UNIVERSITÀ DEGLI STUDI DI ROMA

LA SAPIENZA

Faculty of Civil and Industrial Engineering

Master Degree in Transport Systems Engineering

Use of neural networks for the predictive maintenance of rolling stock.

Candidate
Francesco Calabrese
1722279

Supervisor
Prof. Eng. Gabriele Malavasi

Co-Supervisor
Eng. Alberto Agnoli

Academic Year 2017/2018

The deep essence of sciences is
to understand that problems exist
but also that we are provided
by many tools to solve them.

Summary

Figures

Tables

1. INTRODUCTION

The life cycle of a machine, of a component, is gaining importance in this last period, representing a budget item inside a company. Preventing failures starts being considered from the project of the machine.

A *Failure* is defined as the event in which an entity is no more able to perform its regular activity: this brings not only to repairing costs but also to indirect costs like:

- Low quality of the product;
- Missed supply of the service;
- …

There are many failure typologies in a machine lifetime [Figure 1]:

- INFANT FAILURE: caused by inadequate construction techniques, solved by the break-in;
- ADULT FAILURE: caused by the normal activity of an element, by its solicitations;
- AGEING FAILURE: caused by a not perfect maintenance.

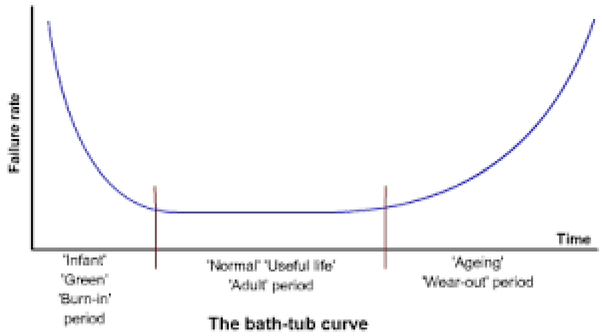


Figure 1: Bath Tub Curve

In this optics the maintenance has gained more relevance; usually it's used to limit the arising of failures and to restore the component once the failure has risen. Now its role is to maximize the life of an element with the minimum global cost: in this contest we speak of *Condition Based Management* (or CBM) that differs from the maintenance at certain time with a maintenance bound to the knowledge of the actual state of the system. This can be done thanks to the Information and Communications Technologies (or ICT), allowing to send data to a calculation central. This system allows managing the so-called Big Data, an enormous amount of heterogeneous data coming from the machine. They need the definition of new tools and methodologies to manage these information in a reasonable time.

Also the study and the implementation of Machine Learning techniques will be useful to manage the Big Data and to predict the behavior of a component and to evaluate the moment in which a component must be changed with a higher saving with respect to the Planned Maintenance.

The aim of this work is the realization of a model able to analyze the Big Data coming from an ICT system, verifying the goodness of the prediction. To realize the model, the program Matlab was used, utilizing some tools able to manage Big Data.

Many different techniques have been developed with example data related to rail system bearings also in collaboration with Bombardier. According to the type of data, a model can fit the problem in the right way or not, this is the problem to cope with, there is not a unique solution that solves any problem, but it depends case by case.

In the first Application Case, general rolling bearings were analyzed for the training of the machine, but the intended use is not known, while in the second Application Case, bearings specific to rolling stock were taken in consideration for the analysis.

2. THE ROLLING STOCK MAINTENANCE

There is not a single maintenance method that fits every situation, but the choose depends on the system characteristics and on the operation conditions. The traditional maintenance, as reported in the Guide Lines of the ANSF [1] (the Italian National Agency for Rail Safety) in collaboration with the CIFI [2],[3],[4] (the Board of Italian Rail Engineers) , is usually divided into:

- CORRECTIVE MAINTENANCE: The set of tasks is destined to correct the defects to be found in the different equipment and that are communicated to the maintenance department by users of the same equipment;
- PREVENTIVE MAINTENANCE: it's characterized by all the processes suitable for maintaining the integrity and the effectiveness of the system, monitoring at the same time the normal degradation. This kind of maintenance is programmed inside the maintenance programs. The critical components are replaced independently from the their conditions. The great limits are the impossibility to replace all the critic components and the possibility to replace an item that has not yet reached its end;
- PREDICTIVE MAINTENANCE: It pursues constantly know and report the status and operational capacity of the installations by knowing the values of certain variables, which represent such state and operational ability. To apply this maintenance, it is necessary to identify physical variables (temperature, vibration, power consumption, etc.). Which variation is indicative of problems that may be appearing on the equipment. This maintenance it is the most technical, since it requires advanced technical resources, and at times of strong mathematical, physical and / or technical knowledge;
- PERIODIC MAINTENANCE: the basic maintenance of equipment made by the users of it. It consists of a series of elementary tasks (data collections, visual inspections, cleaning, lubrication, retightening screws,…) for which no extensive training is necessary, but perhaps only a brief training. This type of maintenance is the based on Total Productive Maintenance;
- CONDITION-BASED MAINTENANCE: it happens just before the break of an element, when an indicator starts signaling that a component is going to break. It uses real-time data to determine system's health in order to act at the correct time. It's a good kind of maintenance, but the costs to install and monitoring are very high;

- REMEDIAL MAINTENANCE: it takes action at the moment in which a malfunction or a more severe failure happens, with consequent stop of the machinery. This is typically the most expensive type of maintenance because the missed use of the system and the repair of the component.

This division of types of maintenance has the disadvantage of that each equipment needs a mix of each of these maintenance types. It's difficult because, when a failure occurs, several steps must be faced:

- IDENTIFICATION: when a problem occurs, identify where and when it happened as well as where and when it did not becomes relevant. Identification problems become relevant not only when trying to understand a situation but also when confusion reigns and the problem is hidden by a mass of effects. The former should be attacked by curiosity and the latter by analysis;
- CAUSE AND EFFECT: typical effects are excessive heat, vibration and noise. A failed bearing or gear is also an effect. Simply changing the component is concentrating on the effect. Forgetting about the reason for the failure is neglecting the cause. Attack the symptom must not forget to unearth the root cause;
- MEANS: these problems are generally characterized by questions .with "how". The problem of selecting a goal or end has already been solved, so now focusing on how to achieve it;
- ENDS: the goals may be very general at first but must be translated into detailed sub-goals to understand what are the critical parts of systems that must be constantly monitored, and how are problems categorized.

The choice of the type of maintenance depends also on the pros and cons; nowadays, the general traditional maintenance has these characteristics:

PROS	CONS
Less risk factor	More money upfront
Follows a schedule	
Longer Life	Over Maintenance
Money savings	
Less energy wastin	More workers
Less disruptions	

Table 1: Pros and Cons of ordinary maintenance

In Italy the guidelines are given by the National Agency for Railway Safety (or ANSF) that, referring to the international standards, defines how to realize maintenance plans, the elements, the documentations…

2.1. Current maintenance operations in rail system

The company carries out several levels of maintenance according to the deepness of the operations, each one made in different centers.

2.1.1. First Level Maintenance

It's carried out in the IMC (or Impianti di Manutenzione Corrente). Preventive and Remedial Maintenances are used, with length or time deadlines: there are periodic controls programmed in the First Level Maintenance Plans that defines also standard times and equipment. Sometimes this type of maintenance can be carried out also when an anomaly is reported, with consequent substitution of the element. Usually it lasts from some hours to several days.

2.1.2. Second Level Maintenance

It's carried out in the OMC (or Officina di Manutenzione Ciclica). The operations are more complex and deeper, sometimes also radical. They are programmed in the Second Level Maintenance Plans, that shows the replacements and the tracking. This type of maintenance is requested between 2 to 6 years, with ten days or also month of works before the restoration.

Several disadvantages are brought by these approaches: first of all, the maintenance is made once the failure has occurred. It's difficult to manage because the weak links between the wearing and the maintenance turns. Also the control systems are expensive.
Also, this type of maintenance obliges the replacement of an element even if it has not reached its lifetime, with other costs also to dismantle the item.

In Figure 2 it's possible to see the deadlines of each type of maintenance for ETR trains, according to the km and to the years and the periods of maintenance.

Cyclic Maintenance Scheme

Visit	Km	Year
Construction	0	0
RT12	600000	2
VIS	1200000	4
RT12	1800000	6
RO	2400000	8
RT12	3000000	10
VIS	3600000	12
RT12	4200000	14
RG	4800000	16
RT12	5400000	18
VIS	6000000	20
RT12	6600000	22
RO	7200000	24
RT12	7800000	26
VIS	8400000	28
RT12	9000000	30

Visit	Main Processes	Working Days	Maintenance Hours
RT12	Gimbals and Valves replacement	9	1250
VIS	Like RT12 + Boogies, Breaking Boxes, Hydraulic Components, Fans and Compressors	45-60	10000
RO	Like VIS + Traction Engines and More Compressors and Fans	50-60	11000
RG	Like RO + Other Particulars	80	12000

Figure 2: Deadlines for maintenance operations ([3])

The codes for the type of operation are reported in Table 2:

CODE	EXPLANATION
RT	Traditional Maintenance
VIS	Intermediate Safety Analysis
RO	Ordinary Maintenance
RG	General Maintenance

Table 2: Maintenance Codes

2.2. The costs of traditional maintenance

Railways produce transit capacity of the rolling stock which principal indicator of the networks is the train-km running on them in a given time interval [4]. But this is not the only indicators, because also pass-km and ton-km must be taken in consideration: it’s possible to say that the direct production is the train-km, while the indirect ones are the pass-km and tons-km. They allow calculating productivity indicators.

The further step is the understanding of the costs rolling stock suffers, especially the one related to the maintenance: these depend on the type of maintenance must be faced, the corrective one or the preventive. The time interval at which the second type of maintenance could be scheduled depends on the expected life distribution of the components, while the corrective maintenance cannot be avoided when a random failure occurs. The total cost will depend on performing both the maintenances. As it was said before, a type of maintenance is not better than another one, but usually the strategic plans take in consideration various combinations, affecting railway safety, passenger comfort and total operation costs.

Every maintenance activity has a certain expiry and duration of operation [Table 3]:

ID	Short Description	Expiry	Duration [h]
I	Pantograph and Bogies Check	Each return to workshop	8
II	Pantograph and Bogies Check	7.500 Km (+/-10%)	8
III	Pantograph and Bogies Check	30.000 Km (+/-10%)	8
IV	Pantograph, Bogies and Wheel Flanges	60.000 Km (+/-10%)	24
V	Check Ultrasonic Flaw Detection	180.000 Km (+/-10%)	72
VI	Wheel Truing	360.000 Km (+/-10%)	108

Table 3: Maintenance operations ([6])

The general problem is solved with the minimum cost (in terms of used units, empty runs and train movements) taking in consideration the timetables, the rolling stock assets and maintenance operations.

A sequential method is developed to solve vary modules linked each other [Figure 3]:

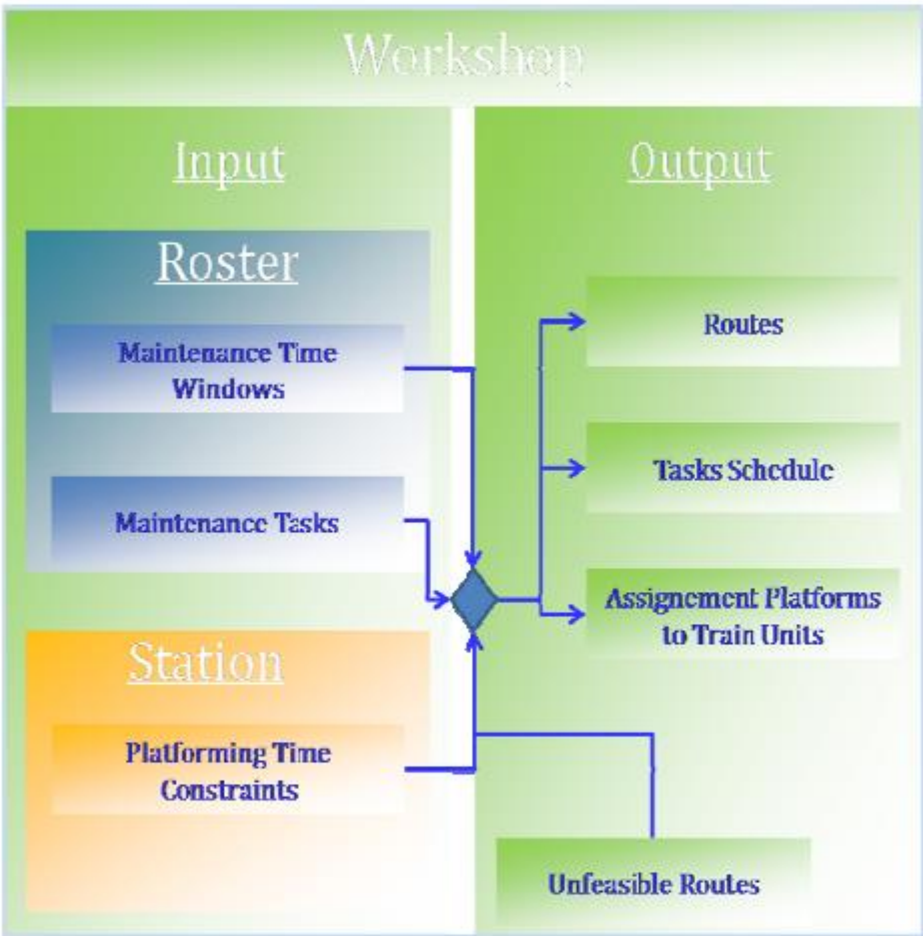


Figure 3: Maintenance Modules ([6])

Usually the rostering problem is the first to be solved and then the other two involving passenger stations and workshop operations.

2.2.1. Rolling Stock Rostering

This module computes rostering to cover a set of commercial services and minimize the costs related to asset units. It optimizes the distance run between consecutive maintenance operations. The problem is solved as a kind of Travelling Salesmen problem with additional constraints and variables to guarantee the respect of maintenance's expiry and guarantee maintenance efficieny. The output is a cyclic roaster including the schedule of maintenance activities and an assigned workshop location.

2.2.2. Station and Workshop Scheduling

There are constraints on maintenance activities that must be performed by each train in the right time interval given by the rostering. In a workshop many units are used at the same time and they must interfere as less as possible in the circulation. So a time window is given that is usually bigger than the sum of all activities and the workshop has some recovery time.

All this procedure must guarantee at every time:

- Safety;
- Train availability;
- Level of Service of the supply;
- Cost control.

As it can be seen, maintenance is a real management activity, monitoring constantly exercise costs and investments, guaranteeing environmental impacts and safety. Maintenance is performed inside the life-cycle of the system, to maintain in time the performances and the efficiency, but also to develop according to the markets.

The maintainability (that it's the probability to perform maintenance in a predetermined time interval) is a parameter that must be well evaluated especially in the project phase of the rolling stock in order to understand:

- Critical areas;
- The right tool availability;
- The development of tools dedicated to maintenance.

A right maintainability must be realized to allow a good efficiency in cost management. In fact the total cost of maintenance is the sum of

- Labor and Tool Cost;
- Service unavailability Cost;
- Replacement Stocking Cost.

In this optic, it’s necessary to maintain low the Life Cycle Cost (or LCC), which is formed by the 40% by the only cost of maintenance.

A problem is related to that rolling stock which haven’t had a maintenance consistent with the maintenance plans in terms of procedures and tools, which recovery will have a higher cost with respect to the estimated one and, usually, without obtaining results.

Also the rise of new technologies has a cost to adapt the tools, the workshop and the labor to them, affecting a lot the maintenance costs.

3. THE MACHINE LEARNING

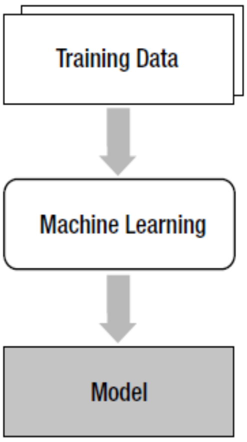


Figure 4: Machine Learning

As it's defined by Phil Kim in his work explaining the many facets of the Machine Learning [7], it is defined as an analytic method able to realize models automatically. The term was coined by Arthur Lee Samuel but the best definition of Machine Learning was given by Tom Michael Mitchell:

"A machine learns from the experience E
with reference to some classes of tasks T
and with performance measurement P,
if its performances in task T, as measured by P,
gets better with experience E."

This sentence says that a machine learns from the experience: there is learning when the performances of the program after the development of a task are better than the past. There are many types of Machine Learning, according to the presence or not of complete example in the learning phase.

3.1.1. Supervised Learning

The machine has datasets as inputs and information regarding the desired results with the aim that the system identifies a general model linking inputs and outputs. Inputs and outputs must be in pairs.

3.1.2. Unsupervised Learning

The machine has datasets as inputs but no information regarding the desired results: the machine must find a scheme and a hidden model, identifying in the inputs a logical structure.

3.1.3. Reinforcement Learning

It's a supervised learning in which also the error is included: the learning process is a sequence of reward and penalty.

3.1.4. Semi-supervised Learning

It's a hybrid model where the inputs are not completed: some inputs have the corresponding result, others no.

3.1.5. Other practical approaches

They are practical applications of Machine Learning, like the *Decision Trees* (with which it's possible to define the outputs graphically), the *Clustering* (where every element is grouped according to mathematical models), *Probabilistic Methods* (basing the calculation on the statistical methods) and also the *Neural Networks*.

It is important to distinct which are the training data to predict the model and which are the input data used to find outputs once defined the model. The distinctness of these two types of data defines heavily how well the generalization is accomplished. One of the primary causes of corruption of the generalization process is *overfitting*: it's the error of division of the data in which the bounder doesn't divide the sets correctly and some data can be not in the correct group [Figure 5]. A model is over-fitted when the trained model brings to low level of performances.

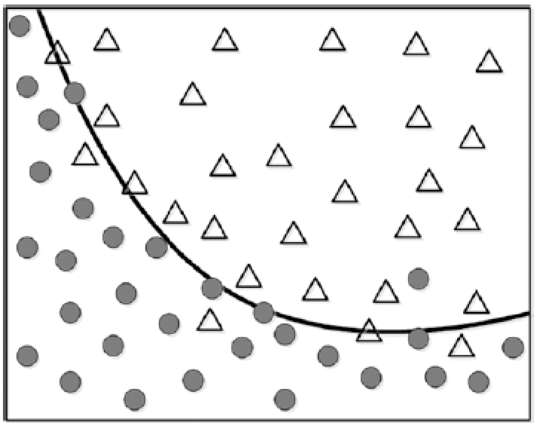


Figure 5: Overfitting

There are two methods to confront this problem:

- **REGULARIZATION** is a method that to construct the simplest numerical model, avoiding overfitting effects at small costs of performances.
- **VALIDATION** is a process that reserves a part of the training data to monitor performances. This set is not used for the training process [Figure 6]. When validation is involved, the process follows these steps:

  - Division of data;
  - Train the model with training set;
  - Validate the model with validation set.

  If the model yield to good performances, then the training is stopped, otherwise another model must be implemented and verified.

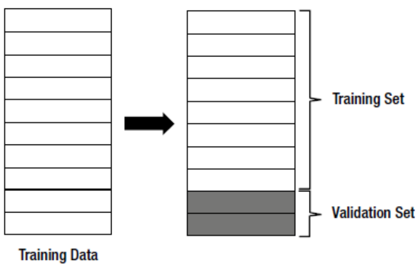


Figure 6: Division of data

Another type of validation is the cross-validation: there is a repetition of division of the data. It maintains the randomness of the division but better detect the overfitting of the model [Figure 7].

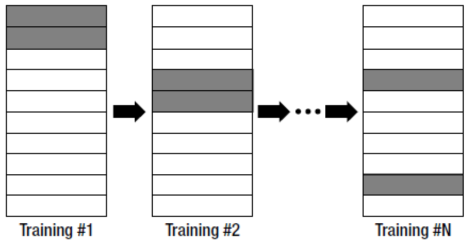


Figure 7: Cross-Validation

3.2. Neural Networks

Neural Networks are a new informatics technology that has a strong analogy with the brain structure, with many neurons strongly connected by synapses [Figure 9]. This analogy is represented by:

- Strong parallelism;
- Strong interconnection;
- Fast communication;
- Simple elaborative units;
- Big memory;
- Not programmed, but trained by automatic learning.

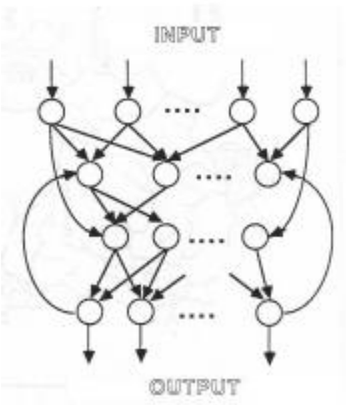


Figure 9: A neural network

The activity of a single unit is simple (represented by a transferability function) and the power of the model is inside connection configuration. A unit is not programmed, but trained by examples coming from the reality.

There are many types of neural networks, according to the model, to the learning method and to the aim. In any case, it's possible to characterize a General Neural Network Model (in Figure 8 they are represented the general characteristics):

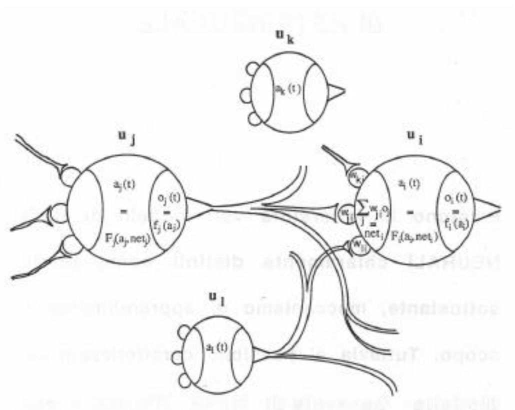


Figure 8: General Neural Network

- $a_j(t)$ = activation value of the j-th unit;
- net_j = propagation function;
- $F_j(a_j, net_j)$ = activation function;
- $o_j(t)$ = output;
- $f_j(a_j)$ = output function;
- w_{ij} = connection weight from unit j to unit i.

Each element is represented by a formula that was elaborated in precedent work, becoming spread inside this topic.

3.2.1. Units

Units are simple and uniform. Its activity consists on receiving an input among a set of inputs and calculates an output value to send to another set of units. The system is parallel because many units can effectuate their computations parallel. There are 3 types of units: input units, output units and hidden units.

3.2.2. Connections

A connection matrix **W** is defined, in which each element w_{ij} is characterized by:

- Departure unit j;
- Arrival unit i;
- Connection force, that can be higher than 0 (excitatory connection), lower than 0 (inhibitory connection) or null value.

3.2.3. Activation Value

Every unit is characterized by an activation value at time t: $a_j(t)$. Different models realize different assumptions regarding the activation values the units can assume:

- **Discrete Values**
  - Binary;
  - Restricted values.
- **Continuous values**
  - Limited;
  - Unlimited.

3.2.4. Output

Units interact each other's transmitting signals to the neighborhood units. An output function is associated to each unit that transforms the current activation value into the output signal $o_j(t)=f_j(a_j(t))$.

The function can assume different values.

3.2.5. Propagation

It's the way in which outputs propagate along the connections. It occurs independently for each type of connection, in case of more different connections.

Usually the generic rule expects simply the sum of the inputs weighted by connection intensity:

$$net_{ai} = \sum_j w_{aji} o_j$$

3.2.6. Activation function

It's a function that takes the current activation value and the propagation value and gives back a new activation value:

$$a(t+1) = F(a(t), net_1(t), net_2(t), \dots)$$

Usually F could be:

- A threshold function;
- A sigmoid function;
- A stochastic function.

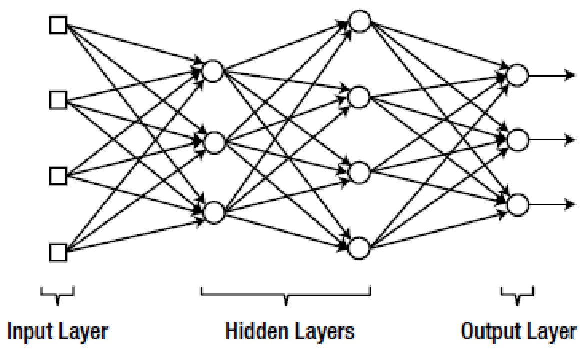


Figure 10: Neural Network Layers

Sets of nodes are structured on various layers [Figure 10]: the input layer transmits the input signals to the next nodes but doesn't calculate the activation function. The exiting results of the output layers are the results of the model. The hidden layers have this name because they cannot be accessed by the outside of the NN.

A simple-layer network is composed only by the input and the output layer, while a multi-layer one includes also one or more hidden layers: in particular, the presence of only one hidden layer is defined as shallow neural network, otherwise *deep neural network* [Table 4].

Single Layer Neural Network	Input Layer – Output Layer

Multi-Layer Neural Network	Shallow Neural Network	Input Layer – Hidden Layer - Output Layer
	Deep Neural Network	Input Layer – Hidden Layers - Output Layer

Table 4: Neural Network Recap

Neural Networks are not programmed but they auto define themselves by an automatic learning. This learning consists in modification of the connections actuated by a learning rule, modification that can occur by:

- Creation of new connection;
- Losing of existing connection;
- Modification of the weight.

All the learning rules can be defined variants of *Hebb's rule* enunciating:

"If a unit u_i receives an input from a unit u_j and if both the units are strongly active, the weight w_{ij} from u_j to u_i must be reinforced"

expressed by a function,

$$\Delta w_{ji} = g(a_i(t), d_i(t)) \cdot h(o_j(t), w_{ji})$$

saying that the change of the connection between the two units is the product of the function g (depending on the activation value a_i and a training input d_i) and function h (depending on the output o_j and weight w_{ji}). The simplest version of this rule there is no trainer d_i and g and h are linear functions to the first argument, obtaining the *classical Hebb's rule* (where η represents the learning speed)

$$\Delta w_{ji} = \eta a_i o_j$$

An important variant is the ***Delta rule***, where the learning is proportional to the variation (or delta) between the current actual value and the desired one, provided by the trainer:

$$\Delta w_{ji} = g(d_i(t) - a_i(t)) \cdot o_j$$

This equation can be generalized as

$$w_{ji} \leftarrow w_{ji} + \alpha \delta_j x_i$$

where

$$\delta_j = \varphi'(v_j) e_j$$

In this equation there are

- e = error of the output node;
- v = weighted sum of the output node;
- ϕ' = derivative of the activation function ϕ.

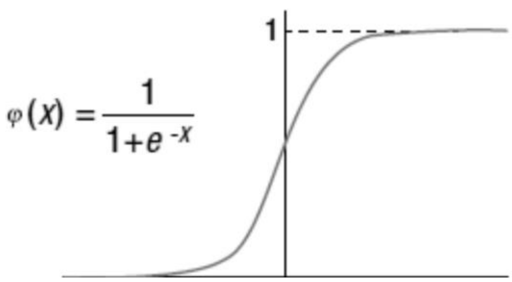


Figure 11: Sigmoid Function

Widely the activation function is represented by the sigmoid function [Figure 11]. The first derivative of this function is given by

$$\varphi' = \varphi(x)(1 - \varphi(x))$$

Substituting in δ it's obtained

$$\delta_j = \varphi(v_j)(1 - \varphi(v_j)) e_j$$

Finally, the following expression is obtained:

$$w_{ji} \leftarrow w_{ji} + \alpha\varphi(v_j)(1 - \varphi(v_j))e_j x_i$$

Unfortunately the delta rule is not good to train a multi-layer neural network because the error, main element of the method, is not defined in the hidden layers (error given as difference between the given output and the neural network output). The introduction of the back-propagation algorithm in 1986 has solved the problem of training a multi-layer network because it provides a method to determine the error [Figure 12].

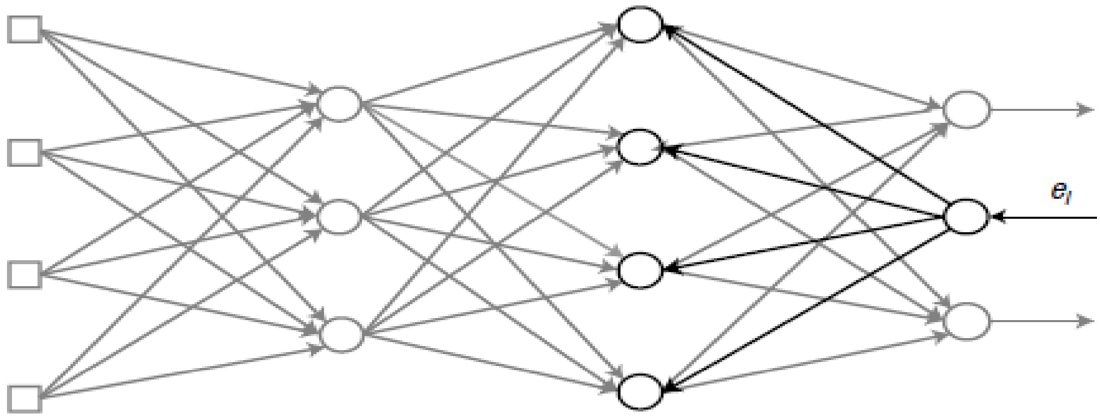


Figure 12: Back-propagation algorithm

3.3. The back-propagation algorithm for neural networks

The delta of the output nodes is defined as

$$e = d - y$$
$$\delta = \varphi'(v)e$$

where v is the weighted sum of the forward signal at the respective node. Forward and backward processes are identically applied to the hidden nodes as well as the outputs nodes. This implies that the outputs and hidden nodes experience the same backward process. In summary, the error of the hidden nodes is calculated as the backward weighted sum of the delta, and the delta of the node is the product of the error and the derivative of the activation function. This process begins at the output layer and repeats for all the hidden layers.

The algorithm will be so composed:

1) Initialize the weights with adequate values;
2) Enter the input from the training data and obtain the neural network's output. Calculate the error of the output to the correct output and the δ;

$$e = d - y$$
$$\delta = \varphi'(v)e$$

3) Propagate the output node delta backward and calculate the deltas of the immediate next nodes;

$$e^{(k)} = W^T\delta$$
$$\delta^{(k)} = \varphi'\left(v^{(k)}\right)e^{(k)}$$

4) Repeat Step 3 until it reaches the last hidden layer;
5) Adjust the weights;

$$\Delta w_{ji} = \alpha\delta_j x_i$$
$$w_{ji} \leftarrow w_{ji} + \Delta w_{ji}$$

6) Repeat from step 2 for each training point until the neural network is properly trained.

3.4. The Deep Learning

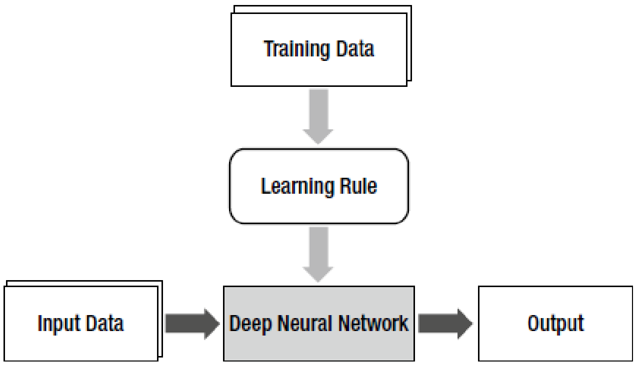


Figure 13: Deep Learning Scheme

The Deep Learning is the multi-layer neural network, composed by two or more hidden layers [Figure 13]. The single-layer network had some known limitations, but it took over 30 years before other hidden layers were added, precisely with the introduction of the back-propagation algorithm. But initially there were problems regarding its performances. There were various attemps to overcome these limitations (as adding hidden layers and adding nodes in the hidden layer), but none worked well, someone also gave poorer results. These limitations sentenced the neural networks to oblivion, till the 2000s in which the concept of Deep Learning was introduced. The current technologies yielded dazzling levels of performances that overcame other learning techniques. The problem in the transition from single to multi-layer network was the lack of a learning rule and the second problem was related to the low levels of performances.

Deep Learning innovation is related to many small technologies improvements. The reason of the late development of this technique was related to three difficulties of the back-propagation algorithm:

- **VANISHING GRADIENT:** the gradient is a similar concept to the delta. The vanishing gradient occurs when the output error is more likely to fail to reach the farther nodes. It follows that the hidden layers closer to the input one are not well trained. The solution is the use of the Rectified Linear Unit (or RLU) function as activation function defined as

$$\varphi(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases}$$

Its first derivative will be

$$\varphi'(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

-

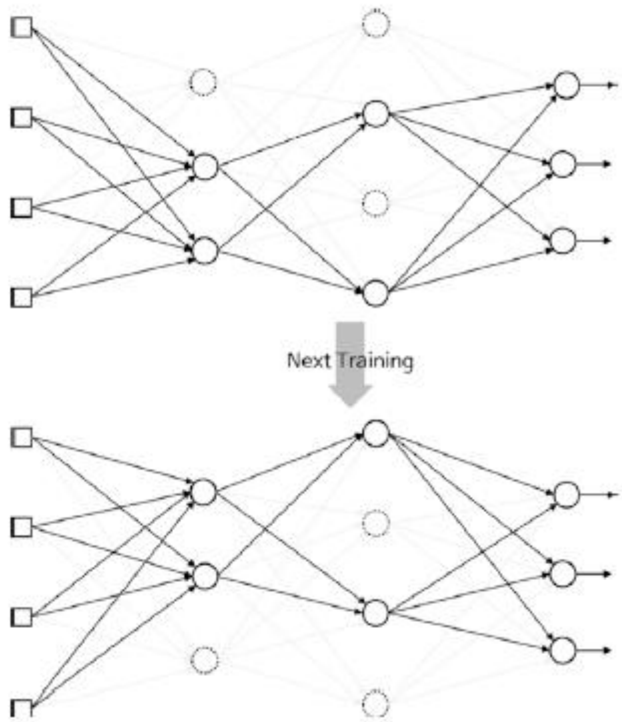


Figure 14: Dropout

- **OVERFITTING:** the model becomes more complicated adding more hidden layers and so more weights. The most used solution is the dropout, which trains only some of the randomly chosen nodes than the entire network. It's effective and the implementation not very complex. On each iteration nodes are randomly chosen and their outputs are set to zero to deactivate them [Figure 14].
  The dropout prevents overfitting as it continuously alters the nodes and the weights. Another solution is adding regularization terms, simplifying the network as much as possible.

- **COMPUTATIONAL LOAD:** it's the time to complete the challenge. The increment of hidden layers (and weights) needs more training data and also the computation time increases.

3.5. The Big Data

Many surveys were written about Big Data and its relationship with Machine Learning [8],[9]: Big Data generally refer to data that exceeds the typical storage, processing, and computing capacity of conventional databases and data analysis techniques. As a resource, Big Data requires tools and methods that can be applied to analyze and extract patterns from large-scale data. The rise of Big Data has been caused by increased data storage capabilities, increased computational processing power, and availability of increased volumes of data, which give organization more data than they have computing resources and technologies to process. The unmanageable large Volume of data poses an immediate challenge to conventional computing environments and requires scalable storage and a distributed strategy to data querying and analysis. A general theme in Big Data systems is that the raw data is increasingly diverse and complex: working with this variety among different data representations poses unique challenges with Big Data, which requires Big Data preprocessing of unstructured data in order to extract ordered representations of the data.

Most of the current technologies that are used to handle Big Data challenges are focusing on seven main issues of that called Volume, Velocity, Variety, Veracity, Validity, Volatility and Value.

- **VOLUME:** we are facing with huge amounts of data that most of traditional algorithms are not able to deal with this challenge (amount that can reach also ZB of data). The definition of high volume is not specified in predefined term and it is a relative measure depends on the current situation of an enterprise;
- **VARIETY:** we are facing with variety types of file formats and even unstructured ones. These data should be unified for further processes;
- **VELOCITY:** data are coming in a very fast manner, the rate at which data are coming is striking that may hang the system easily. It shows the need for real-time algorithms;
- **VERACITY & VALIDITY:** data must be as clean, trustworthy, usefulness, result data should be valid, as possible for later processing phases. The more data sources and types, the more difficult sustaining trust;
- **VOLATILITY:** how much time data should remain in the system so that they are useful for the system;
- **VALUE:** the amount of hidden knowledge inside Big Data.

We also can consider open research problems from another viewpoint as follows, six parameters: Availability, Scalability, Integrity, Heterogeneity, Resource Optimization, and Velocity (related to stream processing).

- **AVAILABILITY:** Means data should be accessible and available whenever and wherever user requests data even in the case of failure occurrence;
- **SCALABILITY:** if a system supports large amounts of increasing data efficiently or not;
- **INTEGRITY:** points to data accuracy. The situation becomes worse when different users with different privileges change data in the cloud. Cloud is in charge of managing databases;
- **HETEROGENITY:** refers to different types of data such as structured, unstructured and semi-structured;
- **RESOURCE OPTIMIZATION:** means using existing resources efficiently. A precise policy for resource optimization is needed for guaranteeing distributed access to Big Data;
- **VELOCITY:** means the speed of data creation and data analysis. The need for real-time analyses is obligatory. These are very application dependent that means can differ for each application to another application.

Big Data area can be divided into three main Phases: Big Data Preprocessing, means doing some preliminary actions toward data with the aim of data preparation such as data cleansing and so on. Big Data storage means how data should be stored. Big Data management means how we should manage data.

3.5.1. Preprocessing

Preprocessing data means transforming, inconsistency, incomplete data that have many errors into an appropriate format for further analyses. In other words, data must be structured prior to analysis stage for further processing and analysis. There are some steps for achieving preprocessing section goal as described as follows:

1. Data cleansing: Removing inaccuracies, incompleteness, and inconsistencies of data;
2. Data transformation: Means doing additional processes like aggregation, or transformation;
3. Data integration: It provides a single view over distributed data from different sources;
4. Data transmission: Defines a method for transferring raw data to storage system such as object storage, data center or distributed cloud storage;
5. Data reduction: reducing the size of large databases for real-time applications;
6. Data discretization: it refers to attribute intervals so that obtained values will be reduced.

The preprocess phase includes the following sub-sections:

3.5.2. Data transmission

It means sending raw data to data storage. One example of proposed method in this area is sending data through a high-capacity pipe from data source to data center. This type of transmission needs to know networks architecture along with transportation protocol.

3.5.3. Data cleansing

It means detecting incomplete and irrational data. It's possible to modify or delete these kinds of data in order to achieve quality improvement for further processing steps. There are five stages in order to achieve clean data:

1) Recognizing types of errors;
2) Finding error instances;
3) Correct error instances and error types;
4) Update data input procedure in order to reduce further errors that may occur;
5) Checking data affairs like limitations, formats, and rationalities.

Data cleansing is an indispensable and principal part of data analysis step.

In brief, there are two main problems in data cleansing step:

i) Data are imprecise;
ii) Data are incomplete.

3.5.4. Stream processing

The stream requirements are completely different with traditional batch processing. In more detail, there are some emerging applications producing large amounts of dedicated data to servers in order to real-time processing. While large volumes of data are received by servers for processing, it’s not possible to use traditional centralized techniques, but there are many open research topics:

1. Data Mobility: the number of steps that are required to get the final result;
2. Data Division or Partitioning: The algorithms are used for partitioning data. In the brief, partitioning strategies should be used in order to achieve better data parallelism;
3. Data Availability: a technique that guarantees data availability in case of failures occurrence;
4. Query Processing: a query processor for distributed data processing efficiently with considering data streams. One possibility of this is doing deterministic processing (always get the same answer) and another one is non-deterministic (the output depends on the current situation) one;
5. Data Storage: Another open research problem in Big Data is how to store data for future usage;
6. Stream Imperfections: Techniques dealing with data stream imperfections like delayed messages or out-of-order messages.

3.5.5. Data storage

Storing data in petabyte scale is a challenge not only for researchers but also for internet organizations. Although Cloud Computing reveals a shift to a new computing paradigm, it cannot assure consistency easily when storing Big Data in cloud storage. It is not a good way to waste data since it may contribute to better decision-making. So it is critical to have a storage management system in order to provide enough data storage, and optimized information retrieval.

3.5.6. Replication

Replication is a big activity that makes data available and accessible whenever user asks. When data are variable, the accuracy of each replicated copy is much more challenging. The two factors to take in consideration are replication sites and consistency. These two factors play more important role in Big Data environment as managing these huge amounts of data are more difficult than usual form

3.5.7. Indexing

Indexing data improves the performance of storage manager. So proposing a suitable indexing mechanism is challenging. There are three challenges in indexing area

1) Multi-variable and multi-site searching;
2) Performing different types of queries;
3) Data search when they are numerical.

A new method uses a tree based index structure and adopts sharing of a single list of event indices to speed up query responses. Index load time is a challenge now same as space consumption. A Support Vector Machine indexing algorithm was introduced: It changes transition probability calculation mechanism and applies different states to determine the score of input data. While it produces a relatively accurate query result with minimum time, it is time-consuming in learning process. A fuzzy-based method can be used for indexing of moving objects where indexing images are captured during object's movements. It provides a trade-off between query response time and index regeneration. The index supports data variables and it is scalable.

3.5.8. Big Data Management and Processing

There are four types of data models in Big Data area:

- Data stored in relational
- Graph data
- Semi-structured data (e.g. XML)
- Unstructured data (texts, hand-written articles)

Most of the relational databases are not designed to scale to thousands of loosely coupled machines. Because of two reasons, companies tended to leave traditional databases: the first one is traditional databases are not scalable and the second one is that using non distributed traditional database along with adding layers on top is very expensive. So companies decided to implement their own file system, distributed storage systems, distributed programming frameworks and even distributed database management systems.

Furthermore, Big Data management is a complex process especially when data are gathered from heterogeneous sources to be used for decision-making and scoping out a strategy. About 75 percent of organizations apply at least one form of Big Data. Big Data management area brought new challenges in terms of data fusion complexity, storage of data, analytical tools and shortage of governance.
In Figure 15 known methods of storage, pre-processing and processing of Big Data are reported, defining problems and proposed solutions while in Figure 16 a comparison of the methods is shown.

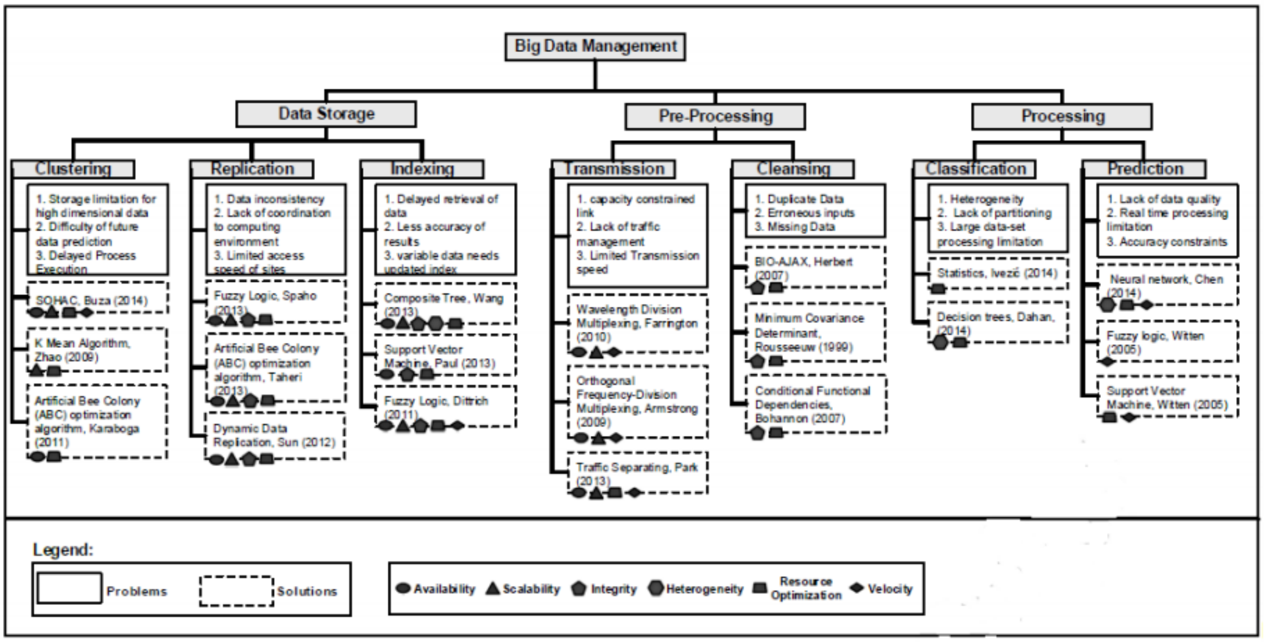


Figure 15: Big Data Management methods ([9])

BDM Activity		Relative Method	Authors	Availability	Scalability	Integrity	Heterogeneity	Resource Optimization	Velocity
Data Storage	Clustering	Storage-Optimizing Hierarchical Agglomerative Clustering- SOHAC	Buza, Nagy et al. (2014)	Yes	Yes	N/A	No	Yes	Yes
		K-Mean	Zhao, Ma et al. (2009)	No	Yes	N/A	No	Yes	No
		Artificial Bee Colony (ABC) optimization algorithm	Karaboga and Ozturk (2011)	Yes	No	N/A	No	Yes	No
	Replication	Fuzzy Logic	Spaho, Barolli et al. (2013)	Yes	Yes	Yes	No	Yes	No
		Artificial Bee Colony (ABC) optimization algorithm	Taheri, Choon Lee et al. (2013)	Yes	Yes	Yes	No	Yes	No
		Dynamic Data Replication	Sun, Chang et al. (2012)	Yes	Yes	Yes	No	Yes	No
	Indexing	Composite Tree	Wang, Holub et al. (2013)	Yes	Yes	Yes	Yes	Yes	Yes
		Support Vector Machine	Paul, Chen et al. (2013)	Yes	No	Yes	No	Yes	No
		Fuzzy Logic	Dittrich, Blunschi et al. (2011)	Yes	Yes	Yes	No	Yes	Yes
Pre-Processing	Transmission	Wavelength Division Multiplexing (WDM)	Farrington, Porter et al. (2010)	Yes	Yes	N/A	N/A	N/A	Yes
		Orthogonal Frequency-Division Multiplexing (OFDM)	Armstrong (2009)	Yes	Yes	N/A	N/A	N/A	Yes
		Traffic Separating	Park, Yeo et al. (2013)	Yes	Yes	N/A	N/A	Yes	Yes
	Cleansing	BIO-AJAX	Herbert and Wang (2007)	N/A	N/A	Yes	No	Yes	N/A
		Minimum Covariance Determinant (MCD)	Rousseeuw and Driessen (1999)	N/A	N/A	Yes	No	Yes	N/A
		Conditional Functional Dependencies (CFD)	Bohannon, Fan et al. (2007)	N/A	N/A	Yes	No	Yes	N/A
Processing	Classification	Statistics	Ivezić, Connolly et al. (2014)	N/A	N/A	N/A	No	Yes	No
		Decision trees	Dahan, Cohen et al. (2014)	N/A	N/A	N/A	Yes	Yes	No
	Prediction	Neural network	Chen, Xu et al. (2014)	N/A	N/A	N/A	Yes	Yes	Yes
		Fuzzy logic	Witten and Frank (2005)	N/A	N/A	N/A	No	No	Yes
		Support vector machine	Steinwart and Christmann (2008)	N/A	N/A	N/A	No	Yes	Yes

Figure 16: Big Data Methods Comparison ([9])

3.5.9. Deep Learning for Big Data

Deep Learning deals mainly with two V's of Big Data characteristics: Volume and Variety. It means that Deep Learning are suited for analyzing and extracting useful knowledge from both large huge amounts of data and data collected from different sources. When we want to apply Deep Learning, we face some challenges:

- *<u>DEEP LEARNING FOR HIGH VOLUME DATAS</u>*
    - We apply Deep Learning algorithms in a portion of available Big Data for training goal and we use the rest of data for extracting abstract representations;
    - Another open problem is domain adaptation, in applications which training data is different from the distribution of test data;
    - Another problem is defining criteria for allowing data representations to provide useful future semantic meanings;
    - Another one is that most of the DL algorithms need a specified loss;
    - The other problem is that most of them do not provide analytical results that can be understandable easily;
    - Deep Learning seems suitable for the integration of heterogeneous data with multiple modalities due to its capability of learning abstract representations;
    - They need labeled data.

- *<u>DEEP LEARNING FOR HIGH VARIETY DATAS</u>*

These days, data come in all types of formats from a variety sources, probably with different distributions. There are open questions in this regard that need to be addressed.

- *<u>DEEP LEARNING FOR HIGH VELOCITY DATAS</u>*

Data are generating at extremely high speed and need to be processed at fast speed. One solution for learning from such high-velocity data is online learning approaches that can be done by deep learning. Only limited progress in online deep learning has been made in recent years.

3.6. Future Progresses

3.6.1. Big Data Preprocessing

One challenge is data integrity that means sharing data among users efficiently. Even though, data integration definition is not much clear in most of the applications. The two challenging research topics in this field are generating and facilitating integration tools. The quality of data is not predetermined. After using data, we are able to find its quality. The more quality data, the better results. Data providers demand error-free data and it is relatively impossible to use only one method of data cleaning to achieve the best quality data is a challenge.

3.6.2. Big Data Analytics

It relates to database searching, mining, and analysis. With the usage of data mining in the big data area, a business can enhance its services. Big Data Mining is a challenge because of data complexity and scalability. The two salient goals of data analyses are: first detecting relationships between data features and second predicting future instances. Additional applications and Cloud infrastructures are needed to deal with data parallelism. Algorithm orders increase exponentially with the increase of data size. There are four types of analyses in simple words:

- Descriptive Analysis: What is happening in data now;
- Predictive Analysis: What will happen in the future;
- Discovery Analysis: Discovering an existing rule from existing data;
- Perceptive Analysis: What should we do in future based on current data.

3.6.3. Semantic Indexing

Another usage of DL and open challenge is using it for semantic indexing with the aim of better information retrieval. It means we should store semantic indexes rather than storing as raw data bytes due to massive amounts of data and low storage capacity.

3.6.4. Data Governance

It is another important core of Big Data Management and it means defining rules, laws and controlling over data. One example is that if Big Data should be stored in the cloud, we must take some policies like which type of data needs to be stored, how quickly data should be accessed, rules for data such as transparency, integrity, check and balances, and last but not the least change management.

3.6.5. Big Data Integration

It means collecting data from multiple sources and storing them with the aim of providing a unified view. Integrating different types of data is a complex issue that can be even worse when we have different applications. Many open research topics are associated with data integration like real-time data access, the performance of the system, and overlapping of the same data.

4. THE PREDICTIVE MAINTENANCE

The predictive maintenance is a process that takes benefits from the machine learning algorithms with predictive capacities. The process is related on the forecasting of the conditions that are going to happen and on the moment in which a machine will break [10].

Ideally, predictive maintenance has to reduce the maintenance frequency, preventing unplanned reactive maintenance, without incurring in the costs associated with doing too much preventive maintenance. Prediction can be done with one of many techniques, which must be effective at predict failures with a sufficient warning time.

To build a failure model, historical data are required, allowing capturing information about events leading to failure. The need to implement data driven strategies emerged in the 90s when relational database were already widespread. Also general static features can provide valuable information, like mechanical properties, average usage and operating conditions. The life span of machines is usually in the order of years, so the data must be collected for long periods to observe the system throughout its degradation process. In this way maintenance work can be better planned, transforming unplanned stops into planned ones, shorter and fewer.

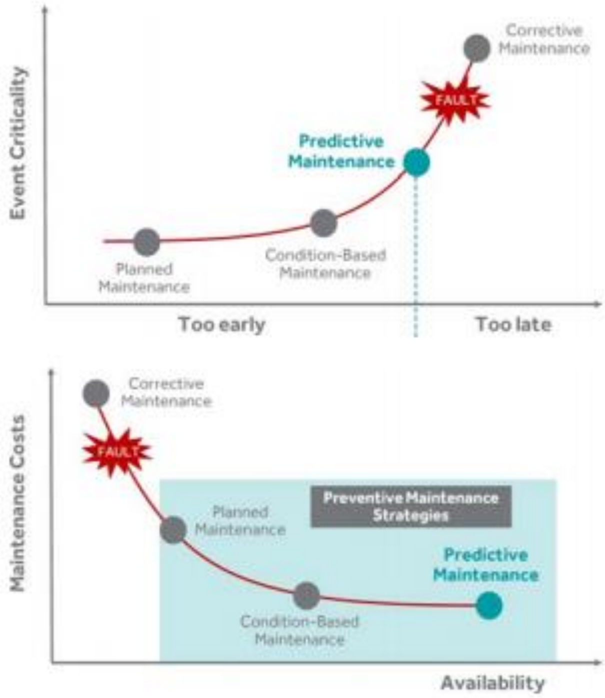


Figure 17: Maintenance comparison

This type of maintenance utilizes non-invasive technologies such as infrared, acoustic, corona detection, vibration analysis, sound level, oil analysis… These tools are used to measure the actual equipment together with measurement of process performance. According to the analyzed item, a technology is better than another one else: for example, a high-speed rotating element can be measured by a vibration analysis, evaluating the condition of the equipment and avoiding the failure [11].

The introduction of the Internet of Things (or IOT) allows the elements to geo-localize themselves, measure their own parameters and communicate to the software, to the analytic systems and to the AI platforms [12],[13]. The Machine Learning uses the data coming from different sources, internal and external, to feed and to train the AI systems, data like historical series of the performances of the components, contextual information, real data inputs. AI platforms learn to determine by statistical calculation the cause-effect relationship and, in this way, forecast (more and more reliably) the behavior of the machines and the items in real and simulated situations, forecasting failures, disservices and malfunctions. This means passing from a protocolled maintenance (based on periodic and, maybe, unnecessary interventions) to a new model, a system that is able to alert what kind of critical issues and with which kind of probability they can happen. It's possible in this way to analyze all the advantages related to this approach:

- Fast identification of anomalies in the machines and avoid economic, environmental and safety consequences;
- Possibility of control and of reduction of the costs with the same quality;
- Possibility to enlarge and maximize the life of the system;
- Possibility to measure the performances and do the maintenance when it's required.

In Figure 17 it's possible to compare the availability and the cost efficiency of the Predictive Maintenance with respect to the other types of maintenance.

4.1. The Predictive Maintenance for the rail system

In rail industry the predictive maintenance uses data collected on equipment during operation to identify maintenance issues in real time. In this way maintenance can be properly planned, avoiding the possibility to take out of service a train because of an emergency or of an unnecessary routine control. If it is possible to forecast which parts are likely to fail in the near future, this will lead to the possibility to achieve a value close to the 100% of availability, because the faults are fixed according to an efficient planning when units are out of service, avoiding breakdowns. These are techniques that have started to be used in rail systems recently, but obtaining good results in terms of performances [14],[15].
As the number and heterogeneity of the rolling stock increases, rail companies have to face the problems related to traditional approaches to maintenance. A trend toward digitalization has emerged to tackle recurrent problems. The first implementation was the Maintenance Information System (or MIS) that allows, in real time, to check the operating status of each individual maintenance plan, to know the progress of each intervention, to build databases with the history of interventions according to the type and to the plant and to obtain statistical data on the various types of operations.

Recently the rail industry, with the increment of new computational and wireless technologies, has added a further step to the digitalization of maintenance with the introduction of the tele-diagnostics, able to show and store data not only on the train but also on the monitors of the control rooms. Usually a tele-diagnostic architecture is composed by two subsystems:

- ON-BOARD SUBSYSTEM: it contains a server that collects data on a local database and analyzes them. It's important the logic inside, able to model the rolling stock events. The system will send diagnostic data to the ground with two different channels:
    - *Communication in near real time:* variables are constantly sent to the ground with protocols, allowing monitoring the fleet in operation;
    - *Batch Communication:* all the signals collected according to the various events are sent at regular interval.
- ON-GROUND SUBSYSTEM: it's a convergent IT solution communicating with the other subsystem, sending data and signals to the various stakeholders and sending maintenance warnings.

Usually a fleet of trains lasts a long time because a strategic focus tends to keep them in service for as long as possible to get value out of the considerable initial investment. In this optic, a technology enabling predictive maintenance, able to reduce operating costs and extend fleet's lifetime, can have the potentiality to deliver huge financial rewards. But this brings also the necessity to renew the technology of the older trains.

It's important to predict the failure of the most critical systems, among the ones that levels huge amount of data to build a consistent predictive model. Also look at system selection would be a good approach, finding where prediction is more effective from a maintenance point of view, for example mechanical and electrical systems that follow a bathtub curve. The choice of the data must remember the objectives behind their collection.

The Predictive Maintenance in a rail system can be performed following two different approaches:

- KNOWLEDGE-BASED: it considers competencies and know-how acquired by designers and maintainers and utilizes Failure Mode, Effects and Criticality Analysis and analysis of Reliability, Availability, Maintainability and Safety. It's possible to identify a priori the train abnormal behavior;

❖ DATA-DRIVEN: the maintenance has an increasing volume of multi-source and heterogeneous data. It's possible to utilize distributed file systems and big data platforms. It has been created the conditions to identify the relationships between apparently independent data, extracting insights, devices' behavioral patterns and dependencies previously unknown and now usable to predict abnormal behaviors.

To perform an effective railway predictive maintenance, four steps must be followed:

1) PREDICTION POSSIBILITIES AND RELATED EFFECTIVENESS: an effective selection of trains' subsystems is fundamental. Critical events that leave enough digital footprints is crucial to build a consistent predictive model: trying to predict everything could lead to misleading results and wasting resources. It's required to map processes to obtain not only the areas in which the occurrence of an event has a higher frequency, but also graphs representing the period in which the predictions are more effective over the useful life of the system [Figure 18]. The goal is to identify the prediction feasibility of the most critical subsystems, with the risk that they can provide little data to build a model;

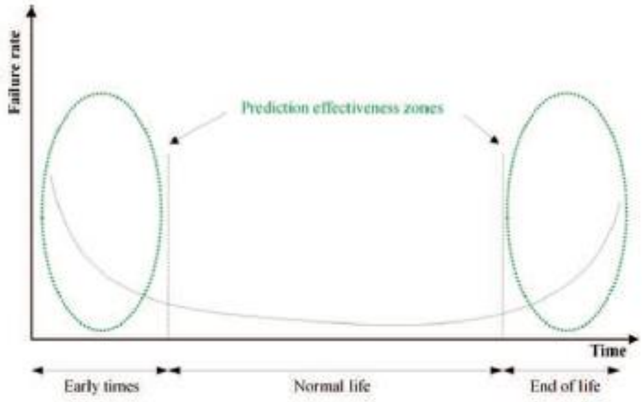


Figure 18: Effectiveness of the prediction

2) EXTRACT THE RIGHT DATA: two factors must be taken to deal with: all the variables potentially assessable and the measurement techniques. There are many elements from which it's possible to obtain data (f.e. the axes, the bogies, the wheel bearings…) and different techniques to collect digital values (f.e. the speed rotation, the temperature, vibrations…). Normal mechanics failure modes degrade at a speed directly proportional to their severity: if the problem is detected early, major repairs can be prevented;

3) The success of the predictive maintenance consists in the careful selection of the train systems to be analyzed, in the construction of an appropriate data ecosystem and in the right combination of experts in railway field. It's essential to identify the evolution of the fault in function of time: this can be achieved identifying mathematical relations able to describe the phenomenon, foreseeing its evolution in function of time and operation. This step can be split in three phases:

a. *Learning Phase:* the available time series are analyzed to identify general rules and patterns;
b. *Formalization Phase:* known and precise rules are created to codify the knowledge of the different failure modes;
c. *Execution Phase:* the models constructed and calibrated are validated through the application to the historical time series in order to test the real predictive effectiveness and evaluating a further fine-tuning.

4) IDENTIFY THE ACHIEVABLE VALUE-ADD: this approach is not only able to bring to an effective predictive maintenance solution to predict failure, but it can also rely on the root cause analysis related to the design of the parts, the construction processes, the life cycle and so on. It can be used to identify various business scenarios and build appropriate prescriptive actions.

4.2. Example of Predictive Maintenance in rail systems

4.2.1. Smart Motors

It's a company that was found in 2009 after a 3-year collaborations between academic realities and the Transports Metropolitans de Barcelona, which goal is to provide new systems to generate useful information about the status of rolling stock in order to improve operation and breakdown management. This company provides software able to record service details and train subsystems information. Events and alarms are controlled and recent analysis has brought to an implementation that can identify complex pathologies and classifies alarms according to their severity. The software takes the historical data of the trains to forecast the moment in which maintenance must be provided.

4.2.2. Cyient

Founded in 1991, Cyient provides engineering, manufacturing, geospatial, network, and operations management services to global industry leaders. It delivers innovative solutions that add value to businesses through the deployment of robust processes and state-of-the-art technology. Its high quality products and services help clients leverage market opportunities and gain the competitive advantage. The company has first developed software for predictive maintenance in air industry, obtaining cost reduction for maintenance up to 8% and, successively, it’s going to use the same technologies in rail transport [16].

4.2.3. Hitachi

One of the older train company, founded in 1910, during the last centuries it has grown up becoming one of the most important company in rail industry. In recent years, with the development of IoT, it has developed predictive maintenance techniques able to reduce costs up to 40%. The system takes information from many sensors planted inside the vehicle, in the wheelsets and in the engine.

5. REALIZATION OF THE PROGRAM

Nowadays there are many tools able to forecast the trend of a component in order to evaluate the moment in which the maintenance must be carried out. Matlab provides many of these tools. The main problem is related to the type and amount of data: the Big Data provoke the so called Out-of-Memory Error because the program is not able to manage the huge amount of data. In addition, it's important to choose the right type of data.

Usually a program follows these steps [Figure 19]:

a. CHOOSE OF THE DATASET: it's possible to choose the dataset and to rearrange it, so that it will be possible to divide the data into Training and Learning Set;
b. CHOOSE OF THE LEARNING PROCESS: it's possible to choose several processes of learning according to the ones allowed by the tall array process, to choose the variables and to predict the goodness of it;
c. SOLUTION OF THE PROBLEM: once found the coefficients of the equation, it's possible to solve the problem equaling to the maximum allowable value;
d. EVALUATION OF THE RESULTS: it calculates the error with respect to the real solution and analyzes it.

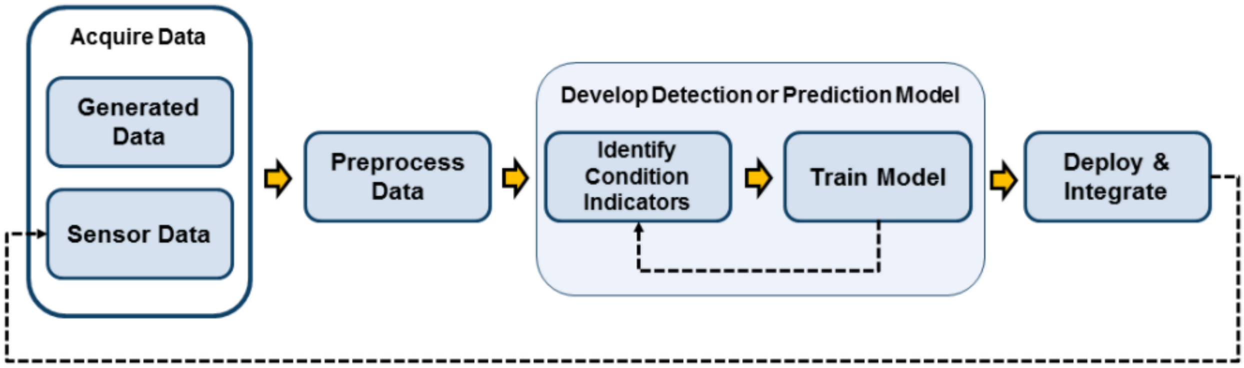


Figure 19: Program flow-chart ([7])

It uses several functions for machine learning:

- k-Means clustering
- Linear regression
- Generalized linear regression
- Logistic regression
- Discriminant analysis

5.1.1. K-means clustering

K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data. The goal of this algorithm is to find K groups composed by data with similarity and represented by a *centroid* and *labels* of the training data. This method allows analyzing directly the groups that have formed organically. The centroid represents a collection of feature values.

It uses iterative refinement to obtain the results. The data set is composed by the features of each data point. Each iteration includes two steps:

1. Each point is assigned to the nearest centroid, according to the squared Euclidean distance:

$$\underset{c_i}{\operatorname{argmin}}\, dist(c_i, x)^2$$

2. The centroids are recomputed, taking the mean of all the points in the cluster data set:

$$c_i = \frac{1}{S_i} \sum_{x_i \in S_i} x_i$$

The choosing of the number of groups K is not defined, but it's possible to have a good solution calculating the mean distance to the centroid as a function of K and plotting to find the elbow point [Figure 20]:

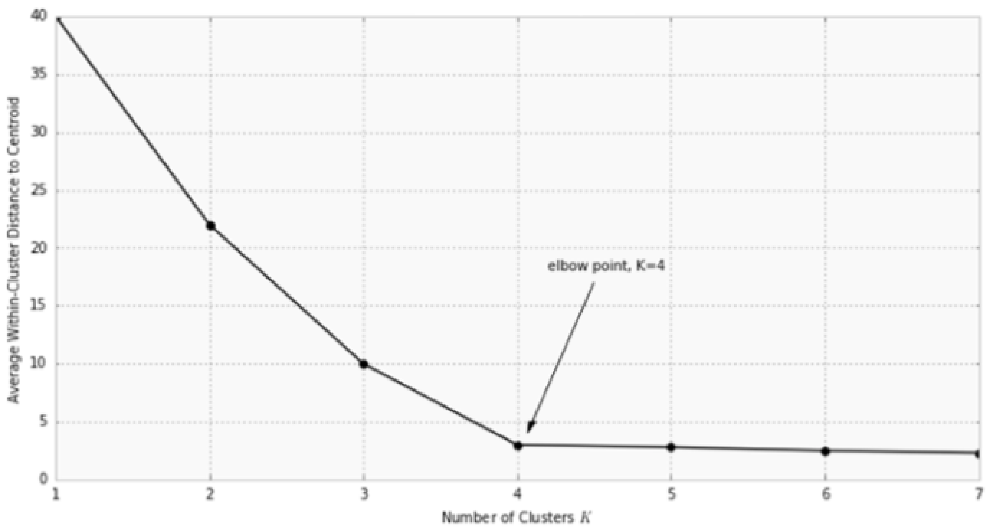


Figure 20: Elbow point for K-mean cluster

5.1.2. Linear Regression

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. Usually the existence of a relationship between these two variables must be verified even if one variable doesn't cause necessarily the other, otherwise this model will not be useful. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. A general equation is:

$$\boldsymbol{y = \beta x + \beta_0}$$

The most common method for fitting a regression line is the method of least-squares. This method calculates the best-fitting line for the observed data by minimizing the sum of the squares of the vertical deviations from each data point to the line.

5.1.3. Generalized Linear Regression

A general representation of a GLR model is

$$\boldsymbol{y = \beta_0 + \sum_i \beta_i \cdot x_i + \varepsilon}$$

This type of model is linear in the parameter β. The error is usually independently and identically distributed as $\varepsilon = N(0,\sigma^2)$.

5.1.4. Logistic regression

It's the method used to solve binary classification problems. In the core of the method there is the sigmoid function, a S-shaped function taking every value and mapping it in a value between 0 and 1 [Figure 21]:

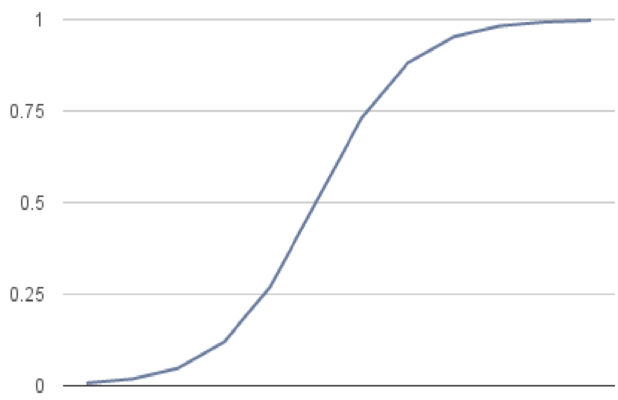


Figure 21: Sigmoid Function

$$\boldsymbol{y = \frac{1}{1 + e^x}}$$

Input value x is a linear combination using weights or coefficient values to predict a value y.

5.1.5. Discriminant analysis

The objective of discriminant analysis is to develop discriminant functions that are nothing but the linear combination of independent variables that will discriminate between the categories of the dependent variable in a perfect manner. It enables to examine whether significant differences exist among the groups, in terms of the predictor variables.

5.2. Estimating the Remaining Useful Life (RUL)

The program was developed to estimate the life status of the bearings in analysis in order to estimate their Remaining Useful Life (or RUL), the usage time left before the machine requires repair or replacement, which prediction is the main aim of the predictive maintenance techniques. RUL estimation models provide methods for training the model using historical data and using it for performing prediction of the remaining useful life.

The term *lifetime* or usage time here refers to the life of the machine defined in terms of whatever quantity you use to measure system life. Units of lifetime can be quantities such as the distance travelled, fuel consumed , repetition cycles performed, or time since the start of operation. Similarly *time evolution* can mean the evolution of a value with any such quantity.

Typically, RUL is estimated by developing a model that can perform the estimation based on the time evolution or statistical properties of condition indicator values, such as:

- A model that fits the time evolution of a condition indicator and predicts how long it will be before the condition indicator crosses some threshold value indicative of a fault condition;
- A model that compares the time evolution of a condition indicator to measured or simulated time series from systems that ran to failure. Such a model can compute the most likely time-to-failure of the current system.

There are three families of RUL estimation models, each one with different solutions [Figure 22]:

- SIMILARITY MODELS;
- DEGRADATION MODELS;
- SURVIVAL MODELS.

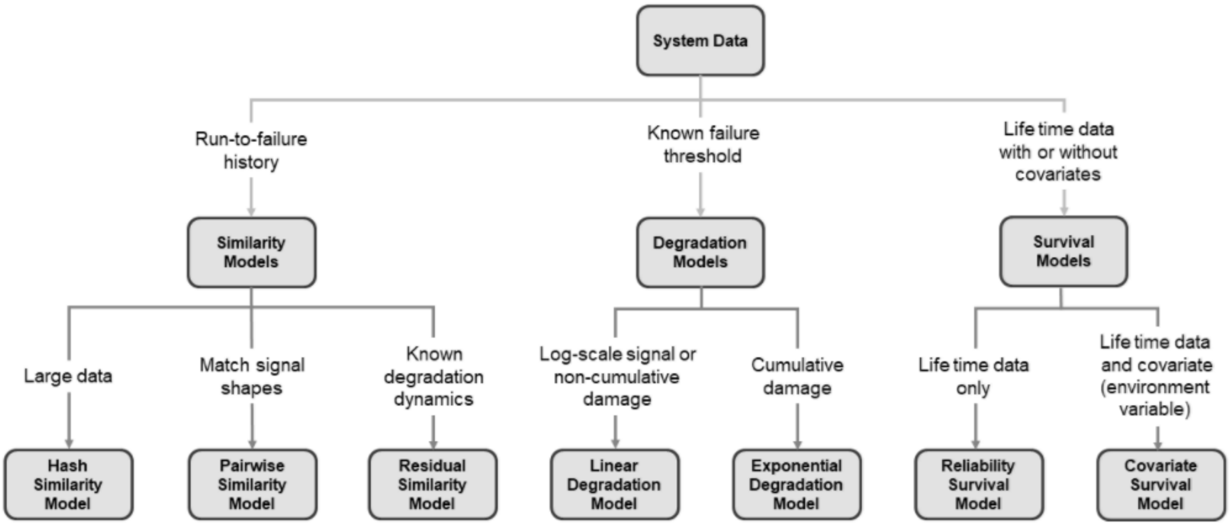


Figure 22. RUL Estimation Models

5.2.1. Similarity Models

Similarity models base the RUL prediction of a test machine on known behavior of similar machines from a historical database. Such models compare a trend in test data or condition-indicator values to the same information extracted from other, similar systems.

Similarity models are useful when:

- Having run-to-failure data from similar systems (components). Run-to-failure data is data that starts during healthy operation and ends when the machine is in a state close to failure or maintenance;
- The run-to-failure data shows similar degradation behaviors. That is, the data changes in some characteristic way as the system degrades.

These models use three estimators to compare the degradation history of a test data set and the degradation history of data sets in the ensemble:

- *Hashed-feature similarity model*: this model transforms historical degradation data from each member of the ensemble into fixed-size, condensed, information such as the mean, total power, maximum or minimum values, or other quantities.

The Hashed-feature Similarity Model is useful when having large amounts of degradation data, because it reduces the amount of data storage necessary for prediction. However, its accuracy depends on the accuracy of the hash function that the model uses;

- *Pairwise similarity model:* Pairwise similarity estimation determines RUL by finding the components whose historical degradation paths are most correlated to that of the test component. In other words, it computes the distance between different time series, where distance is defined as correlation, dynamic time warping or a custom metric that is provided. By taking into account the degradation profile as it changes over time, Pairwise Similarity Estimation can give better results than the Hash Similarity Model;

- *Residual similarity model*: Residual-based estimation fits prior data to model such as an ARMA model or a model that is linear or exponential in usage time. It then computes the residuals between data predicted from the ensemble models and the data from the test component. The Residual Similarity Model can be seen as a variation of the Pairwise Similarity Model, where the magnitude of the residuals is the distance metric. The Residual Similarity approach is useful when the knowledge of the system includes a form for the degradation model.

5.2.2. Degradation Models

Degradation models extrapolate past behavior to predict the future condition. This type of RUL calculation fits a linear or exponential model to degradation profile of a condition indicator, given the degradation profiles in the ensemble. It then uses the degradation profile of the test component to statistically compute the remaining time until the indicator reaches some prescribed threshold. These models are most useful when there is a known value of the condition indicator that indicates failure. The two available degradation model types are:

- *Linear degradation model:* describes the degradation behavior as a linear stochastic process with an offset term. Linear degradation models are useful when the system does not experience cumulative degradation;

- *Exponential degradation model:* describes the degradation behavior as an exponential stochastic process with an offset term. Exponential degradation models are useful when the test component experiences cumulative degradation.

5.2.3. Survival Models

Survival analysis is a statistical method used to model time-to-event data. It is useful when having not complete run-to-failure histories, but instead having:

- Only data about the life span of similar components. Given the historical information on failure times of a fleet of similar components, this model estimates the probability distribution of the failure times. The distribution is used to estimate the RUL of the test component;
- Both life spans and some other variable data (covariates) that correlates with the RUL. Covariates, also called environmental variables or explanatory variables, comprise information such as the component provider, regimes in which the component was used, or manufacturing batch. This model is a proportional hazard survival model which uses the life spans and covariates to compute the survival probability of a test component.

6. APPLICATION CASE 1

In 2012 FEMTO and IEEE Reliability Society organized the IEEE PHM 2012 Data Challenge in order to find a way to calculate the RUL of the bearings [17]. Challenge data sets were obtained by experiments carried out on a laboratory experimental table (PRONOSTIA) that enables accelerated degradation of the bearings in many conditions [Figure 23]; it's composed by three elements:

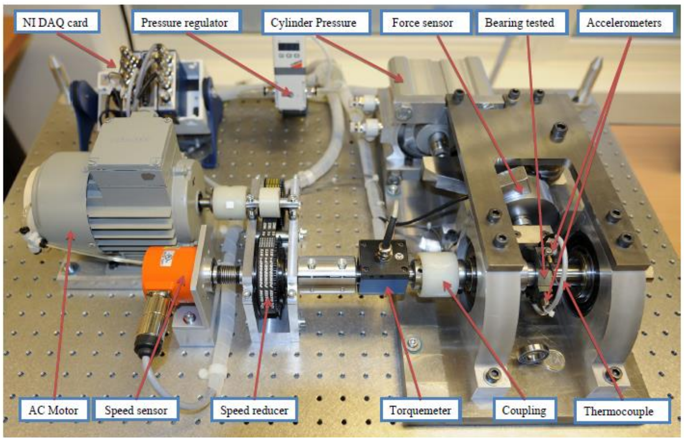


Figure 23: Pronostia table ([17])

- **<u>ROTATING PART:</u>** there is an asynchronous motor with a 250 W power, able to reach 2850 rpm. The bearing support shaft leads the bearing through its inner race. The shaft is composed of only one piece and is held by two pillow blocks. An interface allows setting the speed, selecting the direction of the rotation and to setting the monitoring parameters;
- **<u>LOADING PART:</u>** this part is composed by an aluminum plate supporting a pneumatic jack, a vertical axis, a sensor, the bearing and two pillow blocks with their bearing. It's the core of the system, allow a fast degradation of a test bearing applying a load with a maximum of 4000 N;
- **<u>MEASURAMENT PART:</u>** analyses are measured according to many factors: the radial force, the torque and the speed of rotation. Each bearing is characterized by two elements: vibrations and temperature. The vibration sensor consists of two miniature accelerometers positioned at 90 to each other, while the temperature sensor is a RTD platinum PT10 probe.

A set of 17 bearings was given, divided into Learning Set and Test Set [Table 6] for different conditions (the number of revolutions per minute (or rpm) and the loads), as seen in Table 5.

CONDITION	Revolutions per Minute	Load [N]
Condition 1	1800	4000
Condition 2	1650	4200
Condition 3	1500	5000

Table 5: Conditions of operability of the bearings

	Condition 1	Condition 2	Condition 3
Learning Set	Bearing 1_1 Bearing 1_2	Bearing 2_1 Bearing 2_2	Bearing 3_1 Bearing3_2
Test Set	Bearing 1_3 Bearing 1_4 Bearing 1_5 Bearing 1_6 Bearing 1_7	Bearing 2_3 Bearing 2_4 Bearing 2_5 Bearing 2_6 Bearing 2_7	Bearing 3_3

Table 6: Bearing Set

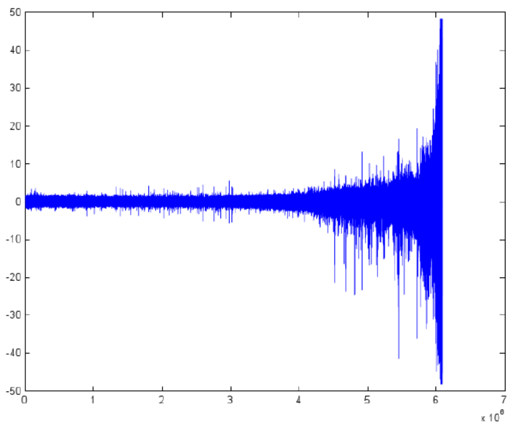


Figure 24: Vibration raw signal ([17])

The test runs until amplitude of 20g is reached in order to avoid propagation of damage to the whole table. A vibration raw signal is represented in Figure 24, but the failure behavior can assume different aspects.

The 6 fail-to-run datasets have a wide spread life duration, between 1 h and 7 h. RUL is defined as the moment in which the acceleration (horizontal or vertical) reaches the value of 20g.

The datasets are given in ASCII files and the information is taken with these modalities:

- VIBRATION (HORIZONTAL AND VERTICAL): a frequency of 25.6 kHz, recording 2560 samples (1/10 s) each 10 s;
- TEMPERATURE: a frequency of 10 Hz, recording 600 samples each minute.

BEARING	RUL [s]
Bearing1_3	5730
Bearing1_4	339
Bearing1_5	1610
Bearing1_6	1460
Bearing1_7	7570
Bearing2_3	7530
Bearing2_4	1390
Bearing2_5	3090
Bearing2_6	1290
Bearing2_7	580
Bearing3_3	820

Table 7: Bearing RUL

The ASCII data are presented as shown in Table 8:

COLUMN	1	2	3	4	5	6
ACCELERATION	Hour	Minute	Second	µ-second	Horizontal Acceleration	Vertical Acceleration
TEMPERATURE	Hour	Minute	Second	µ-second	Temperature	

Table 8: ASCII File

Also the result RULs of the Test Set were given, for each bearing of the Test Set [Table 7].These values are used to convert the predicted values into percent error of prediction. Let note RUL_i and $ActRUL_i$ respectively the RUL of the bearing estimated by the model and the actual RUL to be predicted: the percent error of prediction is defined as:

$$\%\boldsymbol{Er_i} = \mathbf{100} \times \frac{\boldsymbol{ActRUL_i - RUL_i}}{\boldsymbol{ActRUL_i}}$$

Overestimation and underestimation were not considered at the same way: to understand the goodness of the model, the factor A is calculated for each result: in case of negative values of Er a more severe deduction is actuated, while for the other cases an early removal was actuated. These two situations are defined by these accuracy functions:

$$\boldsymbol{A_i} = \begin{cases} \mathbf{exp}\left[-\mathbf{ln}(\mathbf{0.5}) \cdot \left({}^{\boldsymbol{Er_i}}/_{\mathbf{5}}\right)\right] & \boldsymbol{for\ Er_i \leq 0} \\ \mathbf{exp}\left[+\mathbf{ln}(\mathbf{0.5}) \cdot \left({}^{\boldsymbol{Er_i}}/_{\mathbf{20}}\right)\right] & \boldsymbol{for\ Er_i > 0} \end{cases}$$

The final score and the goodness of the model will be represented by the mean value of all the results Ai, which score is reported in Figure 25:

$$\mathbf{Sc} = \frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{A_i}$$

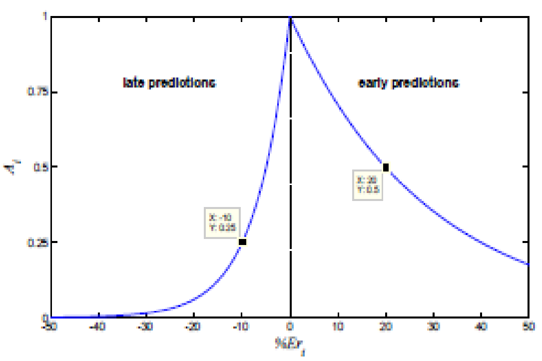


Figure 25: Error-Result Relation ([17])

The huge amount of data provoked the Out-of-Memory error: to solve this problem, Tall Arrays were used: these arrays allow blocking the process until it is not request to do it, so in this way it's possible to clean the data, to split them in Training, Test and Validation Data, reducing the data to be analyzed in the process. For each Learning and Test Set, about 1.5~2.0 million of data are given: all the data inside the Learning Sets are used for the Training phase, while, regarding each Test Set, the data were divided into Test and Validation Dataset, with a ratio of $^2/_3$ for the Test phase (an amount of about 1.0 million of data) and the remaining $^1/_3$ for the Validation phase (an amount of about 0.5 million of data).

To test the program, it was decided to analyze only the first condition specimens, the one in which the revolutions are 1800 rpm and the applied load equal to 4000 N. Furthermore only the time was taken as independent variable, divided into "Hour", "Minute", "Second" and "Microsecond". The Generalized Linear Regression was chosen and the result equation was the one reported below:

$$f[s] = \beta_0 + \beta_1 t_h + \beta_2 t_{min} + \beta_3 t_s + \beta_4 t_{\mu s} + \beta_5 t_h t_{min} + \beta_6 t_h t_s + \beta_7 t_h t_{\mu s} + \beta_8 t_h t_{\mu s} + \beta_9 t_{min} t_s + \beta_{10} t_{min} t_{\mu s} + \beta_{11} t_s t_{\mu s} + \beta_{12} Temp$$

$$f[s] = \vec{\boldsymbol{B}} \circ \vec{\boldsymbol{T}}$$

A number of 20 iterations were chosen for each bearing and both analyses on the vertical and horizontal acceleration were done. A RUL of 20g was chosen.

The provided Learning Sets were used to study the goodness of the code before using it for the Test Sets, of which nothing was known about the cause of breakages. In Table 9 the results of the tests are reported.

BEARING	ACTUAL RUL [s]	RUL [s]	ERROR [%]	A
Bearing 13	5730	5092	16	0.995
Bearing 14	339	878	-147	0.813
Bearing 15	1610	1607	2	0.999
Bearing 16	1460	1192	18	0.994
Bearing 17	7570	1139	85	0.971

Table 9: Results of the specimens

As it can be seen, the model well fits the actual RUL in most of the cases, in the other cases it seems that the model doesn't fit well: the reason of this could be found in the fact that, looking at the actual RULs, the average time of reaching the value is included in a range between 1500 and 5000 s, so for the other specimens a different structural composition could bring to those different values. The results were reported in an Error-Result graph, obtaining a trend similar to the one seen above and shown in Figure 26.

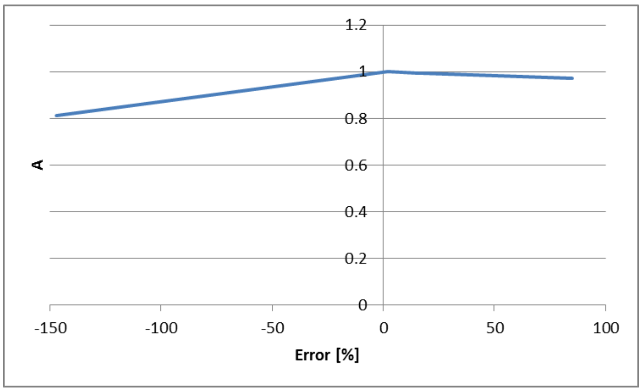


Figure 26: Error-A for Application Case 1

Also the other types of model shown in the previous chapter were used to train the model (specifically the Linear and the Exponential Regression, the K-Means Clustering and the Logistic Regression), but the obtained results were far from the expected ones reported below, in Table 9. In average the predicted RULs found in these ways were increased up to 100% with respect to the Generalized Linear Regression. This represents the fact that it's important to choose the right model to represent the behavior of data.

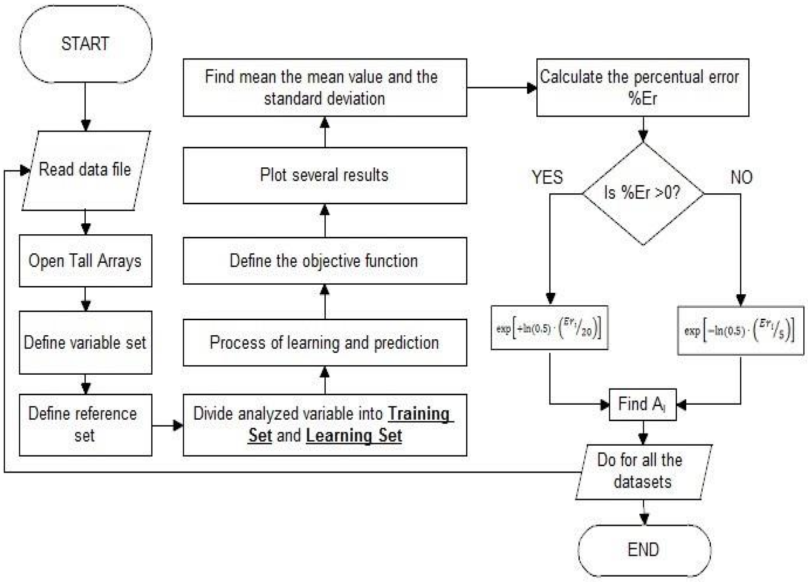


Figure 27: Application Case 1 Flow-Chart

In Appendix A the code of the program for the analysis of this type of bearings is reported, while in **Errore. L'origine riferimento non è stata trovata.** the flow chart of the program is reported.

7. BEARING CLASSIFICATION

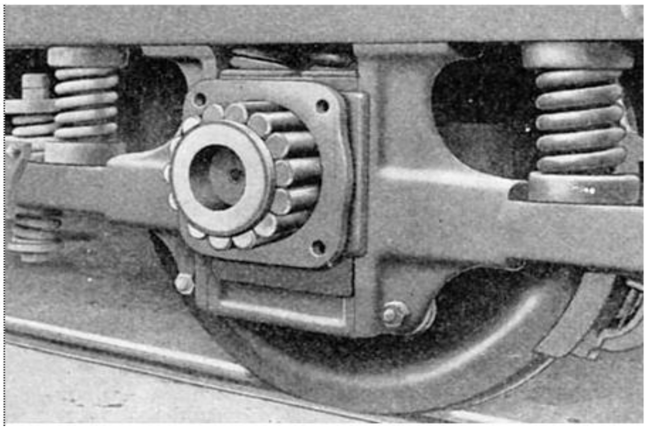


Figure 28: An axle bearing

Rolling stock axle bearings [Figure 28] are subject to radial impact loads caused by rail joints, switches and sometimes wheel flats, as well as to the static and dynamic radial loads of vehicle weight. They are also liable to receive axial loads generated by lateral movement as trains run on curved rails or in a snaking motion. All of these loads together form complex combinations that act on axle bearings. Axle bearings must therefore be designed on the basis of not only dimensional requirements of the axle journal and bearing box geometry, but also these complex load conditions. Additionally, as axle bearings play a critical role in the safety of railroad operation, they are periodically disassembled for inspection. For this reason, simple and dependable procedures for disassembly, inspection and re-assembly are important design factors as well. All types of radial roller bearings, including tapered roller bearings, spherical roller bearings and cylindrical roller bearings, have been used in rolling stock axles based on the particular merits of each type.

To improve operating efficiency, bearings must offer longer inspection intervals, simplified maintenance procedures and increased integration of bearing components and adjacent parts. To meet these needs, unitized bearings with advanced sealing devices have been introduced and are now widely used in modern rolling stock [18].

Axle bearings are divided into six typologies:

- RCT Bearings (Sealed-Clean Rotating End Cap Tapered Roller Bearings);
- RCC Bearings (Sealed-Clean Rotating End Cap Cylindrical Roller Bearings);
- Spherical roller bearings;
- Cylindrical roller bearings combined with ball bearings;
- Cylindrical roller bearings with ribs;
- Tapered roller bearings.

To ensure a good load capacity, usually all these types of bearings are manufactured in double-row configurations.

A. RCT Bearing

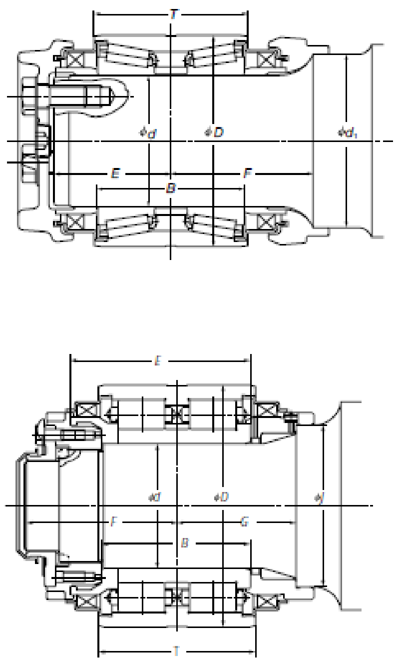
Preventing grease deterioration and leakage, as well as the intrusion of water and other foreign matter into the grease are vital for eliminating bearing trouble and lengthening maintenance intervals. Clearly, bearing seals offer the best way of achieving these objectives. RCT bearings are highly integrated with surrounding components and incorporate advanced sealing mechanisms. They offer outstanding performance, durability and ease of handling.

Figure 29: RCT Bearing ([18])

Generally, RCT bearings consist of an end cap, cap screws, a locking plate for fastening the end cap, a seal wear ring, a double-row tapered roller bearing and a backing ring. The latest variation has a backing ring that also serves as a seal wear ring [Figure 30].

Figure 30: RCT Bearing Component ([18])

Oil seals, mounted in seal cases, are press-fitted onto both ends of the outer ring and are in contact with the seal wear rings with a specified interference and pressure. The seals are spring-loaded contact seals. They are capable of preventing grease leakage and the intrusion of water and foreign matter into the bearing. For the assembly of bogies with axles supported by RCT bearings, saddle-type adapters are used instead of the bearing boxes commonly used for ordinary bearings. The use of such adapters can reduce the weight of the bogie and make assembly work easier.



B. RCC Bearing

This bearing is like the RTC, but the only difference is related to the end cap: in RTC it is tapered, while in RCC it's cylindrical [Figure 31].

Figure 31: RTC and RCC Bearing ([18])

C. Spherical Rolling Bearing

The spherical bearing box [Figure 32] is allowed to move freely in relation to the axle center because of the self-aligning property of the bearing. When a single spherical roller bearing is used, the use of a wing-type bearing box is recommended.

A spherical bearing by itself consists of an outer ring and an inner ring and a locking feature that makes the inner ring captive within the outer ring in the axial direction only. Spherical bearings are used in countless applications, wherever rotational motion must be allowed to change the alignment of its rotation axis.

Figure 32: Spherical Rolling Bearing ([18])

D. Cylindrical Rolling Bearing

Compared with tapered or spherical roller bearings, cylindrical roller bearings [Figure 33] have several strong advantages as journal bearings. These are:

›› The outer diameter is smaller and the weight is lower for the same load capacity;

›› Assembly and disassembly are easier facilitating maintenance and inspection;

›› The speed capability is higher because of the lower friction coefficient;

›› They allow the free setup of their axial clearance.

Figure 33: Cylindrical Bearing ([18])

Usually, the axial loads are borne by a single-row ball bearing such as a deep groove ball bearing or an angular contact ball bearing installed between the bearing box front cover and the axle end. With this type of bearing which is referred as the UIC type and has been standardized in Europe, axial loads are borne by ribs of the outer and inner rings and by the end of rollers. Compared with cylindrical roller bearings combined with ball bearings, this type offers simpler and compact housing construction owing to the absence of the ball bearing.

E. Tapered Roller Bearing

Tapered roller bearings can carry radial and axial loads simultaneously and therefore permit compact design of the bearing and its adjacent parts. This type of bearing, however, requires precise internal clearance adjustment in order to perform properly.


Figure 34: Tapered Roller Bearing ([18])

Tapered roller bearings are used either in sets of two, or in a double-row configuration in which there is one outer ring or one inner ring for the two rows of rollers. There are two types of duplex arrangements: back-to-back and face-to-face. For rolling stock axle applications where heavy moment loads are expected, the back-to-back arrangement, which provides a greater distance between load centers, is preferable. When the rollers are rolling under load, part of their load is transferred to the large rib of the inner ring. The rollers maintain sliding contact with and are guided by the rib. This results in the friction coefficient of these bearings being higher than that of cylindrical bearings. Recently, however, improvements in surface roughness and contact geometry have virtually eliminated the friction problems associated with tapered roller bearings for axles. This type of axle bearing can be designed with a sealed arrangement between the rear cover and the bearing box or, as described in the section on RCT bearings, they can have an internally sealed construction.

7.1. The bearings in railway systems

From the first use in railway systems, bearings have shown the capability to save energy and lubricant. The groups composed by two wheels and an axle rotating on bearings stayed more or less the same. But, because of their use, bearings represent the most critical part of a railway vehicle. The study of tribology and wearing has let know more about the interactions between surfaces in reciprocal movement.

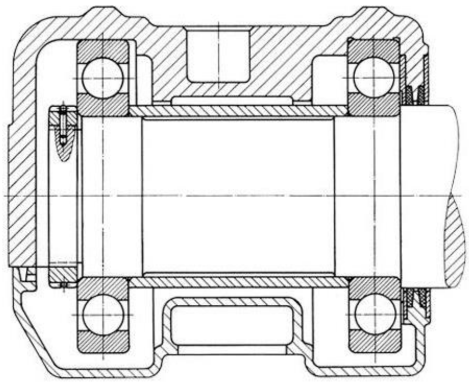


Figure 35: 1903 Bearing ([18])

There were some patent tools, but the first documented use is related to a three-axle passenger coach in 1903, equipped with two spherical radial bearings [Figure 35]: it allows having a reduction of 86% of the traction force needed to move a two-coach set of total weight 33 t (passing from 4.4 kN to 0.62 kN).Other tests were done in 1905 in Syracuse University of New York regarding the energy consumption, testing two trams: one was equipped with bushings, the other with roller bearings. The latter had an energy consumption of 3.10 kWh, while the other of 6.45 kWh, with a saving of 52%. The use of the roller bearings by the Syracuse Rapid Transit on its trams brought to the lack of wear on the bearings after 4 years and 400000 km run, with a save of 260 $ per year per vehicle.

In addition to energy savings, the use of bearing brings also to lubricant savings, both in terms of costs and in terms of environmental costs, reducing the production of oils and greases for the bearings, that must be emptied after several years of use and the products must be disposed of: a reduction of this phase has a good impact on the environment. At the beginning of railway transport, lubricated bearings were used, filled with 1.3 kg of lubricant, divided between 0.5 for the bushings and 0.8 for the fuel-tank. The controls were frequent and a loss of 0.2 kg of lubricant each 1000 km was measured. An important increment was given by the introduction of greased roller bearings, which needed often no re-lubrication. Initially 1.7 kg was used, but some studies analyzed that it was possible to use less product without problems. From the 50s less and less oil was used, till reaching the actual 0.7 kg. Another important step was reached by the pre-lubricated cylindrical roller bearings with incorporated protection (or CRU) that use only 0.2-0.3 kg of grease, reducing working temperatures and increasing in this way the exercise time.

Today trains can reach up to 300-350 km/h, modifying the geography of the Countries and connecting cities in a way that the distance can be count in terms of time and not of distance. Their use will be the future of passenger mobility in terms of speed and sustainability. Usually bearings are standardized, but they can be built according to specific needs; many of them are equipped with tools to monitor the performances. Same bearings are used for the wheel-sets of freight trains; above all bearing units are used with respect to the simple bearings.

A unit simplifies the installation phase and increments also the affordability and the safety because their maintenance is realized by the same companies that sell the product. These units have tapered or cylindrical rollers with specific advantages and used in all the kinds of vehicles. The choice of a unit rather than one other depends on many factors:

- Legislations and rules of the Country;
- Experience;
- Maintenance processes applied in the workshops;
- Type of vehicles.

In some countries practical and lab tests are required before every operation of modification: they are needed for conditioned (and then unconditioned) approval before put in circulation a new vehicle.

The protection system is included inside the bearing, sliding on the internal ring of it, allowing using shorter pivots in order to reduce the deformation on the axle and giving many advantages to the technical designer. The structure is made of polymer despite of steel and bronze, contributing heavily to safety and affordability.

Vehicles have many sensors installed on the bearing and used to measure operational parameters like speed, temperature and vibrations that give information to the control systems, like the braking system or the monitoring one.

7.2. Bearing Failure

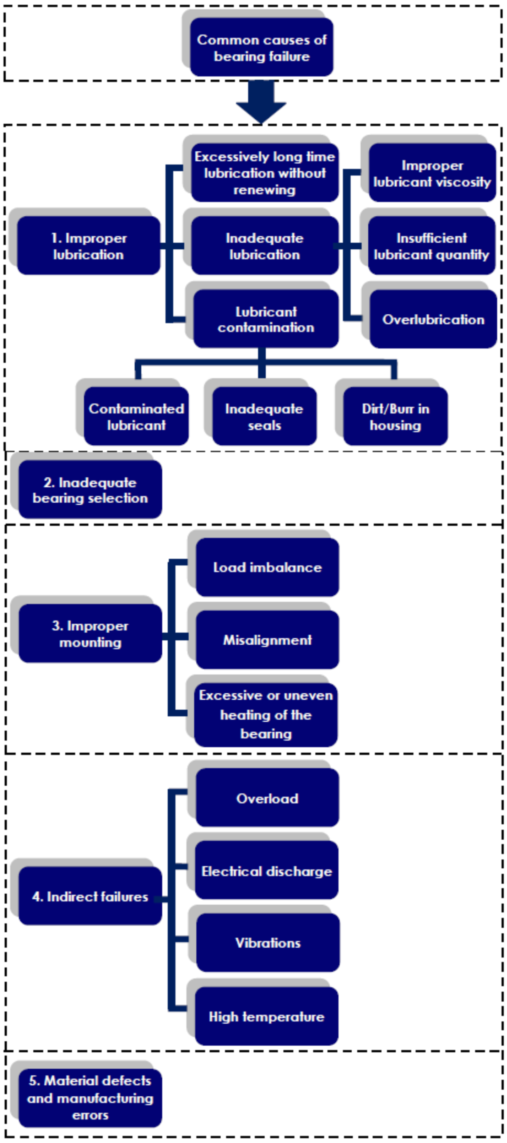


Figure 36: Failure scheme ([21])

As it can be read and understood in the previous chapter and in many work [19],[20],[21], lubrication plays an important role because the failure of one bearing doesn't stop only the vehicle but also the whole system. The amount of grease depends on many factors, above all the temperature and the speed. Damage or failure of a bearing is often the result of several mechanisms operating simultaneously. It is the complex combination of design, manufacture, assembly, operation and maintenance that often causes difficulty in establishing the primary cause of failure.

The evolution of tribological research during recent decades has led to a remarkable increase of new knowledge describing failure mechanisms. The data from this research field show that improper lubrication is the most commonly cited cause of bearing failure and accounts approximately 80% of breakdowns [Figure 37].

In the event of extensive damage to or catastrophic failure of the bearing, the evidence is likely to be lost and then it will be impossible to identify the primary cause of failure. In all cases, knowledge of the actual operating conditions of the assembly and the maintenance history is of the utmost importance. Typically, the causes of bearing failure can be classified in five groups and various sub-groups as shown in Figure 36.

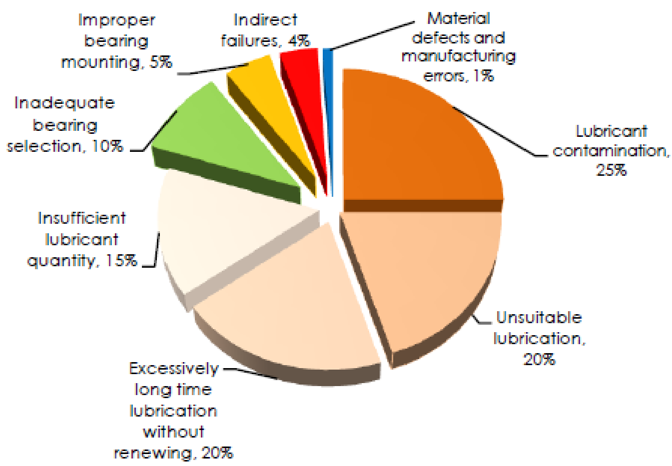


Figure 37: Percentage of failure ([21])

7.2.1. Improper lubrication

The selection of adequate bearing lubricants is based on decisions around whether to choose oil or grease, and determining what kind of additive is required. This decision depends on some factors, such as operating load, speed and temperature. Lubrication is a key factor that can make or break bearing service life. Some researches in the bearing industry have stated that improper lubrication can account around 80% of bearing failures. Failure can be the result of excessively long time lubrication without renewing, unsuitable lubrication and lubrication contamination.

Abrasive wear can be the result of inadequate lubrication. The surfaces become dull to a degree that varies according to the coarseness and nature of the abrasive particles. These particles gradually increase in number as material is worn away from the running surfaces and cage. Finally, the wear becomes an accelerating process that results in the failure of the bearing. Improper lubricant viscosity is one the major causes of bearing failure. As for lubricating oils, viscosity is one of the most important properties and determines oil lubricating efficiency. In order to form an adequate lubricant film between the rolling contact surfaces, the lubricant must retain a certain minimum viscosity while at operating temperature. The bearing life may be extended by increasing the operating viscosity. If viscosity is too low, the oil film will not form, and damage will occur to the bearing contact surfaces. On the contrary, when viscosity is too high, viscous resistance will also be great and temperature due to friction will be high. In either case, the asperities (microscopic machined high points) of the bearing component surfaces may contact each other, initially causing a frosted or smearing condition, followed by adhesion at the contact points. Contamination from water, chemicals, and particles is especially harmful to rolling bearings: when the lubricant is contaminated with wear solid particles, permanent micro cracks on the bearing raceway can be generated when these particles are over rolled. The appearance of these micro cracks can generate local stresses, which will lead to a reduced life of the rolling bearing. When steel, used for rolling bearing components, is in contact with moisture, oxidation of surfaces takes place. Subsequently the formation of corrosion pits occurs and finally flaking of the surface. A specific form of moisture corrosion can be observed in the contact areas between rolling elements and bearing rings, where the water content in the lubricant or the degraded lubricant reacts with the surfaces of the adjacent bearing elements.

Many bearings are simply brought to failure receiving insufficient lubricant quantity or no lubrication throughout their short life. Among others, the most common causes are simple neglect, incorrect lubrication intervals and failed lubrication system. For proper application, bearings must be monitored to ensure that lubricant intervals are not too frequent, causing over-lubrication, and not too infrequent, causing under-lubrication. Over-lubrication occurs when a rolling bearing is greased excessively or when too much oil is added to the housing. Excessive grease or oil quantity may cause internal friction between rolling parts, which generates excessive temperature that can create stress and deformity of the bearing.

7.2.2. Inadequate bearing selection

The selection of bearing made by the original equipment manufacturer is the correct selection for the application. Of course there are some exceptions, and there is always the possibility that a wrong bearing is used for the application.

Another common mistake is the use of a larger or stronger bearing, believing that this arrangement will increase radial load capacity. The larger or stronger bearing will not solve the true root cause problems. On the contrary, this new layout may create additional problems as, for instance that the selected bearing may have a speed lower than the nominal one and may not work properly in an environment with relatively high trust. In some cases, bearings require to be preloaded to facilitate rolling motion and to prevent roller skidding. Replacing the original bearing with a new one may actually lead to failures that are more rapid if it is not properly loaded. As a general rule, the replacement of a bearing has to be done with the same type of bearing, selected by the original equipment manufacturer.

Inadequate bearing selection represents about 10% of all premature bearing failures.

7.2.3. Improper mounting

Improper mounting accounts for about 5% of all premature bearing failures. Improper installation can lead to bearing failure through load imbalance, misalignment or improper load distributions.
A change of misalignment of 0.01/10 mm is enough to cause huge rise in vibration and temperature in the bearing. These sudden changes may introduce heavy wear in the ball or roller pockets where they run. These problems can be detected as a non-parallel running mark of the ball on the outer raceway and as means of extra wide ball or roller pathway on the inner raceway. Improper mounting can also lead to failure due to excessive or uneven heating of the bearing, when this is mounted on a shaft or housing.

7.2.4. Indirect failure

Indirect failures, such as unacceptable operating conditions, transport, storage and handling represent 4% of premature bearing failures. Among other indirect causes, the worst operating conditions are overloading, over-speeding, excessive vibrations, high temperature and electrical discharge. Furthermore, overloading can occur by excessive preloading or incorrect handling during mounting.
Vibration represents a huge problem as for bearing failures. In fact, vibration in a bearing while stationary can cause damage, called false brinelling. The damages can be identified as bright polished depressions or reddish stain common to fretting. These marks left by false brinelling will be equal to the distance between the rolling elements, just as it is in the cases of true brinelling, so these two conditions are often difficult to be distinguished. Electrical discharge is becoming a serious problem for bearings. During equipment operations, drive systems may produce a high level of static electricity that can be dissipated through the bearings to ground, causing pits or fluting to form on the bearing. Initially the surface damage takes the shape of shallow craters, which are closely positioned to one another and small in size. This happens even if the intensity of the current is comparatively low. Flutes will develop from the craters in time. Higher temperatures than those recommended by the manufacturer represent a risk factor for bearing life, no matter what type, quality or amount of lubricant is used. To highlight the importance of this point, consider the fact that a good quality mineral oil begins to oxidize at 71 °C. The same result will occur in greases where such oils are used as the lubricating agent.

Handling starts when a rolling bearing leaves the factory to the point when it is installed on a machine, and continues if the machine is to be transported after it is installed. Proper transportation and storage are essential to prevent damages from occurring before the machine is even placed in service.

7.2.5. Material defects and manufacturing errors

Rolling bearing failures due to manufacturing defects make up less than one percent of overall bearing failures around the world. This percentage is being continuously reduced by improvements in manufacturing process and material technology. Today bearing manufacturers use sophisticated instruments to detect surface and subsurface bearing material defects, eliminating in this way poor quality products during the manufacturing process.

In Appendix B several images of the above failures are reported, according to the type of failure that a bearing can suffer.

7.3. The Shock Pulse Method to measure bearing conditions

It's a method realized in 1969 and now used for monitoring a system. The SPM Institute offers a software that is represented in this chapter.

The bearing consists of two metal pieces contacting each other: when it happens, a shock or propagation wave develops and quickly propagates through the metal: this signal is in the ultrasonic frequency band, with a typical center of 36 kHz. As the signal expands from its point of origin, it is dissipated by carbon and other imperfections in the metal.

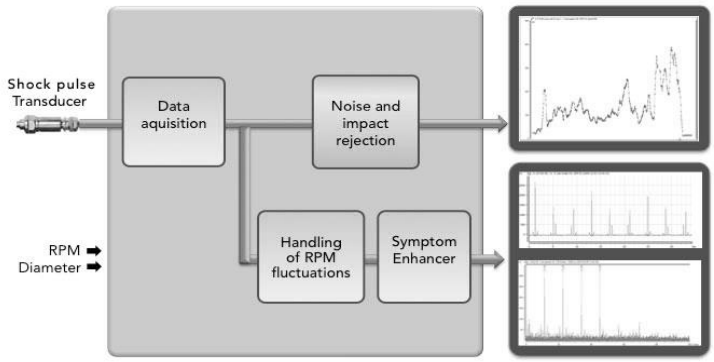


Figure 38: Shock Pulse Method

To measure the propagation, the Shock Pulse Method (or SPM) is used [Figure 38]: it is a patented technique for using signals from rotating rolling bearings as the basis for efficient condition monitoring of machines. The shock is caught by a transducer that converts the shock into an electric signal that will be processed to give a carpet and a peak value. The signal is also magnified because it has very low amplitude, by the use of an accelerometer design for this purpose.

Figure 39: Carpet and Peak Value

The shock pulse meter counts the rate of occurrence and varies the gain until two amplitude levels are determined [Figure 39]:

- *Carpet Value*: in absence of bearing damage, it's the background noise. When lubrication starts to degrade, there are more metal-to-metal contacts and this parameter increases;
- *Peak Value*: in case of bearing damage, the impacts have a periodicity depending on the place and entity of the defect, causing high amplitude waves.

The method requires more precise data on the bearing, because bearing geometry, as well as size and speed, affect the shock carpet and thus the analysis of oil film condition in undamaged bearings. The rpm is needed, plus a definition of the bearing type and size.

Shocks generated by damaged bearings will typically have an occurrence pattern matching the ball pass frequency over the rotating race. Shocks from damaged gears have different patterns, while random shocks from disturbance sources have none [Figure 40].

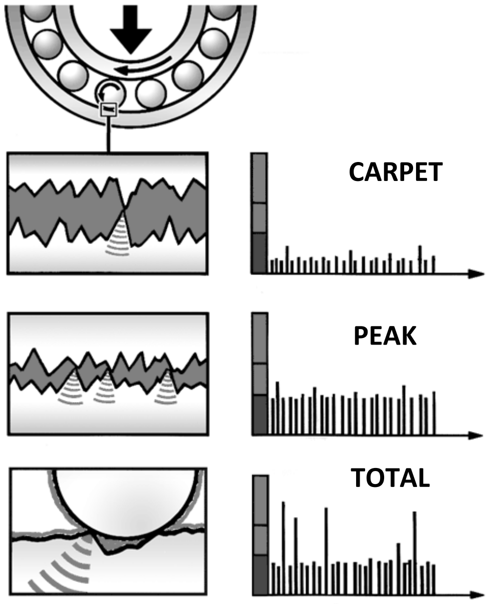


Figure 40: Shock analysis

7.4. Events related to bearing failure

The importance of bearing maintenance is increasing in these last years: many accidents have occurred all around the world related to the failure of bearings. Every accident results in discomforts related to the block of the circulation but often also in loss of human lives.

- ***USA FAILURES:*** in the period between the 2001 and 2010, about 4000 derailments occurred, among which the 7% is related to bearing failures, involving about 7 cars in each accident [22];

- ***NEW ZELAND FAILURES***: between 2007 and 2008 in New Zeland many bearing failures were registered on many freight trains, some bringing to derailment, without fatalities but with discomforts on the lines [23];

- ***FREIGHT TRAIN 50325:*** in 2009 a freight train, carrying 14 tank wagons filled with LPG, suffered the breakup of a bearing for fatigue near the city of Viareggio and the first coach blew up, with the death of 32 people. A lack of maintenance was revealed to be the cause of the missed replacement [24];

- ***CANADIAN PACIFIC RAILWAY TRAIN 119-01***: the train derailed while crossing the bridge over Wanapitei River because of a breakup of the bearings because of the high temperature; there were no injuries but the bridge collapsed [25];

- ***FREIGHT TRAIN 2AD1:*** in 2014 the train, carrying distillate fuel, suffered a bearing failure because of a lack of lubrication; no injuries but only damages to the track [26];

- ***CSX FREIGHT TRAIN:*** in 2015 a freight train, carrying toxic chemicals, derailed near Maryville in Tennessee, suffering the overheating of one of its bearings. 5000 people were evacuated;

8. APPLICATION CASE 2

Hypothetical data were provided data concerning gear train bearings [Figure 41], assumed obtained by two sensors installed inside the transmission box of the trains composing the fleet. The data collected are divided among Measurement Information (the data regarding the vehicle and the sensor and, in general, the physical part of the train) and Detected Values (the information detected by the sensors inside the transmission box):

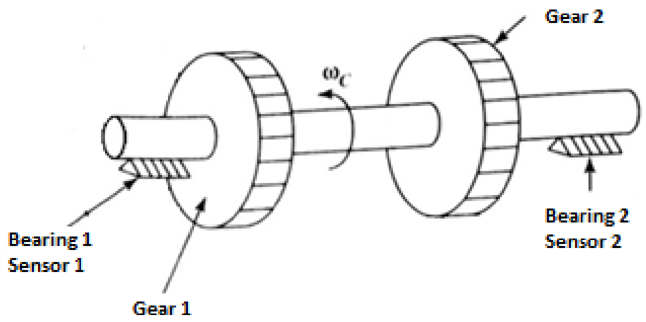


Figure 41: General Representation of a Gear Train

All the data are referred to one year, with an average time step of 2 days between each couple of detection. During this year of analysis, 14 cases of break were registered, preceded by an increment of the carpet value before the event. The availability of these events together with the other cases allows using similarity models, in particular Pairwise Similarity Models because similar behaviors are given but not a maximum value.

For the development of the analysis, only the data related to the sensor S1 of the gearbox where taken. In addition, only the data that are useful for the analysis and correlated to the failure were taken:

- Train;
- Bearing;
- Carpet Value;
- Peak Value;
- Speed;
- Heading.

The data have usually these envelops:

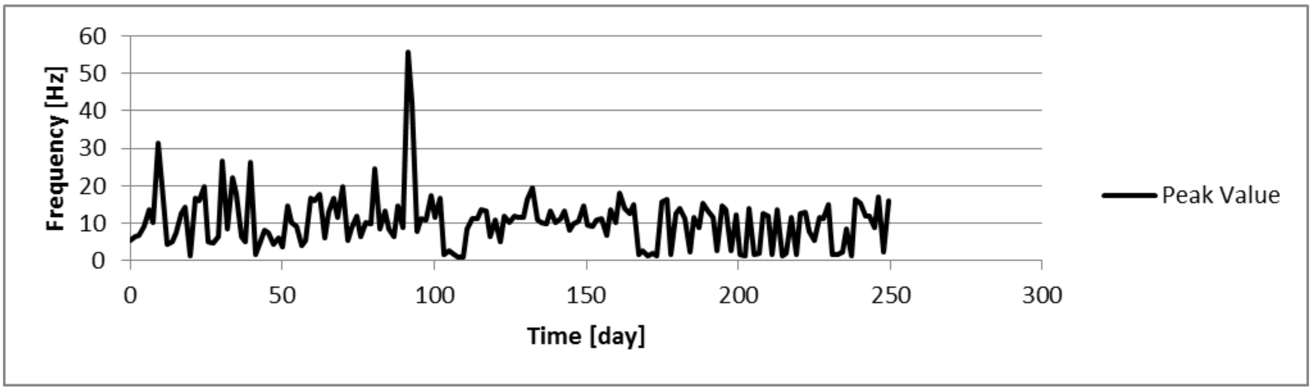


Figure 42: Peak Value

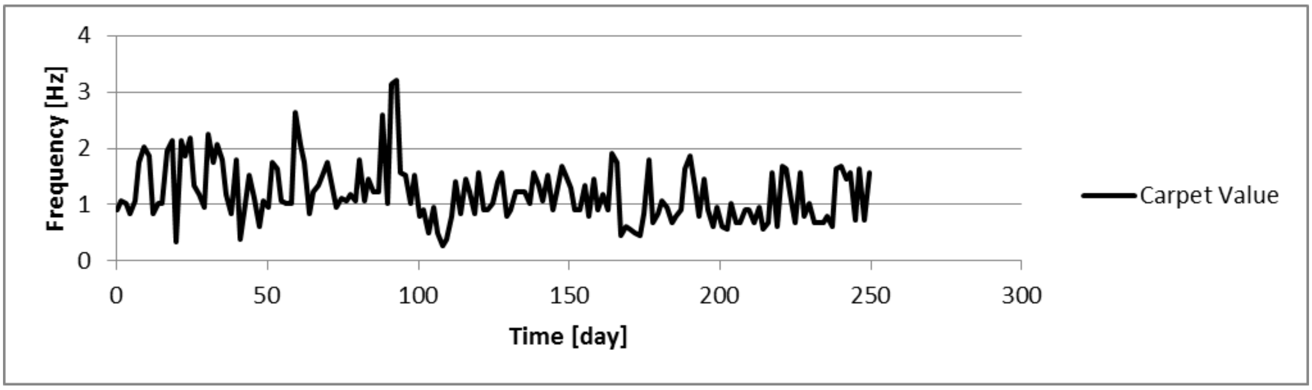


Figure 43: Carpet Value

As it can be seen in the representations of the peak and the carpet values, it's possible to detect an increment of the values before the failure event. This is caused by the reduction of the oil film and the increment of the direct impact of the metals with deterioration of the bearing: this causes the increment of the temperature together with fatigue of the metal and the formation of cracks, until the complete breakup of the bearing.

After the training of the model with all the **POSITIVE** (Yes Break) and **NEGATIVE** (No Break) series, it was decided at first to test the model to predict the breakup of the positive specimen: to do it, taking only data 60 days after the first day was necessary, obtaining the results in Table 10:

Case	Difference between Real RUL and Predicted RUL [day]	Predicted RUL [day]
a	0	78
b	12	50
c	10	66
d	8	55
e	-6	161
f	-8	215
g	2	83
h	-2	127

Table 10: Test on the known specimen

It's possible to see that the error has an expected value of 2 days, with a standard deviation of 1 week: usually in maintenance a 15-day analysis is chosen, in which the possible failure in that period is requested to be evaluated for the next 2 weeks in order to manage maintenance of the bearings.

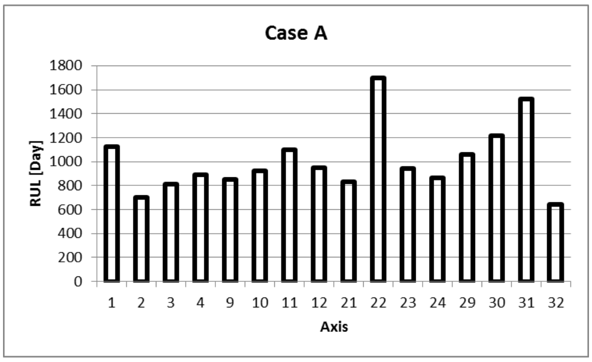
Then the analysis was spread to the other elements with given data. In a first attempt, the training was done with only the positive cases, obtaining the following results [Table 11] with respect to the last detection (after about 250 days from the first one):

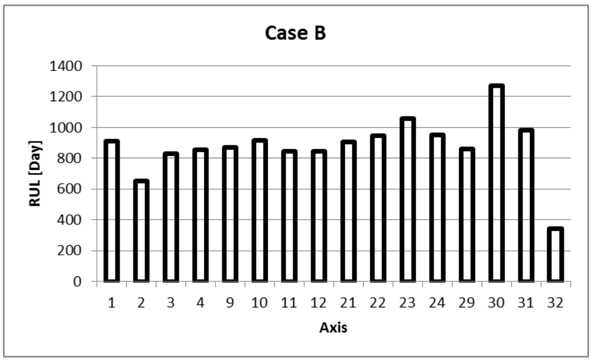
Case	Forecasting RULs [day]
I	216
II	163
III	141
IV	144
V	157
VI	125

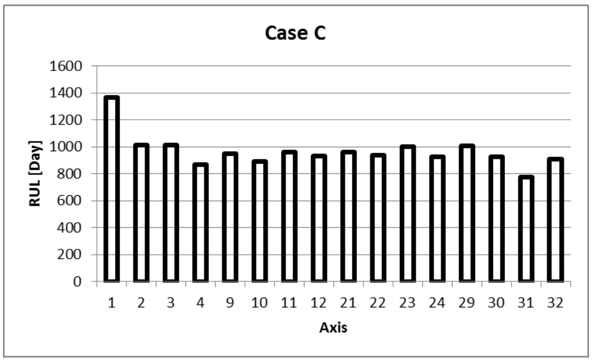
Table 11: RUL Estimation with only positive element testing

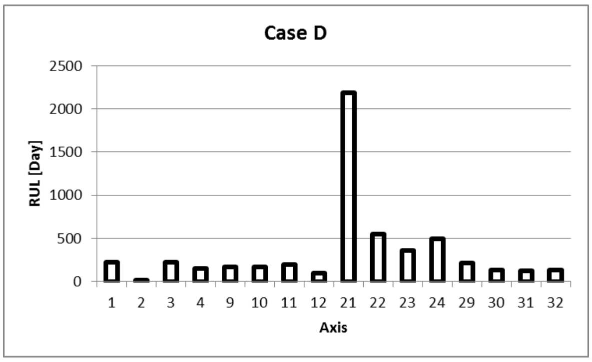
The results could seem to be right, but in the reality this is not true: the type of bearings realized for that type of train could last for years, while, with this training, they could break after a couple of years that is problematic because stop a train for maintenance presents a high cost and there are also clauses for not guaranteeing a certain operability of a train.

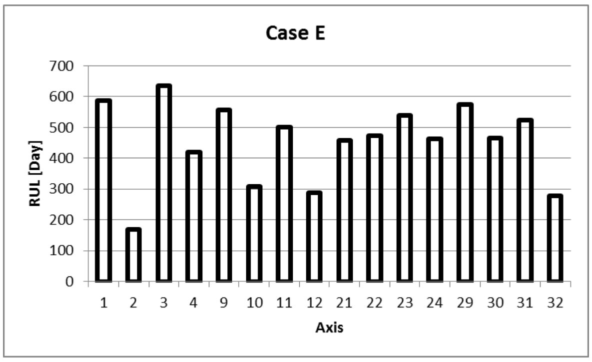
These results are caused by the wrong training of the model with only the positive series: in this way the machine is "induced" to bring the bearing to breakup in the fastest way possible, because it cannot "accept" that the item can have a longer life and not break in some months. So the model was trained with also negative series, obtaining some results of RULs reported below, with respect to the last detection:

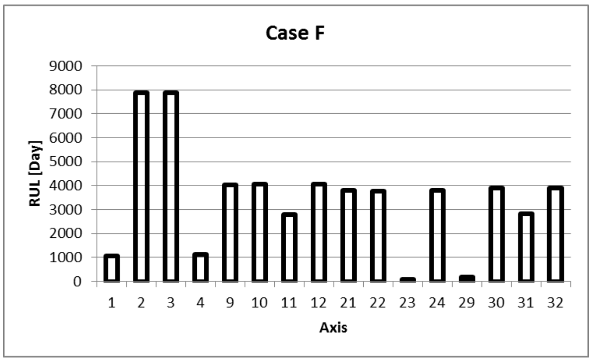












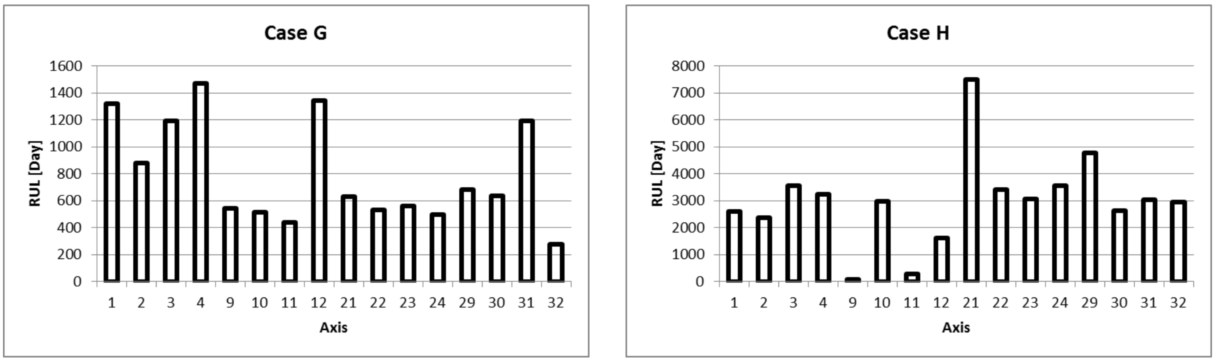


Figure 44: RUL Example

The RULs have a huge variety of values, from some months till many years, with an average RUL of 2 years. The results seem to be reliable with the expected life of the bearing, but they should be monitored each week because the situation could vary very fast in few weeks: the evolution of degradation of a broken bearing is shown in the pictures below, with an increment of 1 week:

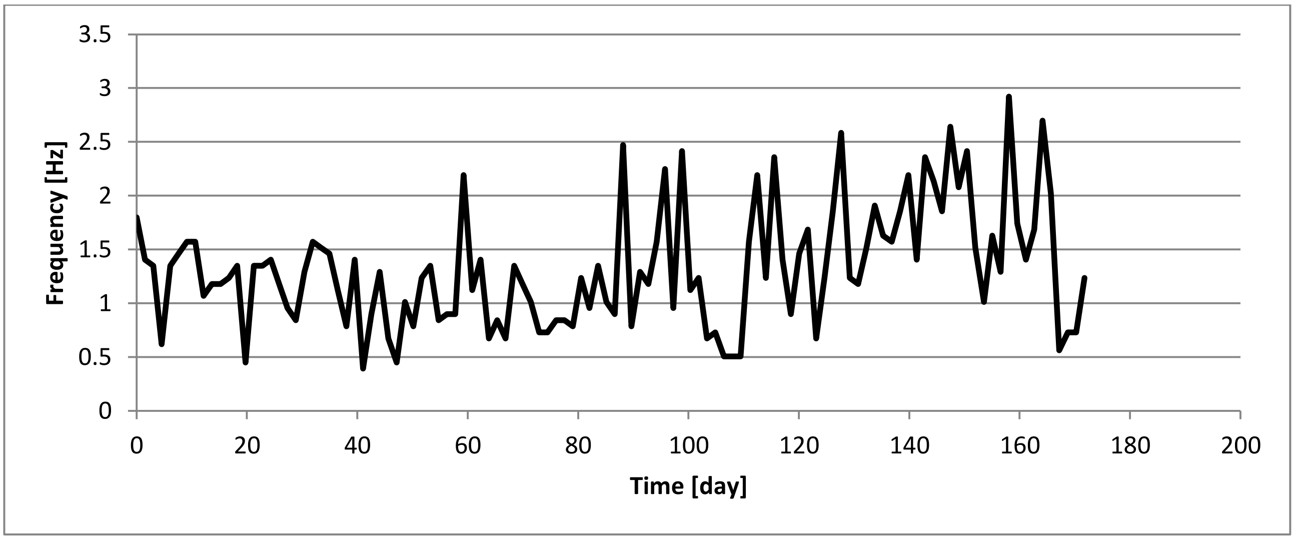


Figure 45: Breakup

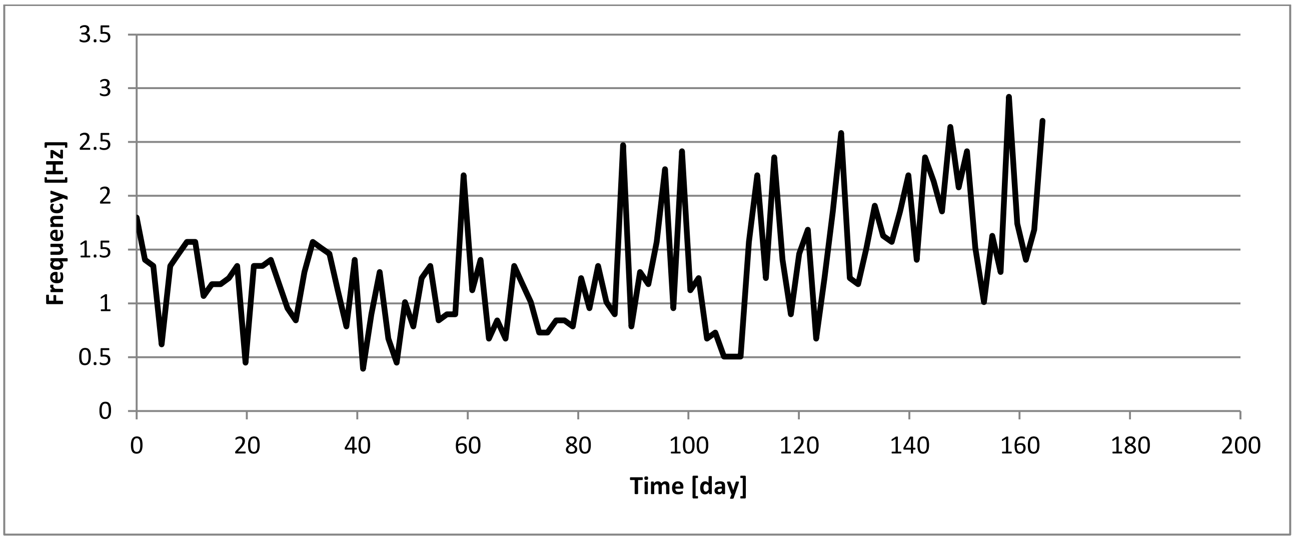


Figure 46: One weeks before breakup

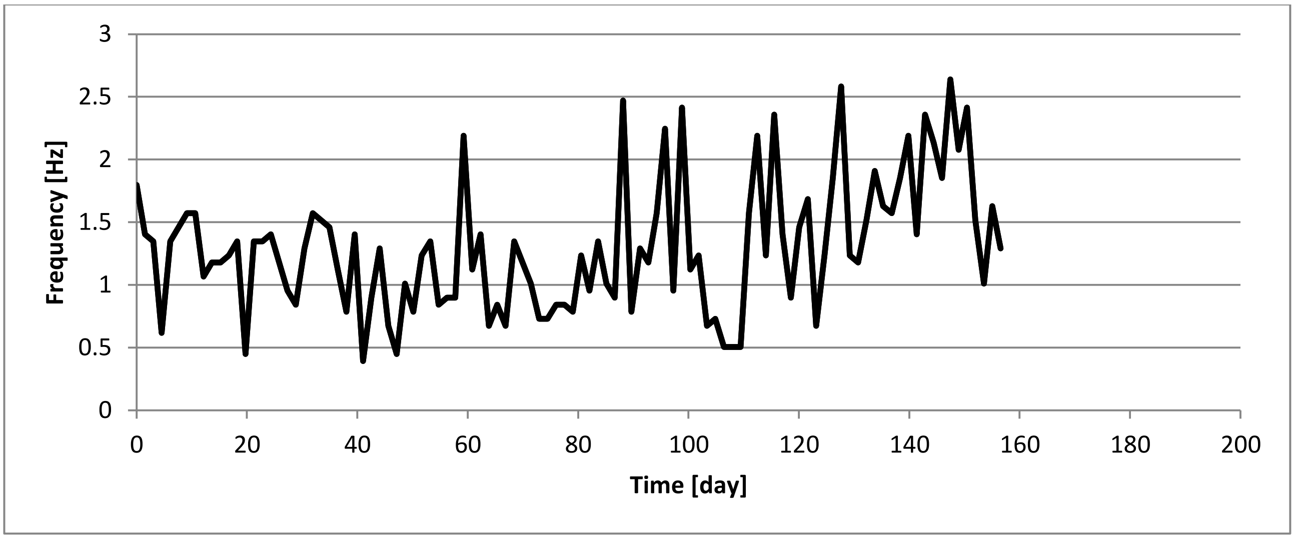


Figure 47: Two weeks before breakup

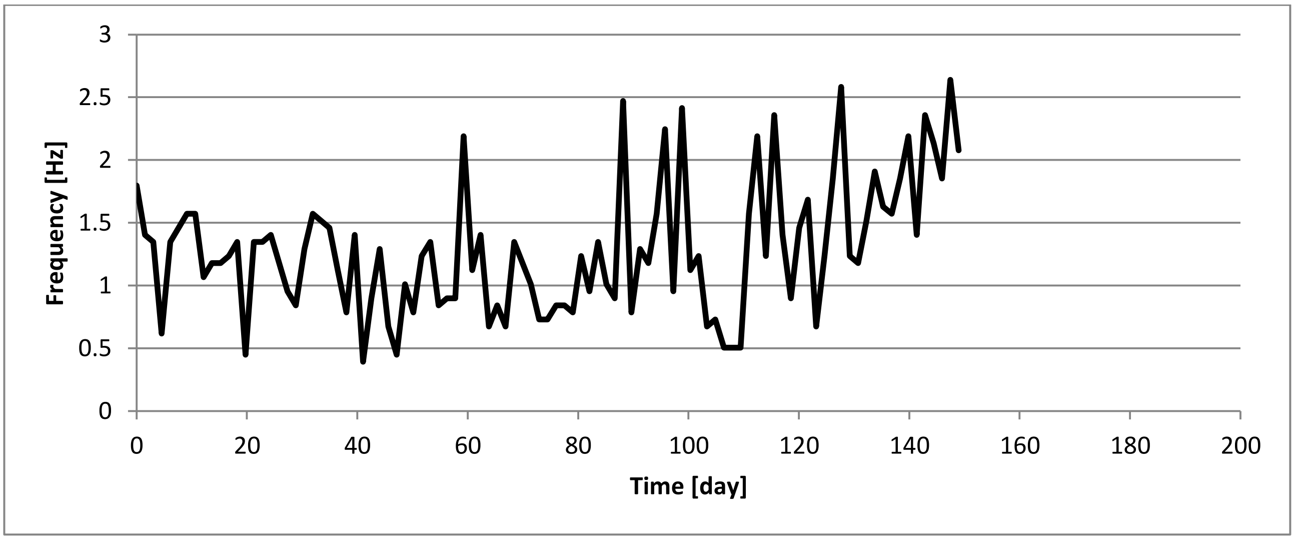


Figure 48: Three weeks before breakup

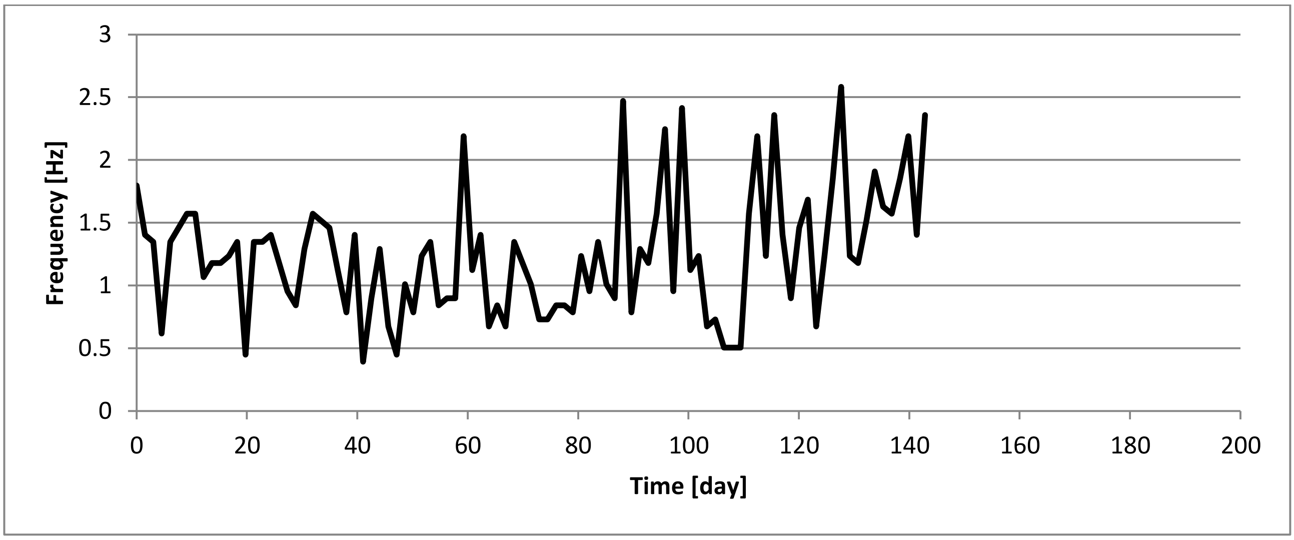


Figure 49: Four weeks before brakeup

The easy of the evolution of the degradation can be seen easily in the graphs above: the carpet value follows a normal trend till eight weeks before the breakup, swinging around a central value, but then it starts to increase in the following weeks, with a peek in the increment in the last three, increasing of the 200% before the breakup. This demonstrates the necessity to continuously monitoring the trend of the RUL week by week to have the right time to manage maintenance operations and take the right countermeasures. In Figure 50 the model gives the variation of the RUL during the weeks from the first day of detection, analyzing the bearing conditions week by week: as it is reasonable, after a first period in which the specimen seems to last many years because new, the RUL starts to decrease, reaching the same RUL registered at the end of detection.

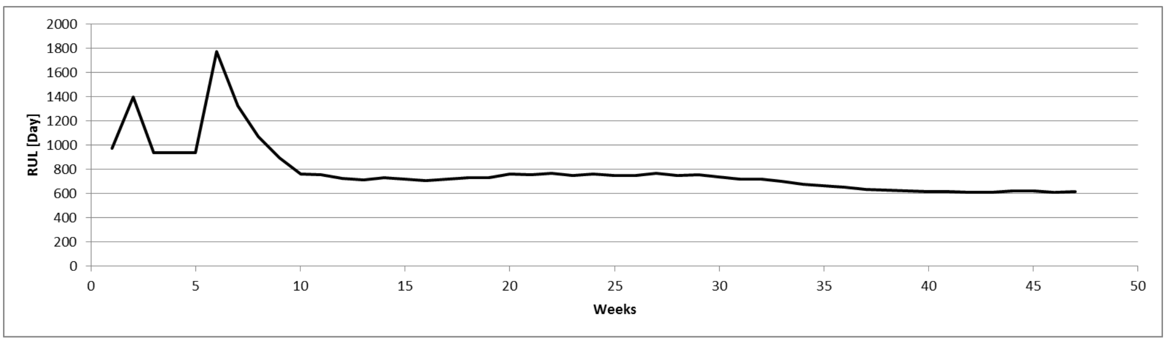


Figure 50: RUL Variation

9. OTHER SYSTEMS ANALYZABLE BY THE NEURAL NETWORKS

It is in the common interest of travelers and train operators try to avoid rolling stocks failures, for the passengers to experience comfortable trip and for the railway undertaking to save the costs of corrective maintenance activities. Maintenance is not only the one related to the mechanical components of the train, but also related to other systems inside the vehicle: the *converters* and the *refrigerator compressors.* These two components were analyzed in other thesis works by Filippo Rea in 2016 [27] and Lorenzo Roazzi in 2017 [28].

9.1. The Converters

From the evidences of the analysis, the converters are the third direct cause of warnings. Nevertheless, the large amount of warnings on the other many train components supplied by the converter shows the real impact of this typology of failures: in fact a converter provides energy to many subsystems of the train, like the door system, the plug system, the light system and so on [Figure 51].

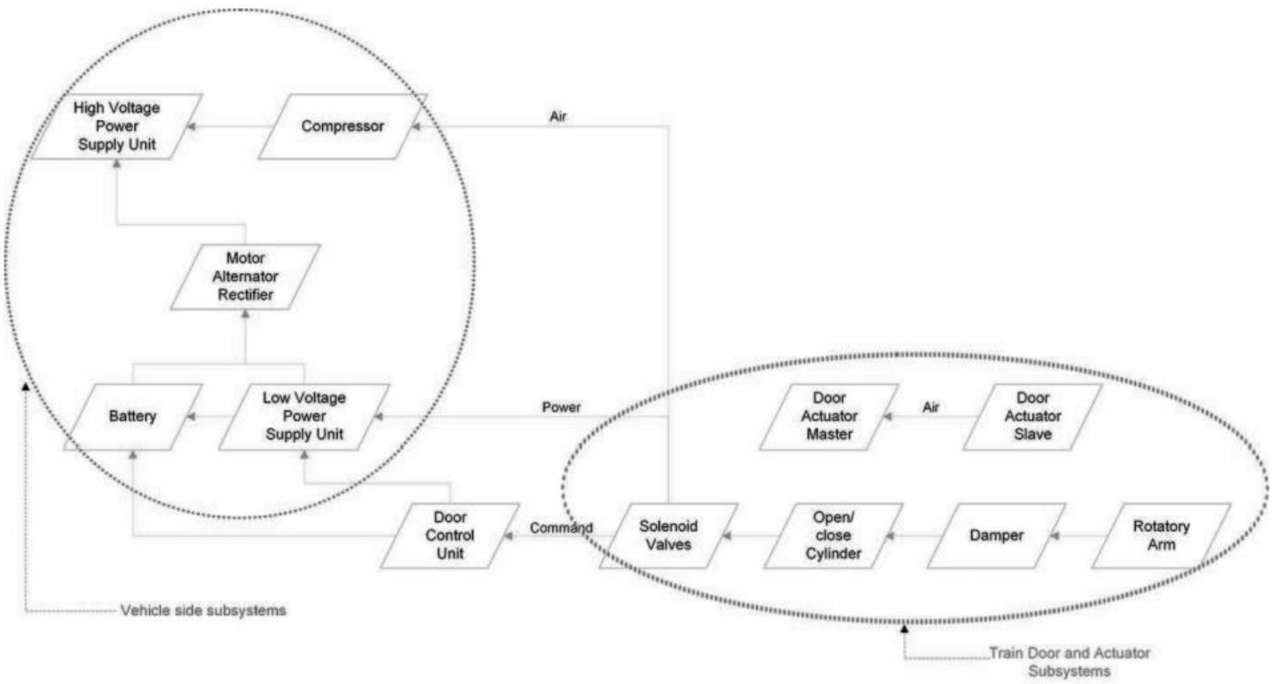


Figure 51: Train sub-systems [28]

Analysis were done related to the occurrence of failures of the converter system in the Universal Service coaches according to the FMCA analysis, affecting time by time different components. The corrective maintenance interventions are temporary solutions, the fact that the failures are very spread risks to make it difficult to face the problem with a reactive approach, and in the long term to extend the repair times. There are many small defects influencing the systems performance, but they are now immediately detected once happened. The maintenance operations found were estimated in around 1000€/coach, especially due to the time of non-operability of the vehicle.

Monitoring the right indicators is the first step, as it was reported in the work. For a bearing, the indicators shown in **Errore. L'origine riferimento non è stata trovata.** were used because they're the ones that better represent the bearing conditions. For the converter system, other types of indicators should be chosen:

- Current Intensity;
- Temperature;
- External Temperature;
- Number of operations of opening and closing door;
- Number of passengers;
- Number of stops;
- Voltage.

The failure analysis has brought to low reliability of the model: the use of self-learning simulation models can result a useful tool to prevent the failures. Thanks to on-board sensors, the acquisition and processing of a large amount of data will possible. The data contain the information on the component condition and behavioral range, in order to link the description of the effects of the failures with its real physical parameters. In this way the replacement and the maintenance activities of converter components will be optimized by checking the actual status of an item.

9.2. The Compressor Refrigerators

For long journey times, comfort plays a key role in the quality of service, and between the various aspects, to be the most critical is the HVAC (heating, ventilation and air conditioning) system [Figure 52]: the strong variability of the seasons and microclimates as well as temperature excursions between day and night, require a great adaptability and versatility of the system, fundamental characteristics for the maintenance of a pleasant temperature inside the carriage and therefore for passenger comfort, but that result in frequent failures of climate system, especially in the summer and winter peaks where the effort is maximum, resulting in detention of carriages and replacement of faulty elements that cause in turn lack of carriages available for the service and great amount of money for corrective measures. In particular, the compressors are system parts which have a greater anomaly in the frequency of failures because sometimes they have a life cycle drastically lower than that expected, even with repeated replacements in the same system in a short period of time, due to internal and external factors. Several types of failure were found:

- Mechanical Failures caused by Liquid Shots, occurred by the use of virtually incompressible liquid in the compressor chamber that provoked hammering and breaking of the aspiration valve;
- Mechanical Failures caused by Lubrication, occurred by the excessive dilution of the oil with result an improper lubrication;
- Mechanical Failures caused by Refrigerant Circuit Contamination Problems, caused by the presence in the refrigerant circuit oil of rust, corrosion, refrigerant decomposition, mud deposits, provoking friction and overheating;
- Electrical Failures, usually linked to other irregularities of the systems;
- Indeterminate Failures that cannot be determined by the maintainers because of many external factors (lack of time, freedom of operation…).

The specific cause of frequent failures of the compressors seems to be in the relationship between these and static converters, another key element of the carriage, equally dated, and despite some elements are modernized and changed, the converter remains the same, and with time it can't afford the increasing request of electric current.

There were proposed various possible solutions: to increase, with minimal cost, the converter period of transition in order to ensure full satisfaction of the starting of the compressors; with higher cost but better results, to replace the low-power converters with other powered, or to replace existing compressors with other little. More difficult to realize, but great for a future design, it is the installation of a continuous conduct that could carry heat and cool to the adjacent carriages in the event of compressor failure.

These solutions could represent good solutions for the old coaches analyzed, but cannot be the only ones in the future development: a monitoring of the temperature, of the energy expenditure, also of the vibrations, through few sensors rightly located, could be used for training a machine able to predict, with a certain warning, a possible future breaking of the system and avoid more costs than could be saved.

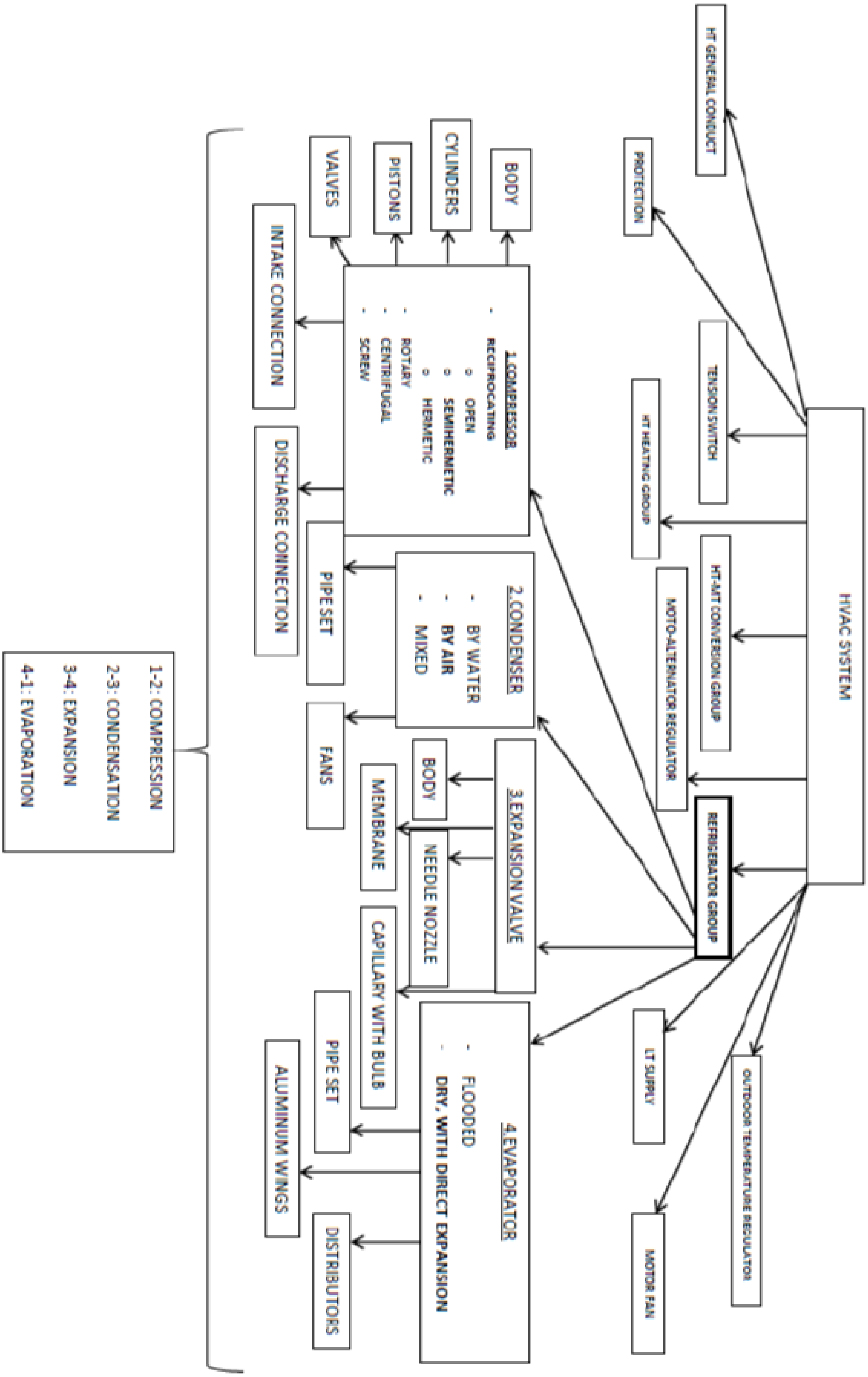


Figure 52: Conditioner System [27]

10. CONCLUSIONS

The Machine Learning is starting looking out the railway systems and, in general, in the other transports systems too. The actual traditional maintenance suffers many problems, related to its efficiency but, above all, to the costs: unnecessary replacements present costs that could be avoided, but, in the other hand, costs related to the corrective maintenance occurred for a not performed maintenance because the vehicle hasn’t reached the expiry date, taking into account also the costs for the not granted service and, unfortunately, the cost in terms of injuries and fatalities in case of accidents.

Also the importance of lubrication has been shown off: all the bearings have suffered breakup after a reduction of the oil film during its utilization, suffering meanwhile the formation of creeks and breaks on the surface, increasing during time till the failure is reached and the element breaks. This usually goes along with the increasing of the temperature due to the friction between the metal surfaces.

The use of Machine Learning in this work has shown that a different and better approach to the life state of a component is possible to measure its Remaining Useful Life: the most difficult step is the definition of the parameters to be analyzed in their historical trends, for the programmer, and the cost to install the specific sensors inside the vehicles, for the transport company. These are not two completely different views of the same problem, as it’s written, they are the SAME problem and only with a synergic work is possible to bring to a good resolution. It’s obvious that the initial cost will be high with long time for installing the mechanical components and that these new techniques will not be introduced inside the old rolling stock, but the future advantages will be a able to compensate them.

The results have shown that there is a good response, demonstrated by the application of the models to the cases, in which the failure event was occurred, with variations every week: this fact assumes, anyway, that a constant monitoring of the data must be granted because a fast variation of the conditions could occur in 7-15 days. In addition, it’s important, in the learning phase, to use not only the cases in which a failure is occurred, but also the other cases: like a human brain, a machine learned with only the first type of data could have a “pessimistic” view, forcing a component to end its life too early, with useless replacements. On the contrary, using only the negative results will give the model an “optimistic” view, giving a life longer than the real one and risking un-estimated failures. But learning with all the possibilities will give a “realistic” view of the component life.

The analysis is related only to the bearings, but, as it was reported in chapter 9, there are other subsystems inside a vehicle that can be interested by the use of Machine Learning and that cannot be represented by the traditional maintenance models.

But there are many cons related to the Neural Networks: first of all, every tool will not provide each passage, each weight used for the learning phase, an available equation of the model found by the machine, but it will only provide a black box, in which inputs are put and outputs are obtained. It’s possible to decide the type of model to use, but nothing else.
Another problem related to the programming phase is the yet reported difficulty in indicator choice: different subsystems have different indicators for their own and this provides a large availability of input but this is related to the set up cost of the sensors for a transport company: in this point programmers and companies must find a common ground, between the type of data to collect and the sensors to set up.

Bibliography

1. ANSF, *Linee guida Documentazione Iniziale relativa alla manutenzione (Dossier di Manutenzione) e Piano di Manutenzione dei veicoli,* 13th June 2014;
2. Bianco L., *La manutenzione dei veicoli ferroviari,* for "Congresso CIFI" in Verona, 12th February 2016;
3. Serpelloni G., *Manutenzione dei Veicoli Ferroviari*, for "Congresso CIFI" in Verona, 12th February 2016;
4. Facini M., *Sistemi di Gestione della Manutenzione dinamica (DMMS, E464)*, for "Convegno CIFI" in Verona, 5th December 2016;
5. Arrigo U., Di Foggia G., *Produzione, costi e performance delle principali reti ferroviarie dell'Unione Europea*, 2014;
6. AAVV, *Short-term rolling stock rostering and maintenance scheduling optimization,* Report RT-DIA-209-2014, February 2014;
7. Kim P., *MATLAB Deep Learning*, Seoul, Apress, 2017;
8. AAVV, *Deep learning applications and challenges in big data analytics,* in *"Journal of Big Data"*, Springer, 24th February 2015;
9. Bhuiyan Z.A., Gheisari M., Wang G., *A Survey on Deep Learning in Big Data,* July 2017;
10. Banjevic D., Jardine A.K.S., Lin D., *A review on machinery diagnostics and prognostics implementing condition-based maintenance*, in *"Mechanical Systems and Signal Processing"*, Vol 20, pp. 1483-1510, 3rd November 2005;
11. Adriani P.G., Paoli M., *La Diagnostica e la Manutenzione Predittiva*, for "Congresso Italiano di Manutenzione" in Verona, 22nd October 2008;
12. Lugarà A., *The railway predictive maintenance and the enabling role of the "Internet of Things",* in "Ingegneria Ferroviaria", n° 5, pp. 434-463, May 2018;
13. Forafò A., *Strumenti per un'analisi predittiva ai fini di una manutenzione efficace*, for "Manutenzione delle Macchine" in Parma, May 2014;
14. Malavasi G., Ricci S., *Simulation of stochastic elements in railway systems using self-learning processes*, in *"European Journal of Operational Research"*, n° 131, pp. 262-272, 2001;
15. Subrahmanyam M., Sujatha C., *Using neural networks for the diagnosis of localized defects in ball bearings,* in "*Tribology Internationl"*, Vol 30, n° 10, pp. 739-752, 1997;

16. CYENT, *Engineering Solutions for Predictive Maintenance in Rail*, https://www.cyient.com/fileadmin/cyient.com/media/downloads/150224-predictive-maintenance-in-rail-EN.pdf, 2015;
17. IEEE PHM, FEMTO, *IEEE PHM 2012 Prognostic Challenge: Outline, Experiments, Scoring of Results, Winners*, 2012;
18. NSK, *Bearings for railway rolling stock,* www.nskeurope.com, 2010;
19. Graney B.P., Starry K., *Rolling Element Bearing Analysis*, in *"Materials Evaluation"*, Vol 70, n° 1, pp. 78-85, January 2012;
20. AAVV, *Vibration Analysis of Rolling Elements Bearings Defects,* in *"Journal of Applied Research and Technology"*, Vol 12, pp. 384-395, June 2014;
21. Radu C., *The Most Common Causes of Bearing Failure and the Importance of Bearing Lubrication,* in *"RKB Technical Review",* February 2010;
22. Barkan C.P.L., Liu X., Rapik Saat M., *Analysis of Causes of Major Train Derailment and their Effect on Accident Rate,* in *"Transportation Research Record: Journal of the Transportation Research Board",* n° 2289, pp. 154-163, 2012;
23. Transport Accident Investigation Commission, *Rail Inquiry: Derailment caused by a wheel-bearing failure, Huntly, 19 October 2007, and, 11 subsequent wheel-bearing failures at various locations during the following 12 month period,* Report 07-114, November 2010;
24. Organismo investigativo del Ministero dei Trasporti, *Relazione di indagine sull'incidente ferroviario del 29 Giugno 2009 nella stazione di Viareggio*, 12th March 2012;
25. Transport Safety Board of Canada, *Canadian Pacific Railway Freight Train 119-01. Mile 112.70, Parry Sound Subdivision, Wanup, Ontario, 02 June 2013,* Railway Investigation Report R13T0122, 30th October 2014;
26. Australian Transport Safety Bureau, *Derailment of freight train 2AD1,* Report RO 2014-025, 7th December 2015;
27. Rea F.M., *Intercity Coaches HVAC Performance and Reliability Improvement through Engineering Failure Analysis of Refrigerator Compressors: Effects on Transport Service Availability*, Master's Degree Thesis in Transport Systems Engineering, 2016;
28. Roazzi L., *Performances and Reliability Improvement of InterCity Coaches DC/AC Converter based on FMECA Analysis*, Master's Degree Thesis in Transport Systems Engineering, 2017;

APPENDIX A: Matlab Code for Application Case 1

```
clear
clc

%Number of observations
noss=input('Define the number of observations = ');
%Number of relevament for each observations
nril=input('Define the number of relevament = ');
%Define the total number of rows
prod=noss*nril;

%Open the folder where there are the results
cd('C:\Users\mario\Dropbox\Tesi
Magistrale\Pronostia\Training\Learning');

%Define the RUL to test the code
Solmax=input('Define the RUL = ');
%Define the iterations
jmax=input('Define the number of iterations = ')

%Read the csv file with the data
Z=csvread('TEST13.csv',0,0,[0,0,prod-1,5]);
i1=1:prod;
%Create the arrays
u1=Z(:,1); u2=Z(:,2); u3=Z(:,3); u4=Z(:,4); u5=Z(:,5); u6=Z(:,6);
t1=table(u1,u2,u3,u4,u5); %Create the table
tt1=tall(t1); %Create the tall arrays
idx1=tt1.u5>0; %Define the constraints
tt1=tt1(idx1,:);
tt1=gather(tt1);

for j=1:jmax

    j %Counter
%Divide the data into training and test data
c = cvpartition(tt.u5,'HoldOut',1/3);
dataTrain=tt1(training(c),:); %Define the training data
dataTest=tt1(test(c),:);; %Define the test data
mdl1=fitlm(dataTrain,'interactions','ResponseVar','u5') %Realuze
the model
pred1=predict(mdl1,dataTest); %Predict the model
err1 = pred1 - dataTest.u5; %Evaluate the error

syms tm
K=mdl1.Coefficients.Estimate
%Define the angular coefficient
SU=K(2,1)*3600+K(3,1)*60+K(4,1)+K(5,1)*10^-
6+K(6,1)*3600*60+K(7,1)*3600 ...
+K(8,1)*3600*10^-6+K(9,1)*60+K(10,1)*60*10^-6+K(11,1)*10^-6;
%Define the equation
TE=K(1,1)+SU*tm==200;
```

```
%Solve the problem
Sol=solve(TE,tm);

%Define the coefficients
K1(j)=K(1,1);K2(j)=K(2,1);K3=K(3,1);K4=K(4,1);K5=K(5,1);K6=K(6,1);
K7(j)=K(7,1);K8(j)=K(8,1);K9=K(9,1);K10=K(10,1);K11=K(11,1);

end

Sol=gather(Sol); %Start the calculation of the tall arrays
Sol=vpa(simplify(Fol),4);

%Plot the result
plot(Sol),xlabel('Iterations'),ylabel('Seconds [s]'),...
    title('Test at n = jmax')

%Evaluate the mean and the standard deviation
m=mean(Sol)
SM=(Sol-m)^2;
SD=sqrt(sum(SM)/(length(SM)-1))

%Evaluate %Er
Er=(Solmax-mean)/Solmax

%Er
if Er>0
    A=exp(log(0.5)*(Er/20))
else
    A=exp(-log(0.5)*(Er/5))
end
```

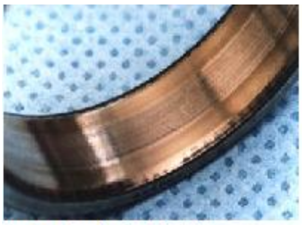
APPENDIX B: Matlab Code for Application Case 2

```
clear
clc

%Read each column of the data, converting the dates into numbers
TR=xlsread('NAMEFILE.xlsx','Sheetname','a2:a677624');
AS=xlsread('NAMEFILE.xlsx','Sheetname','c2:c677624');
[n,pi,valo]=xlsread('NAMEFILE.xlsx','Sheetname','d2:d677624');
TI=datenum(datetime(pi,'inputformat','dd/MM/yyyy'));
RE=xlsread('NAMEFILE.xlsx','Sheetname','e2:h677624');

clear n valo pi %Clear useless variables

%Divisions of the data according to the train and the axis
ji=50; %Trains
ki=16; %Bearings

for j=1:ji
for k=1:ki
data=[TR AS TI RE];

data=data(data(:,1)==j,:); %Filter for the train
data=data(data(:,2)==k,:); %Filter for the axis
po=data(:,4:7);
t=cumsum([0;diff(data(:,3))]);
sim=[t po];
[~,idx] = unique(sim(:,1));
sim = sim(idx,:);
string=['Sit' num2str(j) num2str(k)];
v=genvarname(string);
eval([v '=sim;']);
end
end

clear  idx j  k  po sim string v t %Clear useless variables

%Put in cell format the results
for j=1:ji
  for k=1:ki
        sit=['Sit' num2str(j) num2str(k)];
    col(j)={eval(sit) };
  end
end
cell=col';

clear AS col data j k RE sit TI TR %Clear useless variables
```

```
%Learning Phase
mdl = pairwiseSimilarityModel;
fit(mdl,cell)


%RUL Estimation
for j=1:ji
cas=['b' num2str(j+3) ':' 'q' num2str(j+3) ];

for k=1:ki
sit=eval(['Sit' num2str(j) num2str(k)]);
d=sit(:,1);
s=mean(d);

y=round(s*(predictRUL(mdl,sit)));

q(k)=[y];

xlswrite('RESULT.xlsx',q,'Risultati2',cas)
end
end
```

APPENDIX C: Typologies of bearing failure

IMPROPER LUBRICATION



Wear for lack of lubricant

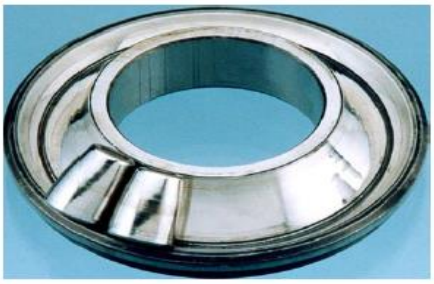
Wear for contact among asperities

Wear for low oil viscosity

Failure for water contamination

INADEQUATE BEARING SELECTION

IMPROPER MOUNTING



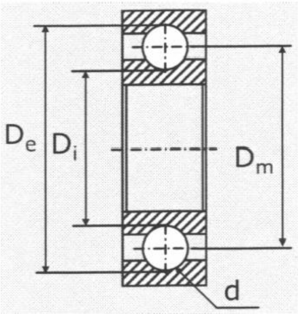
INDIRECT FAILURES

Overload deformation

Current leakage deformation

False brinelling

APPENDIX D: FEMTO Testing Bearing Characteristics



D_e = Outer Diameter = 29.1 mm

D_i = Inner Diameter = 22.1 mm

D_m = Mean Diameter = 25.6 mm

d = Rolling Element Diameter = 3.5 mm

Z = Number of Rolling Elements = 13

APPENDIX E: FEMTO Sensor Characteristics

Accelerometer Sensor

SPECIFICATION	VALUE	UNITS
PHYSICAL		
WEIGHT	2.5	grams
SIZE, HEX x HEIGHT	.281 (9/32) x .33	inches
MOUNTING PROVISION, 3035B	5-40 integral stud	
MOUNTING PROVISION, 3035BG	flat surface for adhesive mount	
CONNECTOR, RADIALLY MOUNTED	5-44 coaxial	
MATERIAL, HOUSING AND CONNECTOR	300 Series Stainless Steel	
PERFORMANCE		
FREQUENCY RANGE, ± 5%	0.5 to 10k	Hz
RESONANT FREQUENCY, NOM.	45	kHz
EQUIVALENT ELECTRICAL NOISE FLOOR	.007	g rms
LINEARITY [2]	± 1%	% F.S.
TRANSVERSE SENSITIVITY, MAX.	5	%
STRAIN SENSITIVITY	.002	g/με @ 250με
ENVIRONMENTAL		
MAXIMUM VIBRATION/SHOCK	600/3000	± g pk
TEMPERATURE RANGE, 3035B/BG, 3035B2/B2G	-60 to +225	°F
3035B1/B1G, 3035B3/B3G	-60 to +250	°F
SEAL, HERMETIC	Glass-to-metal and TIG welded	
COEFFIEICNT OF THERMAL SENSITIVITY	.04	%/°F
ELECTRICAL		
SUPPLY CURRENT [3]	2 to 20	mA
SUPPLY COMPLIANCE VOLTAGE RANGE	+18 to +30	Volts
OUTPUT IMPEDANCE, TYP.	100	Ω
BIAS VOLTAGE	+11 to +13	Vdc
DISCHARGE TIME CONSTANT	0.5 to 1.2	seconds
OUTPUT SIGNAL POLARITY FOR ACCELERATION TOWARD TOP	Positive	
CASE GROUNDING	Case is grounded to electrical power ground	

Temperature Sensor

Nominal Resistance = 100 Ω

Usage Range = -200 to +600 řC

Diameter = 2.8 mm

Length = 25 mm

	Tolérances									
Temp	Classe B		Classe A		1/3 DIN		1/5 DIN		1/10 DIN	
(°C)	± °C	± Ohms	± °C	± Ohms	± °C	± Ohms	± °C	± Ohms	± °C	± Ohms
-200	1,30	0,56	0,55	0,24	0,44	0,19	0,26	0,11	0,13	0,06
-100	0,80	0,32	0,35	0,14	0,27	0,11	0,16	0,06	0,08	0,03
0	0,30	0,12	0,15	0,06	0,10	0,04	0,06	0,02	0,03	0,01
100	0,80	0,30	0,35	0,13	0,27	0,10	0,16	0,05	0,08	0,03
200	1,30	0,48	0,55	0,20	0,44	0,16	0,26	0,10	0,13	0,05
300	1,80	0,64	0,75	0,27	0,60	0,21	0,36	0,13	0,18	0,06
400	2,30	0,79	0,95	0,33	0,77	0,26	0,46	0,16	0,23	0,08
500	2,80	0,93	1,15	0,38	0,94	0,31	0,56	0,19	0,28	0,09
600	3,30	1,06	1,35	0,43	1,10	0,35	0,66	0,21	0,33	0,10
650	3,60	1,13	1,45	0,46	1,20	0,38	0,72	0,23	0,36	0,11
700	3,80	1,17								
800	4,30	1,28								
850	4,60	1,34