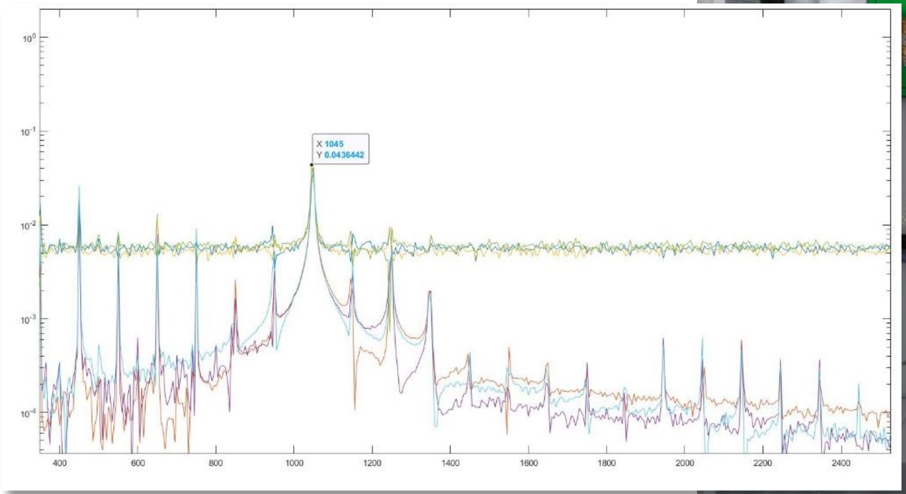
Federico Gargiulo

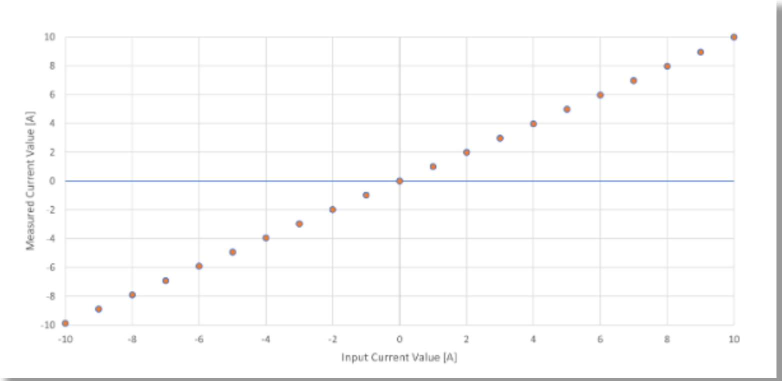
Proposed Method in Fault Detection and Failure Prediction in Asyncronous Motors



Device Setup

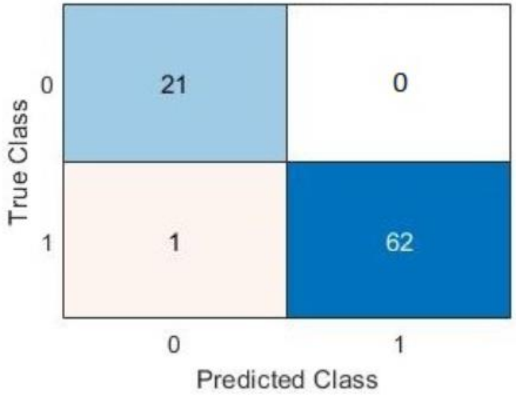
Figure 5.9: MCR1101-20-5 Package

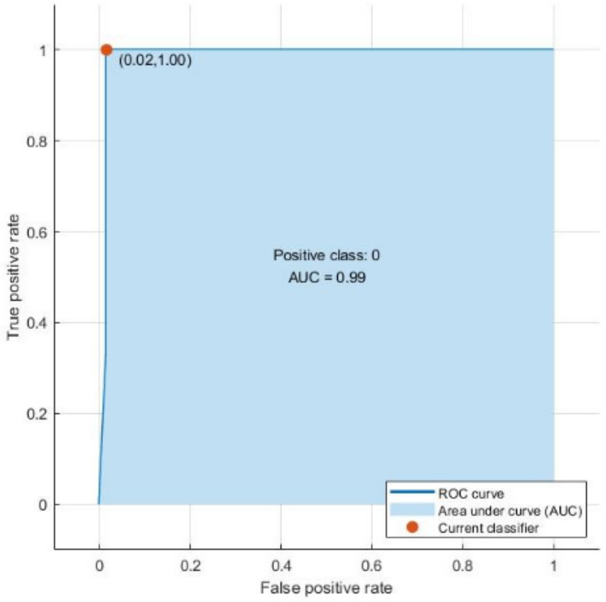




Case Study

Classes	Number of Motors	Class'dimension
Healthy	7	21
Faulty	21	63





Hyperparameter	Range
Number of Fully connected Layer	1 – 3
First Layer Size	1 – 300
Second Layer Size	1 – 300
Third Layer Size	1 – 300
Activation	ReLU; Tanh; Sigmoid
Regularization Strength (Lambda)	1.1905e-07 – 1190.4762
Standardize Data	Yes; No

Conclusions

- Model-driven approaches in Predictive Maintenance comes into play in the management of ordinary and extraordinary maintenance tasks;
- Machine learning techniques have proved to be very useful for realtime diagnostics to reduce machine downtime;
- The connection of devices and the growing computing capacity (cloud computing/edge computing) create new frontiers in the technological transformation of manufacturing companies;
- The new availability of large volumes of data has opened up scientific approaches in the design and organization of industrial processes.

Pubblications

- Gargiulo, F., Duellmann, D., Arpaia, P., & Schiano Lo Moriello, R. (2021). Predicting Hard Disk Failure by Means of Automatized Labeling and Machine Learning Approach. *Applied Sciences*, *11*(18), 8293.

- Arpaia, P., Cesaro, U., Chadli, M., Coppier, H., De Vito, L., Esposito, A., Gargiulo, F., Pezzetti, M. (2020). Fault detection on fluid machinery using Hidden Markov Models. *Measurement*, *151*, 107126.

Q&A

Thanks for the attention!
Questions?

Email: federico.gargiulo@unina.it
Skype: gargiulo.federico