POLITECNICO DI TORINO

Department of
Control and Computer Engineering (DAUIN)
Master Degree in Mechatronic Engineering

Development of a predictive maintenance system for electrodes dressing in welding guns

Giuseppe De Bernardo
Matr.-Nr.: s265814

Mentors: Prof. Alessandro Rizzo
Prof. Maurizio Schenone

Supervisor: Ing. Dario Cambiano - ISI-WELDING

Academic Year 2020/2021

Ho una serie di nomi incisi sul cuore,
GIUSEPPE DE BERNARDO è il nome del nonno,
ma chiamatemi anche:
AMALIA,
ROSARIA,
VITTORIO
...

Contents

List of Figures

List of Tables

1 Introduction

In these years marked by global pandemic and economic crisis, humanity has been challenged to radically change their habits and to find modern solutions that fit the context.
The market was also heavily affected: companies are unwilling to buy new goods, rather they are intent to maximize the useful life of the industrial machines. For this reason, the business is settling to the production of services act to extend goods maximum life.
The presented work of thesis arises from the necessity of the ISI-Welding Company to obtain a predictive maintenance system for their welding guns. Hand in hand with the use of data science, predictive maintenance is one of most interesting topics of last years. The main promise is to allow a convenient scheduling of corrective maintenance, and to prevent unexpected equipment failures, saving goods, money and production time.

1.1 ISI-Welding

ISI-GF EQUIPMENT (WUHAN) CORP., LTD. is a professional manufacturer of intelligent robot welding equipment and supplier of integrated welding robot technology, with major business of research, development, production and distribution of intelligent robot welding products and welding robots. Since its establishment in 2007, the company has been focusing on the application of robot welding technology to automobile manufacturing industry, making it one of the few manufacturers in automotive industry in possession of integrated solution for intelligent welding robot for customers. As a manufacturer of intelligent robot welding equipment and assembler of welding robots, they are committed to become a first-class "supplier and service provider of all-in-one solution for intelligent robot-based welding operation", this means to design professional intelligent solution adapting to customer's special requirements on welding assembly,to supply corresponding intelligent robot welding equipment,to complete welding procedures and to provide the customer with satisfactory after-services.
By business optimization and integration, the company is transforming to a high-tech enterprise be able to deliver the customers with all-round services from technical solution, design, processing, production, manufacture, installation, delivery to training and consultation. ISI have constructed industrialized production base in Wuhan, where it's equipped with complete production, monitoring, testing and experimental installations for production and manufacture of welding control system, integrated welding machine, MF welding system, robotic automatic welding system, nonstandard welding fixture and its packages and complete automatic welding production line. ISI-GF EQUIPMENT was delisted in August 2015 and have acquired ISI-Italia (Original Italian GF Welding S.p.A.), a company specialized in welding technology with a history of 50 years, acquiring world-leading robot welding technology and development capacity, and effectively expand to global market. Acquired by SI-GF Equipment (Wuhan) Co., Ltd, ISI-Italia (Former Italian GF) is a professional firm of welding technology, a producer among few world players able to supply complete welding technology, with all know-how derived from experiences for over 50 years, full line of core products are independently developed and designed, also, it's an exclusive supplier of FIAT Italia, key supplier of Volkswagen Deutschland,

key supplier of Renault France, as well as a supplier for automobile engine manufacturers, including French PSA. Wuhan ISI-GF Eagle Automotive Equipment Co., Ltd, a controlling subsidiary of SI-GF Equipment (Wuhan) Co., Ltd, was established in 2015, the business scope of which includes: development, design and manufacture of automatic production line as well as mechanical electrical equipment; installation and refitting of mechanical equipment; manufacture, wholesale and retaining of position apparatuses, fixtures, molds and gages (Special equipment not included); development and technical transfer of automotive assembly technology; wholesale and retaining of automotive assembling tools, wires cables, electronic products, metallic materials, steel structure members, automotive parts, integrated mechanical electrical products and accessoriesISI.

1.2 Objective

The request of the ISI-Welding is to have a reliable predictive maintenance system for their welding guns.
The collaboration of the Polytechnic of Turin allowed on one hand, two students to experience the work of the company as experience for their master's thesis, while it gave the company the opportunity to take advantage of the human resources of the Polytechnic. The team of students, followed by the director of the R&D sector of ISI-Welding, eng. Dario Cambiano, were able to tackle the proposed challenge based on the study of previous works and publications and continuous comparison.
In fact, the first part of the work was a shared study: the current maintenance policies, the structure of the welding guns were addressed and the main methodologies for data processing and machine learning were studied. Later, models were produced that could approximate the phenomenon of welding in order to become aware of the phenomena on which one wants to act. After that, the works took two different paths, albeit always focused on predictive maintenance:

- The analysis of an algorithm capable of anticipating the spray of metallic material during welding
- The analysis of an algorithm capable of anticipating the need to revive the electrodes

1.3 State of art of welding guns

The spot resistance welding is a welding methodology born in the XIX century. The teacher Elia Thompson, during a physics lesson at the Franklin Istitute in 1877, by chance, invented the resistance welding. Thompson was illustrating to his student how a simple electric circuit works: an induction coil with some capacitors on the secondary winding. Once the capacitors were charged he tried to close in short circuit the primary winding, the current melted the connected ends of the wire, joining them together. This was the first resistance welding of history.
Since then, resistance welding has made enormous progress up to current technology.
The welding gun is an equipment provided by mechanical and electrical elements. In order to obtain a resistance welding spot, this machine should be able to compress the sheets to be welded and to release high currents that melts the metal through Joule effect. An actuator generates the compression force, a transformer generates high currents (tens of kA) with low voltages (15-20 V) from mains voltage (380 or 500 V). In order to avoid the melting of the gun electrical components and to better cool the weld core, the welding gun is provided by a cooling system (usually the refrigerant fluid is water).
The welding guns can be manual or robot. They obviously have the same mechanical and electrical characteristics, but the robot welding gun is the control unit of a robot and it automatically moves and performs a sequence of welding spots. The manual welding gun has to be carried to the working positions by a human operator. Welding guns can also be classified as:

- fulcrum welding gun
- slider welding gun

The difference consists in the arm movements toward the welding spot: a welding gun is provided with a fixed arm and a mobile one. In the fulcrum gun both arms are hinged to the same fulcrum, the mobile arm rotates around this fulcrum to reach the welding spot. In the slider gun the fixed arm is connected to the frame, the mobile one perform a translation movement in order to compress the sheets in the designed spot.
A fulcrum welding gun is composed from the following functional groups :

- Elctrical part:
  - **Shunt (01)** form the electrical connection between transformer and welding part.
  - **Welding part (02)** includes electrode (tips) , electrode holders and arms.
  - **Transformer (03)** inserted between the brackets and equipped with all the sensors to measure current and voltage.
- Pneumatic part:
  - **Equlizing system (04)** with valve group.
  - **Actuator cilinder** just for welding guns with pneumatic implementation.

- Mechanical part:
  - **Fixed support (05)**.
  - **Fixed joint (06)**.
  - **Moving joint (07)**.
  - **Robotic arm connection (08)**.
- Handling system:
  - **Electric motor (09)**.
- Cooling system:
  - **Hose fittings**

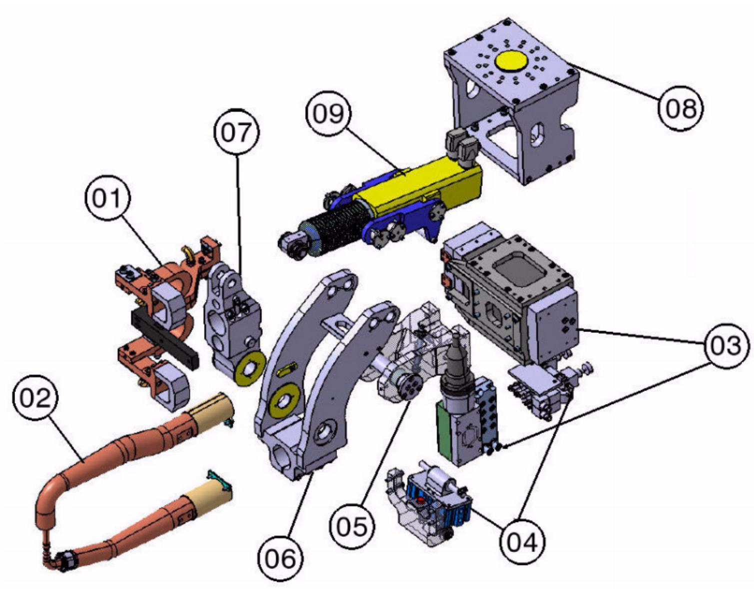

Figure 1: Decomposed fulcrum spot welding gun

1.3.1 Shunt

Flaps group is composed from RIGID CONNECTIONS realized with electrolytic copper (a) by fusion or using commercial slabs, they are screwed on the transformer case and are called fixed shunts. The shunts are individually electrically isolated using a bakelized canvas (b) and also all the bushings and washers are insulators. Then there are FLEXIBLE CONNECTIONS, the lamellar packs (c),they are called shunts and are silver-plated (to improve conductivity) copper bundles. Flaps guarantee electrical continuity allowing arms movement. Flaps are connected to the arms through clamps called brides (d), these components realize a mechanic, electrical and fluidic junction at the same time. For this reason surface tolerances (e) for these components are really restrictive. On the same surfaces the holes for the cooling system are placed.

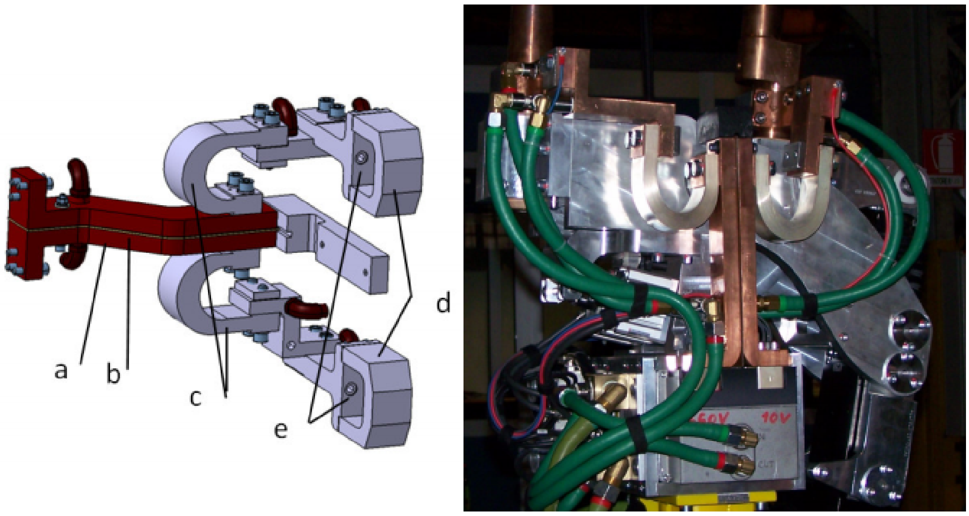

Figure 2: Flaps and shunts

1.3.2 Welding part

Welding part is composed by the arms (a) realized in copper alloy (CuCrZr) starting from commercial cylindrical sections.These components have both structural and electrical function because have to admit force trasmission to the electrodes (b) and guarantee electrical continuity between shunts and metal sheets.
On the arm is placed the electrode holder (c) according with standards required by car manufacturer companies.
On the electrode holder is positioned the electrode, often with a conical section acts to maintain the cooling fluid and to easily perform replacement operations. Inside the arm and the electrode there is the cooling circuit realized in copper (d). Fluid is managed by a brass pawl (e) with o-rings (f), the cool water travel inside the arm while the heated one has an external path. The caps (g) close the hole used for liquid insertion.

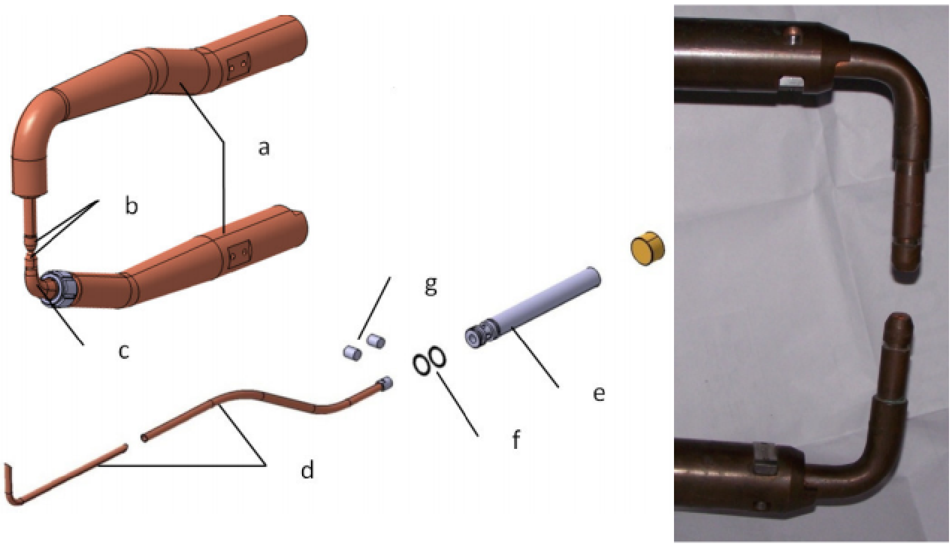

Figure 3: Arms, electrodes and cooling system

1.3.3 Transformer

The transformer (a) is correlated with 4 brackets (b) that form its cage. The transformer cage is both a support element and a robot coupling. Flaps are connected on a side of the transformer (c), on the other side there is the signal strip and power connection.

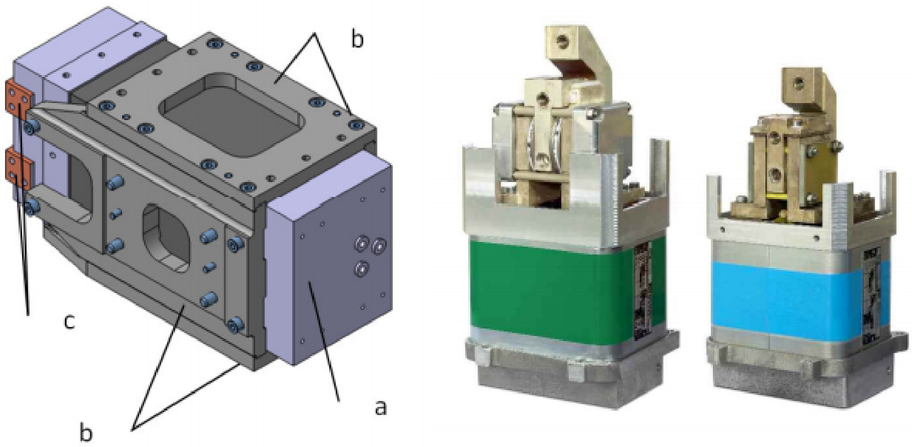

Figure 4: Transformer

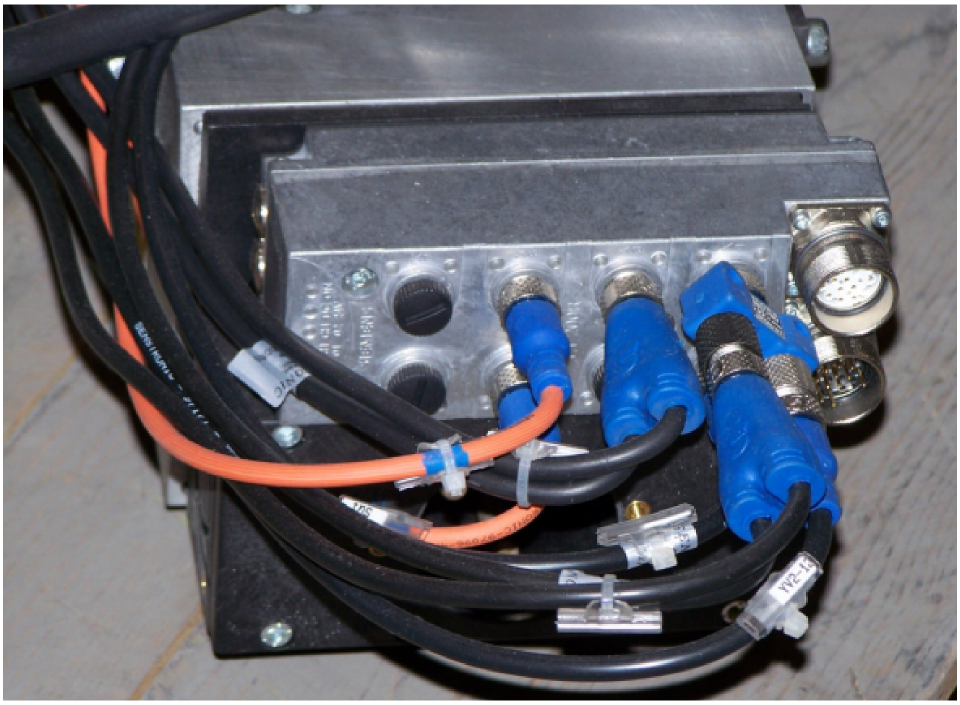

Figure 5: Signal stip detail

1.3.4 Pneumatic equalizing group

The pneumatic balancing group consists of one or two pneumatic cylinders (a).
The cylinders are hinged on one side to a bracket (b) connected in turn with the lower support, on the other side the cylinders are connected to the fixed support. The balancing assembly also carries an end pad stroke (c).
Balancing requires a control valve group that also contains manometer indicators. This valve assembly is typically placed on the transformer bracket assembly in the space-saving position.

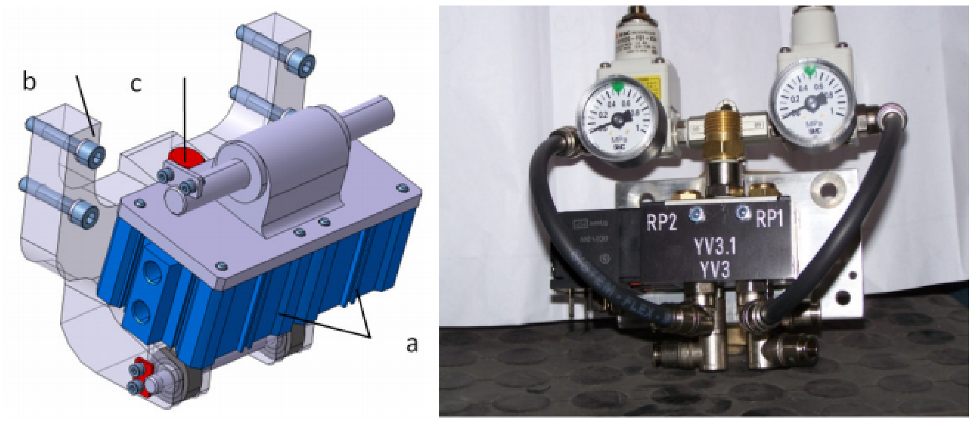

Figure 6: Pneumatic balancing group

In recent times, in order to lighten the welding gun, the pneumatic group has been removed and replaced with a software function able to equalize the force on each tip. An equalizing system compensates for welding conditions in which the closing weld tips are offset from the plane of the workpiece. As one of the tips first touch the workpiece, a force is created that slides or rotates the gun to a position that centers the gun tips about the workpiece. [?]

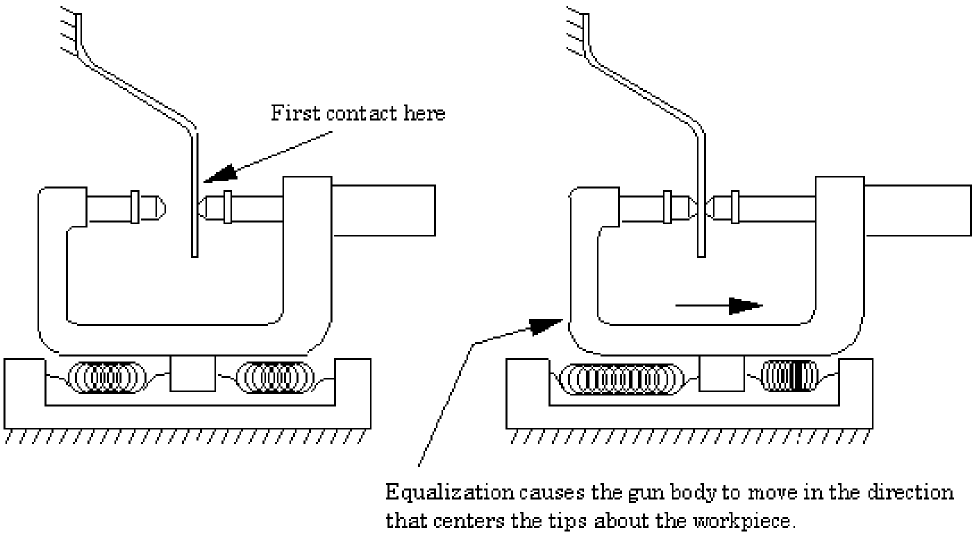

Figure 7: Equalizing system

1.3.5 Fixed support

It is the main structural element of the fulcrum clamp on which all the forces and the moments are discharged.
The fixed support is made of aluminum alloy, obtained by fusion; the support houses the pin (a) on which the joints are hinged. The support is connected at the rear to the transformer cage. It also contains the screw (b) that allows the adjustment of the balance and the seat for the cylinder connection pin balance (c).

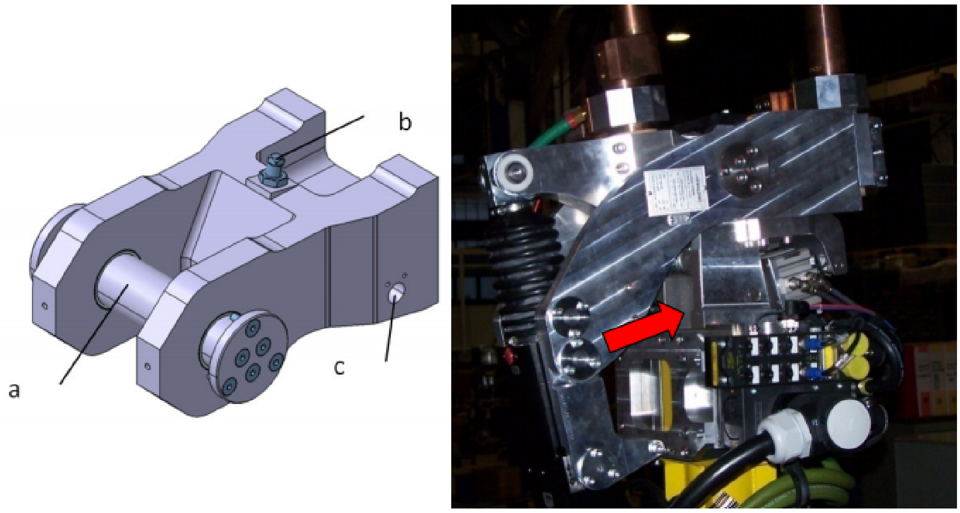

Figure 8: Fixed support

1.3.6 Fixed articulation

It is again an assembly of components in Ergal 7075 consisting essentially of two sides (a) that carry the seat of the fulcrum (b) and the seats of the connections to the handling system (c). In the lower part a support (d) is connected to the sides. In the upper part there is a reinforcement (f) which prevents the sides from twisting. The plate (i) is the guide to move the mobile arm.
To avoid the rotation of the arm during the application of the welding force it has been added an anti-rotation system consisting of a leveling on the arm and a corresponding plate screwed on the back.
The figure on the right highlights the anti-rotation system (e), the insulating bush (g) and the adapter bush (h) for the different arm diameters.

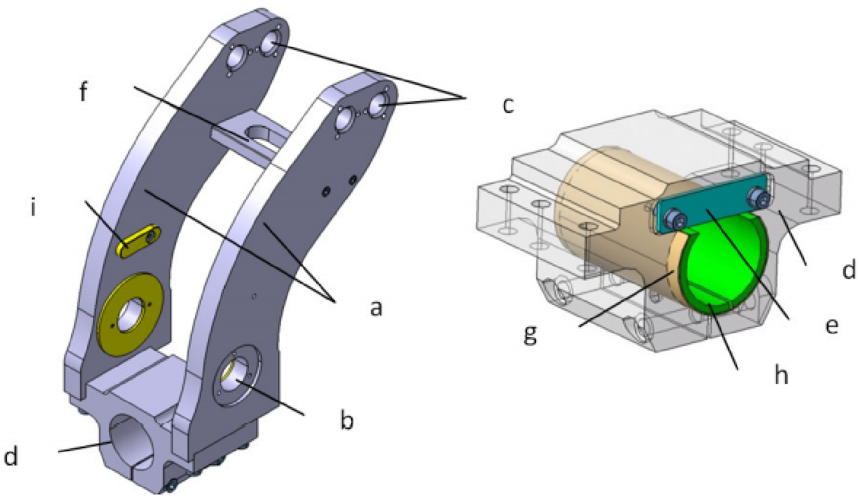

Figure 9: Fixed articulation

1.3.7 Mobile articulation

The mobile joint is inserted at the bottom of the main fulcrum, at the top it is inserted in the pin of the tenon by means of a rotating joint that allows for small misalignments of the pin axis with respect to the motor or cylinder rod.
The figure shows in yellow the insulating washers (a) in Ertalite TX, fixed with pins for prevent its rotation; these serve as insulators and help eliminate gaps between the joints. To uniquely guide the arm, pads of polymer (b) are inserted on the sides of the joint; these slide on the guides on the fixed joint.
The closing of the arm (c) in the clamp is carried out in the same way as the lower arm.

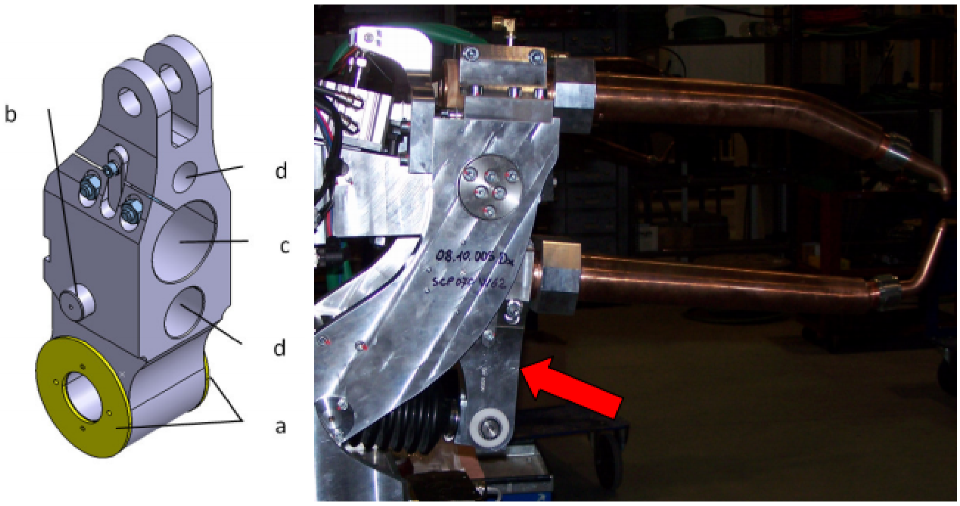

Figure 10: Mobile articulation

1.3.8 Consolle

The consolle or robot attachment is a structure in Ergal 7075 that is used to connect the robot wrist to the transformer cage. The shape of the robot side flange depends on the manufacturer's robot standards (in the figure a Comau standard for Smart series robots); the shape of the other plate must instead mate with the brackets of the transformer cage. The most common variants are two. The first variant (I), made with screwed plates, is related to upper robot attachment. In that case it is necessary to override the actuator with the console. The second variant (II), with welded plates, it is used instead when the attachment is lower or rear. Attacks can present different distances and angles between the plates. The yellow disk (a), visible in the figure, is a device of centering. Sometimes a lateral attack is also possible.

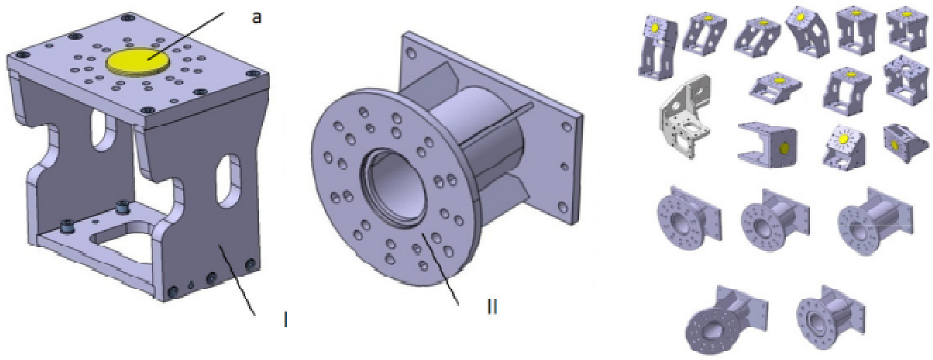

Figure 11: Consolle

1.3.9 Handling system

The most used handling systems are electric or pneumatic. The electric actuator system (a) is connected to the tenon (a component not shown which carries the seat of the attachment pin to the mobile joint); the engine, through a subframe (b), connects to the fixed joint. The pneumatic cylinder connects to the other components in the same way, but does not require auxiliary frames. However, the cylinder obviously has the so-called "bar kit" (c) that is the system of fittings, valves and filters for the management of compressed air. The pneumatic cylinder it also has micro brackets (d) for stroke adjustment, if it does not have a servo control.

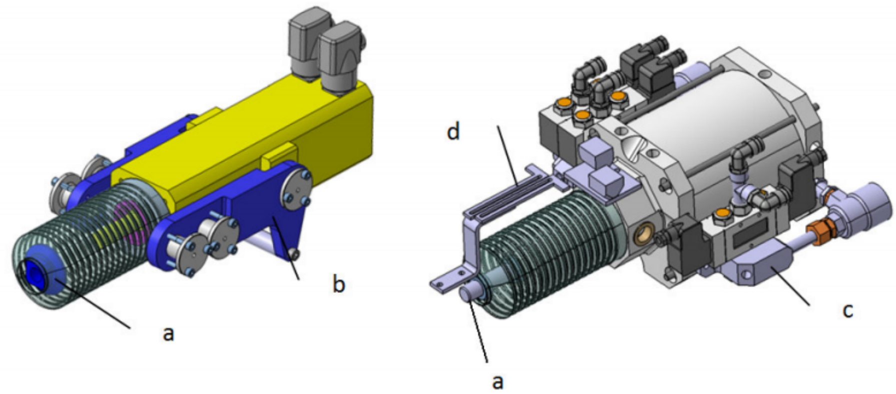

Figure 12: Handling system

1.3.10 Cooling system

The cooling system consists of two main circuits. In the clamps with 50 Hz transformer, a circuit cools one arm and the flaps, while the other circuit cools the other arm and the transformer. In the MF welding guns the cooling circuit of the transformer is independent, this is because the transformer MF requires more heat dissipation. [2]

1.4 State of art of maintenance

Maintenance is a set of technical, operational and managerial actions with the aim to guarantee the availability, cost-effectiveness and safety of systems and the optimal use of resources. A first phase of the organizational development of maintenance can be located in the '60s-'70s, when the importance of a maintenance planning and improvement was felt in sectors with growing market, such as steel, chemical, petrochemical and aeronautical one. In particular, thanks to the aeronautical industries, the reliability theory was developed. This probabilistic theory was based on mathematical and physical theories and it was aimed at estimating the remaining life of a component. A second important phase, during the '80s, was characterized by the overcoming of the maintenance endorsement, for example transferring maintenance resources to production departments and teaching the basics of maintenance to human operators. This path led to a third phase in which production appropriates the maintenance culture until the complete integration of production and maintenance strategy.

1.4.1 Lean manufacturing

Lean manufacturing, or lean production, is a production method derived from the Toyota strategy of the 1930 and it was defined from Womack and Jones as "the way to do more and more with less and less", the way to give to the customer exactly what he wants using less effort, less time, less equipment and less space. This way of thinking can be resumed in 5 key principles:

- **Value:** specify the value of the product as it is desired from the customer.
- **Value Stream:** identify the value stream for each product
- **Flow:** make the product flow continue, without interruption.
- **Pull:** introduce pull between steps to make the flow continue.
- **Perfection:** improvement

The continue improvement is then realized with the kaizen philosophy (composed of the 6 S principles).

Figure 13: 6 S flow

The modern concept of maintenance arises from 2 fundamental assumptions:

- **Increasing automation** human operator now-day is the manager and supervisor of the machine.
- **Growing competitiveness :** the imperative is to serve excellence to the customer.

In this contest, the entire production process is divided in elementary productive unit: the mini-factory. Every mini-factory carries out only one of the transformations that lead the raw materials to the final product.
The process is managed independently: the mini-factory has full responsibility also in the search for suppliers and any other customers.
Human operator in the mini-factory is the supervisor of an activity and he has to be able to guarantee the quality through self-certifications.
The self-certification defines the critical variables of the process: the aspects to which particular attention must be paid to have a satisfactory quality.
This category also includes the parameters that most influence customer satisfaction even if they are not considered, from the producer, fundamental for process quality.
The fundamental concept in this philosophy is "The Empowerment": to exploit as much as possible the resource of the human intellect.
The Empowerment provides to transfer more and more elementary functions to the driver of the machine. For this reason, modern maintenance starts with cleaning the machine. In this way the worker can learn how the machine is done, how it works and where are placed the critical aspects.
Over time, by cleaning, the human operator can learn about the places that get dirty the most but also can notice, with the experience, parts that are deteriorating and wearing out.

The worker become an integral part of the maintenance system, he will suggest to expert the critical issues to be analyzed.
After this step, obviously, the next ones are:the establishment of an information system with suitable diagnostic tools, the planning of cyclical interventions and the optimization of the life cycle cost.
The goal of the modern maintenance philosophy is to preserve the heritage but avoiding temporary actions that damage machinery in the long term.
Also increased availability and quality are values to be pursued, new technology must be viewed with distrust especially if it has not been tested for a long enough period. History is full of example of how new technologies, too reckless, have brought disastrous and sometimes even catastrophic results.
The last but not least point is costs reduction which is leading to an increasing 'Outsourcing'. [3]

1.4.2 Maintenance policies

The maintenance policy indicates the overall attitude that the organization assumes in relation to maintenance problems, which can then be explicit in the use (depending on the departments, the single machine, the convenience economic etc.) of various strategies. One milestones of maintenance approach was the *Total Productive Maintenance*, developed in Japan in fairly recent times, defined as "production maintenance carried out by all workers of the company organized in small groups of activities". It is a comprehensive approach to organisational issues with a view to improving the performance of production equipment and plants, which takes into account the Japanese matrix and the application experiences made in the Italian industry[4].
The main innovation was to bring the responsibility of the line maintenance and the quality control to the coordinator of a production segment at the operational level. So TPM's main contribution to maintenance theory is given by the attempt to break down the existing demarcation line, within a company, between maintenance and production departments. In this context, the TPM acknowledges the existence of several maintenance situations which may require different techniques to achieve a good result, and consequently it uses different methodologies which can differ from plant to plant or from machine to machine, provided that they are effective in a given situation.
Many of the strategies used are certainly not new: what is innovative is the Japanese culture, the commitment it provides for all employees. Business maintenance policy optimization should be pursued in the context of improving business profitability and the service provided and, in particular, the continuous improvement in operating income. This improvement is the expression of a close synergy between maintenance and production which takes the form of production maintenance.

1.4.3 Maintenance strategies

The maintenance activity aims to obtain a certain continuity of the production process. In the past this objective was pursued through operational and functional redundancies or applying an aggressive program of review and replacement of critical systems. All

these approaches have proved to be partially inefficient, as redundant systems and excess capacity freeze capital that could be more profitably used for productive activity, while a political revisions excessively prudent has proved to be a rather expensive method to obtain the required standards.
Maintenance has therefore transformed from operational repair activities to a complex management system with the point of preventing failure.

Breakdown maintenance:
Breakdown Maintenance is certainly the most spontaneous and simple way to work: maintenance action is taken when the failure occurs. In the presence of non-critical and easy-to-replace systems at low cost, it is convenient wait for the failure to occur before intervening. Unfortunately, this strategy has many questionable aspects: a serious and unexpected failure on a component may have deleterious consequences on other elements of the system, compromising its functionality with an additional amount of costs, moreover unscheduled repairs often take a long time to obtain spare parts and assign the appropriate technician, stopping the production and poorly employing human operators. Finally, a sudden or catastrophic failure is a condition that a good maintenance activity should avoid a priori.

Preventive maintenance:
Preventive Maintenance is based on the belief that the average life of some component is determinable and that it is possible to anticipate the failure of a system (machine or production line), predefining the moment of intervention, usually replacement, depending on the expected life time of the component itself. This concept was a great success in the 1960s and 1970s with the spread of the reliability theory, because it gave a basis of scientific nature to the maintainers. It is a type of maintenance that is one step higher than the previous one, because in this case the mechanical system is still working but its performance deteriorate to the state of imminent failure. There are two philosophies to implement a failure avoidance:

- Condition-based, that promotes maintenance only when necessary by means of a shallow observation of the system and the detection of the deterioration,
- Time-based, that schedule the interventions at constant interval on the basis of reliability, safety and performance.

The weak point of this strategy is that the reliability theory is a probabilistic theory, so a failure can happen also before the scheduled part replacement, but mainly there are no chances to increase the mean time between two subsequent failures of the system.

Predictive maintenance:
A modern view of maintenance problems led to the use of non-destructive techniques for testing systems for the purpose of identifying with a consistent advance the presence of faults, so it is possible to schedule a review only when the condition of the machine determines its necessity. This maintenance strategy does not use probabilistic methods

for making a prognosis of the failures, but it uses the trend of tracked parameters to predict potential failures. This is the predictive maintenance: a diagnostic process that, by providing information on the health status of machine allows to plan revisions based on the actual conditions of the components rather than on the operating time. Unlike the earlier described condition based preventive maintenance, this new philosophy has significant implications on design: in fact, to reduce to minimum passive times due to frequent checks, the mechanical system should be equipped with a whole series of devices necessary for the determination of the status efficiency of components.

Table 1: Benefits of the predictive condition-based maintenance

Safety	Predictive maintenance allows machine downtime before reaching critical condition
Increase in availability, lower costs maintenance	The intervals between two successive revisions may be increased. Downtime can be reduced by preparing maintenance resources
Better chance of negotiation with manufacturers	Because the conditions are measured on new machines, at the end of the warranty and after the review it is possible have some comparison data
Better relationships with customers	Knowing in advance when a failure will occur, it is possible better organize production
Opportunities to design better future plants	The experience properly collected in historical files can be useful for this purpose

The limit of predictive maintenance can be identified as being failure-oriented: it is more effective than traditional approaches, but leaves wide areas of improvement in terms of reliability and cost reduction. This strategy tries to provide the operator an sufficient warning alert to organize the necessary repairs and the downtime. This depends, of course, on the monitoring program and the time needed to obtain the results of the analyses: if this time is large, an incipient failure conditions may transform in imminent failure one, bringing the system into much more worrying conditions.

Proactive maintenance:
All these maintenance strategies can be defined as 'reactive' strategies. In the Proactive (or productive) Maintenance, the term 'proactive" is opposed to the concept of reaction, in the sense that it refers to an action that takes place before the critical event. It is a pre-alert activity that is carried out before any damage relating to the equipment or performance of the system, a series of actions with the aim of correcting those conditions which may lead to deterioration of the system. Instead of analyzing material or performance alteration to evaluate incipient or imminent failure conditions, the proactive maintenance is proposed to detect and correct abnormal values of primary causes of failure that could lead to conditions of operational instability, the so-called 'failure roots', and report that first level of malfunction, the 'conditional failure'. This maintenance practice is the first line of defense against the degradation of material (incipient failure) and the consequent weakening of the performances (imminent failure) which finally lead to the breakdown. Moreover, intervening with such advance make possible to avoid the occurrence of secondary failures

that may arise on the elements adjacent to that in examination (for example because of vibrations). Summing up, the proactive maintenance requires the following actions:

- monitoring of the key parameters indicative of the health of the system (failure roots),
- definition of threshold values, that is the maximum acceptable values for each parameter,
- recognition and interpretation of any outlier of these key parameters, which indicate some instability in operating conditions,
- Specification of the methods to be used to correct primary failure causes and restore system stability

Table 2: Failure classification

Failure type	**Description**
Catastrophic failure	A condition of sudden and complete cessation of operations and a total deterioration of functions.
Sudden failure	A condition of accelerated degradation of both material and performance, which results in a partial weakening of functions.
Imminent failure	A condition of perceptible degradation of the material in the presence of a serious deterioration in performance.
Incipient failure	A condition in which the use of appropriate means of investigation allows to identify the first signs of degradation of the material, without the user experiencing any change in the performance of the machine.
Conditional failure	A condition of pre-alert in which it has not yet occurred a deterioration neither of the material nor of the performance but such that, if the situation persists, it will inevitably lead to a functional failure.

The evaluation of the failure roots is not always possible, sometimes there are no ways to detect them or sometimes it is too expensive. Implementation of the maintenance policy requires design criteria based on the logic of minimizing the overall cost. The first step is analyzing a specific failure mode and verifying the existence of measurable signals that can help its detection. If the signal exists, it is possible to perform a predictive condition-based maintenance by monitoring the degradation of the component. If the signal does not exist, then the analysis moves on the theory of reliability and the estimated life of the component. If there are enough information about this topics, a preventive maintenance can be implemented activating planned inspections or performing replacements at scheduled times. When there is no signal e no estimated life of the component, the breakdown maintenance is the only possible strategy.

1.5 State of art of machine learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.
Machine learning starts with the collection of a great amount of data, they can be measured on a real plant or simulated using an identified model of the phenomenon.
In this way the algorithm can look for common patterns in data and make better prevision in future based on the examples that we provide. Machine Learning algorithms can be categorized as supervised or unsupervised.

- **Supervised machine learning algorithms** the data sets provided to train the algorithm are labeled examples. The learning algorithm produces an inferred function to make prediction about the output values and compare its output with the given label to modify the parameters inside it accordingly.

- **Unsupervised machine learning algorithms** are used when the information is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output but can describe hidden structure from unlabeled data.

- **Semi-supervised machine learning algorithms** between supervised and unsupervised learning, since this kind of algorithm uses both labeled and unlabeled data for training. Typically a small amount of labeled data and a large amount of unlabeled. This method is able to improve learning accuracy. Semi-supervised learning algorithms are chosen when to have labeled a great amount of resources are required. Otherwise, acquiring unlabeled data generally does not require additional resources.

- **Reinforcement machine learning algorithms** is a learning methodology able to interacts with its environment by producing actions and discover errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. A simple reward feedback is required for the agent to learn which action is the best; this is known as the reinforcement signal.

Machine learning allow the analysis of massive quantities of data and, to be trained properly, it may requires additional time and resources.

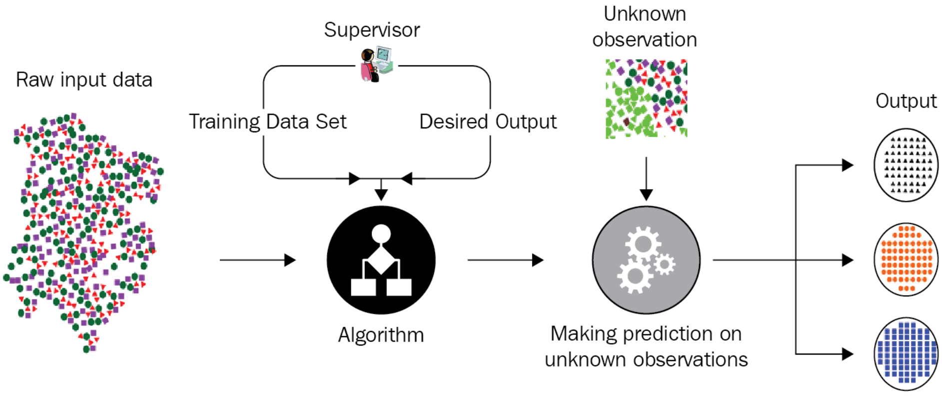

Figure 14: Machine learning flow

1.5.1 Classification Problem

In statistics, classification is the problem to identify to which, of a set of categories, a new observation belongs. Today, for example, this approach is widely used to image recognition applied on autonomous vehicles. The success of this kind of algorithm is entrusted by the presence of a sufficiently wide data set, most of the time labeled, that allows the solution of a supervised problem. In the terminology of machine learning the corresponding unsupervised procedure is known as clustering, and involves grouping data into categories based on some measure of inherent similarity or distance. Often, the individual observation are analyzed into a set of quantifiable properties, known as explanatory variables or features. These properties may be categorical, ordinal, integer-value or real-valued.
Other classifiers work by comparing observations to previous observations by means of similarity or distance function. An algorithm that solves a classification problem, is a classifier. The term "classifier" sometimes also refers to the mathematical function, implemented by a classification algorithm to map the data into a category.
Classification and clustering are examples of the more general problem of pattern recognition, which is the assignment of some sort of output value to a given input.
A subclass of classification is probabilistic classification. This kind of approach uses statistical inference to find the best class for a given instance.

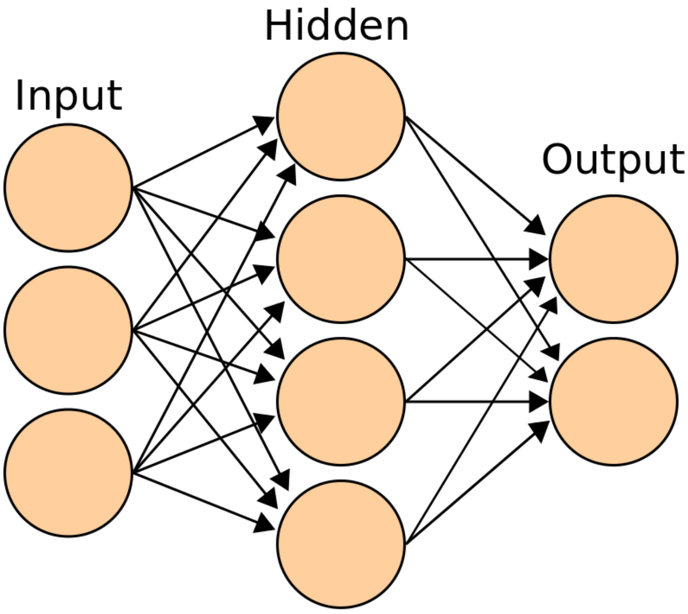

Figure 15: Neural network simple graphical structure

The most commonly used classification algorithms are:

- **Naive Bayes classifier:** makes use of simple "probabilistic classifier" based on applying Bayes'theorem with naive independence assumptions between features.
- **K-nearest neighbor:** k-NN classification has a class membership as output. An object is classified by a plurality vote of its neighbors, with the oobject being assigned to the class most common among its k nearest neighbors (k is typically a small integer value). If k=1, then the object is simply assigned to the class of that single nearest neighbor.
- **Decision Tree:** in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.
- **Neural network:** ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection can transmit a signal to other neurons. AN artificial neuron receives multiple signals as input and processes a single output that would become the input of other neurons. Neurons typically have a weight that adjusts as learning proceeds.

1.5.2 Regression problem

Regression algorithms belong to the family of Supervised Machine Learning. Regression algorithms predict the output values based on input features from the data fed in the system. The go-to methodology is the algorithm builds a model on the features of training data and using the model to predict the value for new data. Today, regression models have many applications, particularly in financial forecasting, trend analysis, marketing, time series prediction and even drug response modeling. Some of the popular types of regression algorithms are linear regression, polynomial regression, lasso regression and multivariate regression.

- **Simple Linear Regression model:** it is a statistical method that study relationships between two continuous (quantitative) variables. In linear regression, a model assumes a linear relationship between the input variables (x) and the single output variable (y). In this way the output can be computed from a linear combination of the input variables. When there is a single input variable, the method is called a simple linear regression. When there are multiple input variables, the procedure is referred as multiple linear regression. Sometimes this algorithm is affected by underfitting problem when a linear relationship is not enough to estimate the output.

- **Polynomial Regression model:** the main difference between this algorithm and the previous one is that the model is not linear, it is slower but has a greater accuracy. The underfitting problem is thus avoided, on the contrary an overfitting one can arise. The overfitting is due to an excessive adaptation to the training set with the loss of ability to correctly estimate new data.

- **Lasso Regression:** LASSO stands for Least Absolute Selection Shrinkage Operator. Shrinkage is defined as a constraint on parameters. Lasso regression is aimed to obtain the subset of predictors that minimize prediction error for a quantitative response variable. The algorithm starts imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward a zero, then variables with a zero-regression coefficient are excluded from the model. Variables with non-zero regression coefficients variables are strongly associated with the response variable. This lasso regression analysis is basically variable selection method and it helps analysts to determine which of the predictors are most important.

- **Multivariate Regression:** this algorithm is used when there is more than one predictor variable in a multivariate regression model, so it is implemented to predict the response variable for a set of explanatory variables.

2 Model

2.1 NARX model

The nonlinear autoregressive network with exogenous inputs (NARX) is a recurrent dynamic network with feedback connections enclosing several layers of the network. This particular application is very useful with time-series data. It can be used as a predictor, for nonlinear filtering and for the modeling of nonlinear dynamic systems. The defining equation of the NARX model is:

$$y(t) = f(y(t-1), ..., y(t-n_y), u(t-1), ..., u(t-n_u)) \tag{1}$$

where the next value of the dependent output signal $y(t)$ is regressed on previous values of the output signal and previous values of the independent input signal. It is possible to design a feedforward neural network to approximate the function f. So, the output of the network is an estimate of the output of the examined system [7].

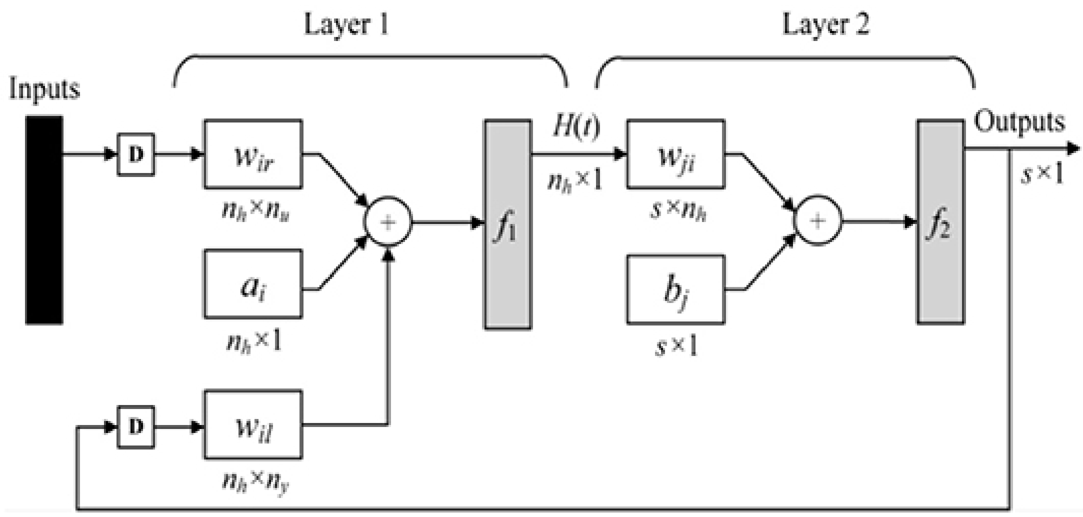

Figure 16: Narx architecture

During the training, it is more convenient to use that the true output as delayed input instead of the estimated one, in an open-loop architecture. In this way the inputs of the feedforward network are more accurate and the neural network is better trained.
After the training, there are two different possible ways to implement this algorithm according on the available measurements. If the output is measured during the process, it is possible to feed the network with its real values, without feeding back the estimated ones. On the other hand, if the output is not measured or its measures are not readily available, it is possible to close the loop and to feed back the estimated output as new network input.
Different types of neural network architecture have been tested, changing input delays and layers size. The performance have been evaluated through a fit parameter, defined as:

$$fit = 1 - \sqrt{\frac{MSE}{\frac{1}{N}\sum_{t=1}^{N}(y(t)-\bar{y})^2}} \tag{2}$$

where N are the samples of the output, $\bar{y}$ is the mean of the output values $y(t)$ and the Mean Squared Error (MSE):

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(y(t)-\hat{y}(t,\theta))^2 \tag{3}$$

where $\hat{y}(t,\theta)$ is the estimated output. The MSE is a measure of the quality of the estimator, the closer its value is to zero, the better the estimator is. In fact the MSE takes into account the variance of the estimator and the its bias with respect to the real output. However, it can not be considered as a reliable index since it has the squared measurement unit of the output and it strongly depends on numerical values of data. This is why the fit parameter has been chosen as the best index for performance evaluation.

2.1.1 Results

Different tests have been performed using the NARX methodology. At first, short-circuit data have been used, in particular short-circuit data after the electrode dressing. This data represent the most 'ideal' conditions that it is possible to obtain on welding guns, without disturbances introduced by metal sheets and electrode pollution, so they are useful to build and evaluate a model.
Against this background, a neural network with 10 delayed states (both external input and feedback output) and two hidden layers with 30 and 5 neurons has been implemented. A training set of 70 welding spot data has been fed to the network. A first test is made with the open-loop network. This is a realistic choice because the input current and the output voltage are measured and readily available in order to detect disturbances. In addition, it is clear that the network has a higher accuracy 'feeding back' real values than the estimated ones.

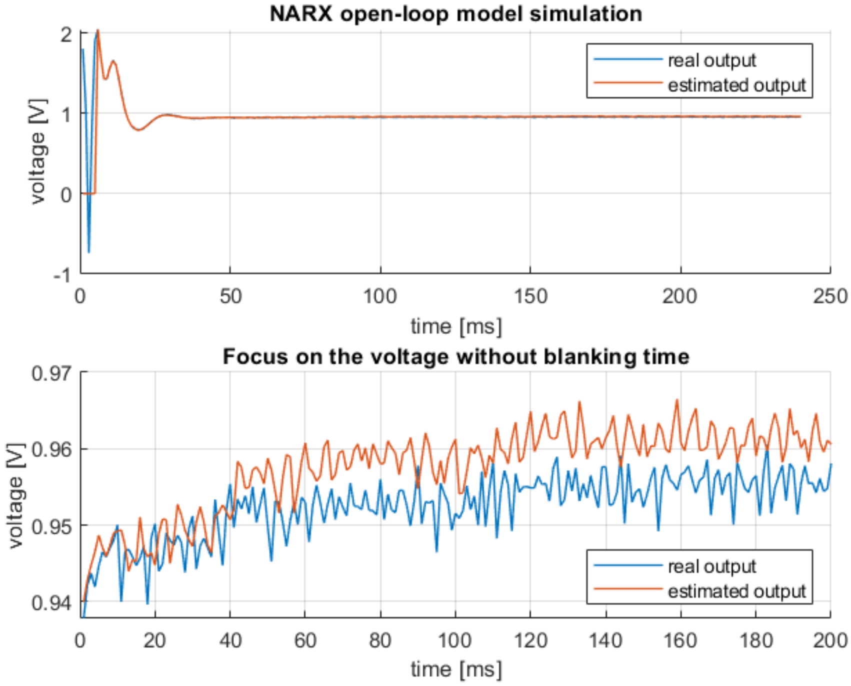

Figure 17: Comparison between the real and the estimated output

In this simulation, the MSE is $3.73 \cdot 10^{-5}$ while the fit is about 95.2%. The simulation starts with a 10 ms delay, this is due to the initial acquisition of the delayed input and output by the neural network that can begin to estimate only once acquired these data. An interesting application of this algorithm is obtained by closing the loop of the neural network. This can be a choice when the output is not available while the system is working and the neural network takes his own estimated output as input for the estimation of the subsequent step. Unfortunately, in this case the simulation does not lead to good results:

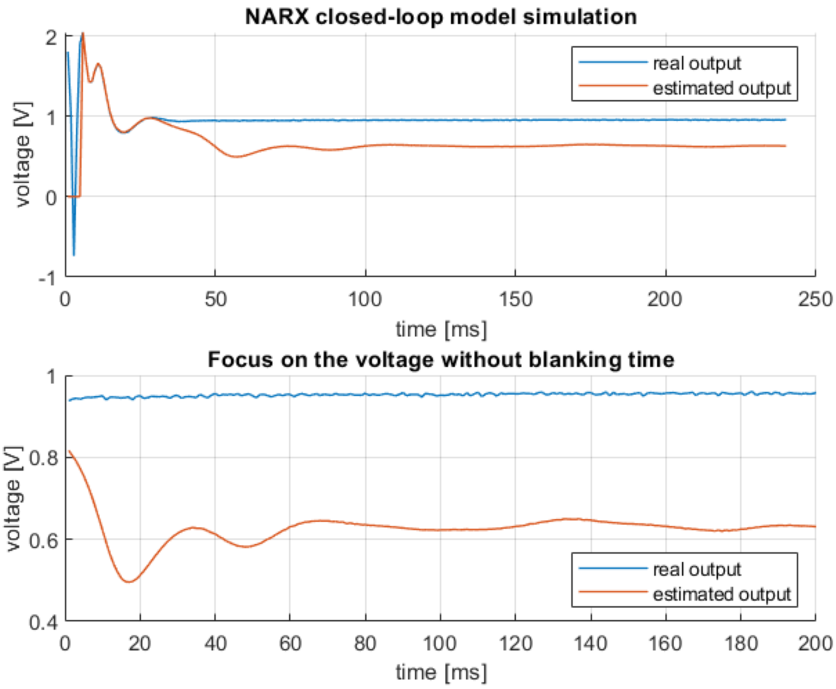

Figure 18: Comparison between the real and the estimated output

It is clear that the estimated output is not able to approximate the real one. This 'failure' may be due to the incapability of the electrical data to fully describe the phenomena involved in a welding process.
Anyway, it can also be considered that the most significant data are the ones collected after the blanking time (first 40 ms). A better estimation with the closed-loop is obtained by removing both in training and test data those values collected during the blanking time and re-designing the neural network for the new setting. The following figure shows the result with 5 delayed states and two hidden layers with 20 and 10 neurons:

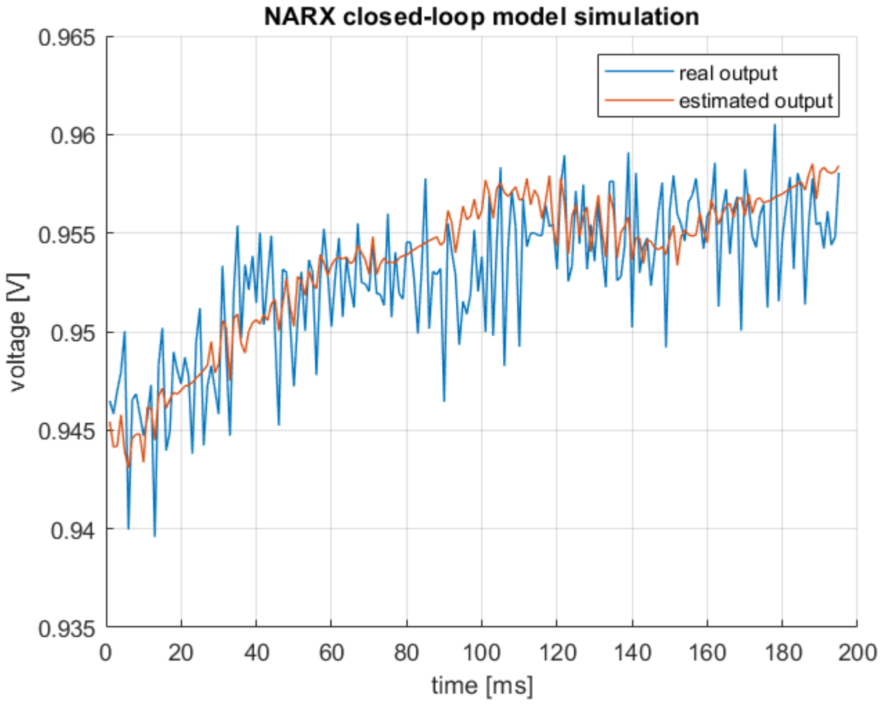

Figure 19: Comparison between the real and the estimated output

The estimation is still not excellent but it is able to follow the real output also in closed-loop, without any information on its real values.
Another attempt is made using data collected from short-circuit points before the electrode dressing, after 100 real welding points. In this case the disturbance introduced by metal sheets are still avoided, but the electrode is polluted. A neural network with 10 delayed states and two respectively 40 and 10 neurons hidden layers has been designed and trained with the new dataset. Performance are similar to the previous case, with a MSE of $2.79 \cdot 10^{-5}$ and a fit of about 93%.

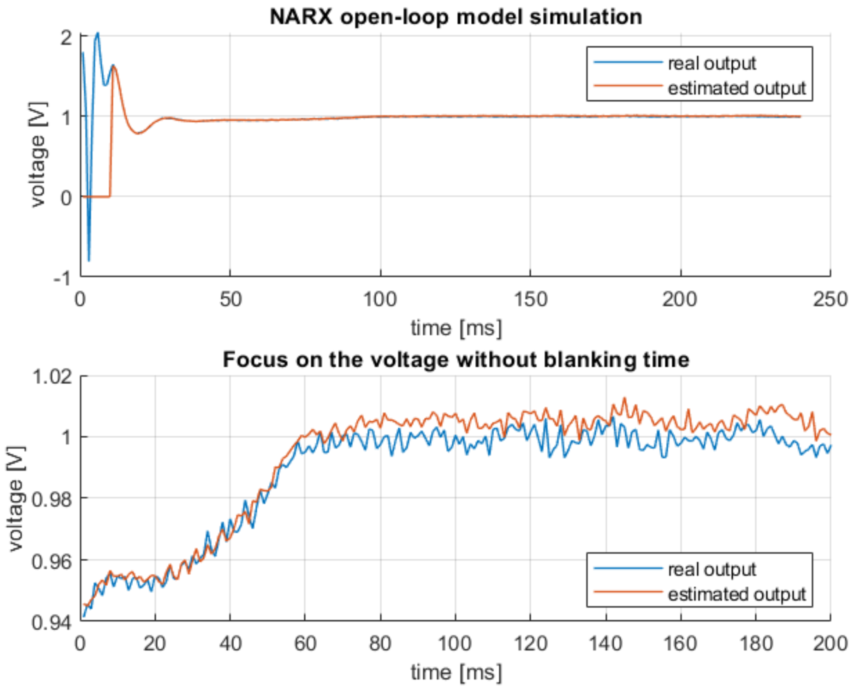

Figure 20: Comparison between the real and the estimated output

The last possible attempt is made using real welding data, taking into account disturbances and discovering if these data can be adequate for a model. 10 delayed states, a 30 and a 5 neurons hidden layers are the characteristics of the neural network. There is higher availability of real welding points with respect to the short-circuit ones, in fact a set of 500 points has been selected as training set. There are points affected by splash. A first test has been performed on a welding point that has not presented the splash:

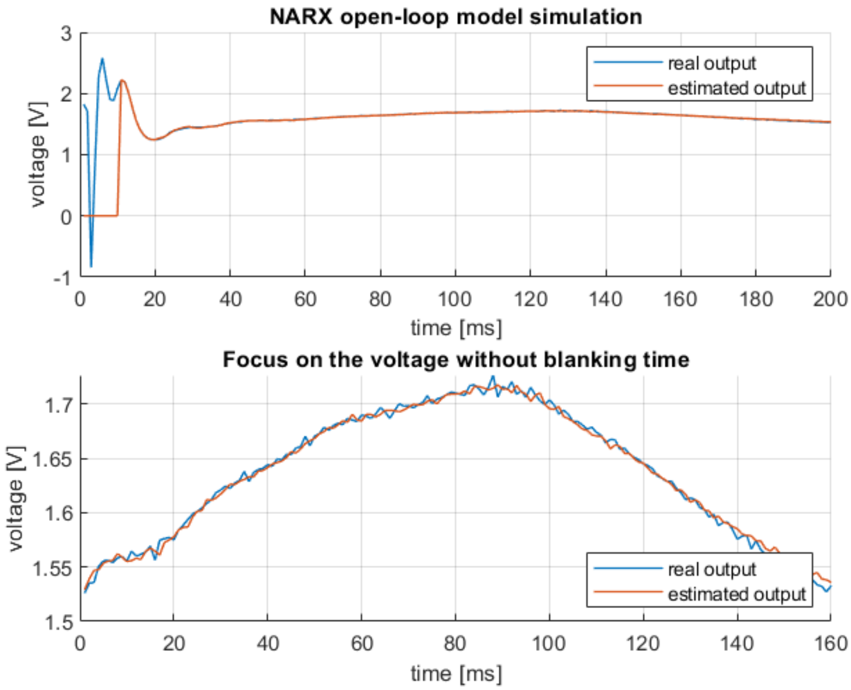

Figure 21: Comparison between the real and the estimated output

With a MSE of $3.1 \cdot 10^{-5}$ and a 95.4% fit, the estimation is able to approximate the real output.
The following figure represents the test implemented using a point affected by splash:

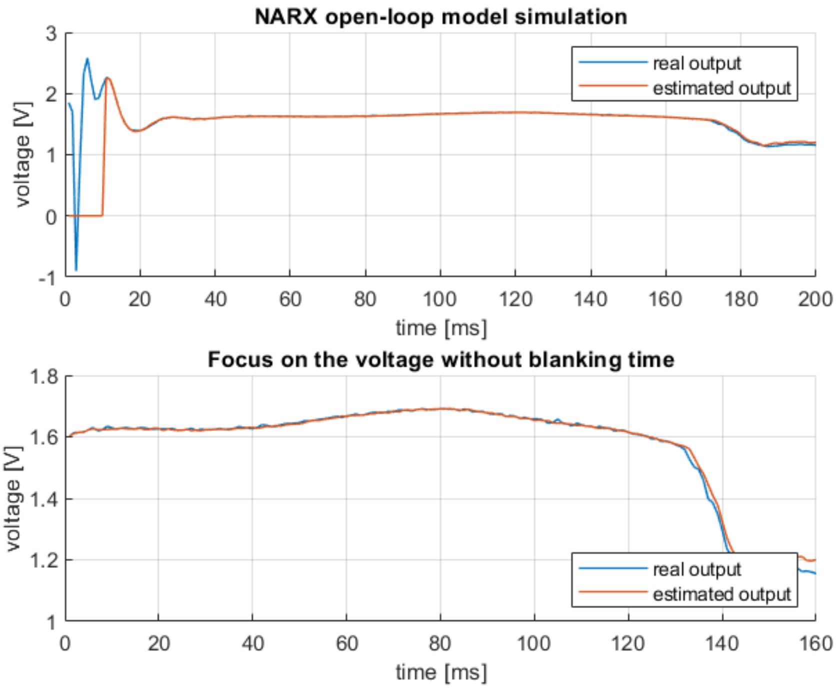

Figure 22: Comparison between the real and the estimated output

The indices are worst (MSE=$1.86 \cdot 10^{-4}$, fit=92%) but the general trend is still acceptable. Actually, the splash is well estimated and this open the way to a possible forecasting attempt using the same theoretical basis of this algorithm, with the necessary adaptation for the different kind of problem. A final closed-loop test is performed:

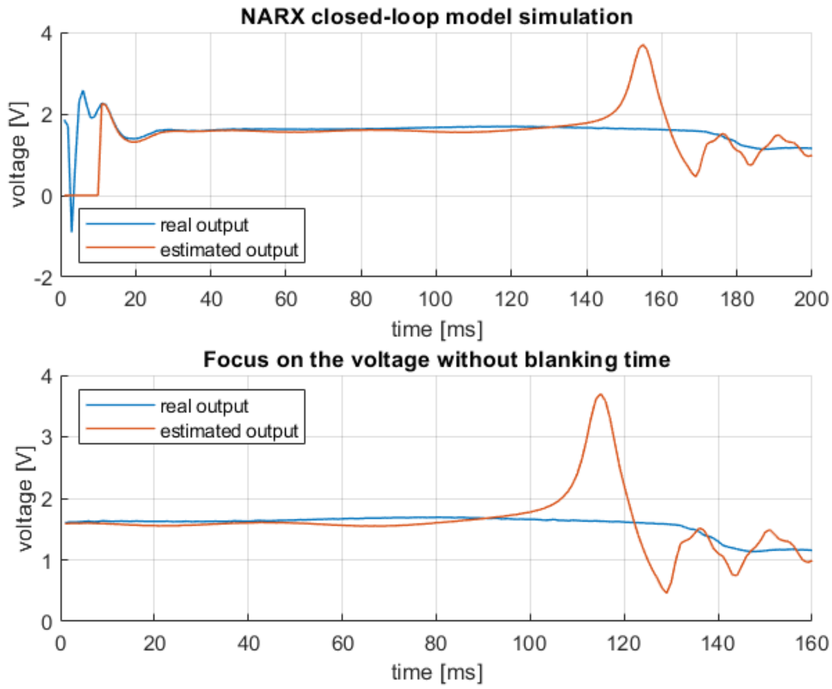

Figure 23: Comparison between the real and the estimated output

As expected the estimation is completely inappropriate. There are too many factors not considered and not yet available that affect the welding process.

System identification is aimed at constructing or selecting mathematical models M by dynamical data, generated by a system S, to serve certain purpose (forecast, diagnostic, control, etc.).
The first step is to determine a class m of models within to search the most suitable model. There are 2 possible models to find :

- **transfer-function models**
- **state-space models**

The system identification problem may be solved using an iterative approach:

- Collect the data set :
  - design the experiment so that the data can be maximally informative.
  - pre-filtering technique of the data.
- Choose the model set or the model structure:
  - physical model with some unknown parameters may be constructed by exploiting the possible a priori knowledge and insight.
  - black-box model may be employed, in this case the given data are elaborated without a physical reference.
  - gray-box model may be used, with adjustable parameters having physical interpretation.
- Determine a suitable complexity level of the model set or model structure.
- Tune the parameters to pick the 'best' model in the set (guided by data).
- Perform the model validation test.

At the end of this model development cycle, if the found model greatly approximate the behavior of the real data it is possible to use it, otherwise there is the necessity to restart from the beginning criticizing the data, the model orders or the other choices made in the development phase.
One of the approaches used to find a relationship between input and output is the polynomial identification. In this way we assume to have a completely unknown system (black-box) with only the measured data. This approach has been used in order to try to see if the polynomial relationship could suggest something about the physical model.

2.2 Polynomial model

Principal families of dynamic model can be considered as a particular case of:

$$y(t) = G(z)u(t) + H(z)e(t) \tag{4}$$

where $y(t)$ is the measured output, $u(t)$ is the command input and $e(t)$ is the error. $G(z)$ and $H(z)$ are transfer functions, given from the relationship between polynomials, in which parameters have to be estimated with precise identification methods.

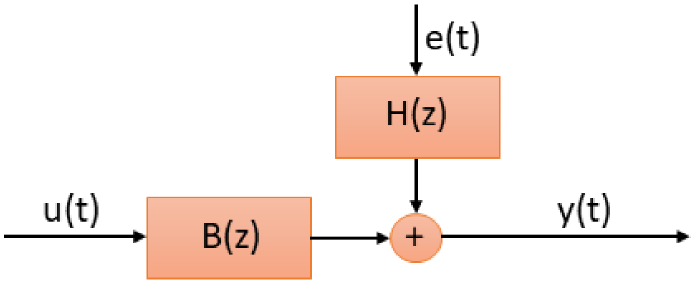

Figure 24: Example of the system to identify

The signal $v(t) = H(z)e(t)$ can be seen both as a noisy agent on the overall system or as a phenomenon not modeled from $G(z)u(t)$.
Moreover, it is possible to suppose that $H(z)$ has all the poles with modulus less than 1 and roots of the numerator with modulus less or at least equal to 1.
Specializing $G(z)$ and $H(z)$ in particular structures, it is possible to obtain different families of black-box models[6].

2.3 ARX model

An 'AutoRegressive eXogenous' model has the form:

$$y(t) = -a_1 y(t-1) - ... - a_{n_a} y(t-n_a) + b_1 u(t-1) + ... + b_{n_b} u(t-n_b) + e(t) \tag{5}$$

The noise enter as a direct error.
If z^{-1} is denoted as the unitary delay operator such that $z^{-1}y(t) = y(t-1)$ and $z^{-2}y(t) = y(t-2)$, is possible to define:

$$\begin{aligned} A(z) &= 1 + a_1 z^{-1} + a_2 z^{-2} + ... + a_{n_a} z^{-n_a} \\ B(z) &= b_1 z^{-1} + b_2 z^{-2} + ... + b_{n_b} z^{-n_b} \end{aligned} \tag{6}$$

then, the above relationship can be written as:

$$A(z)y(t) = B(z)u(t) + e(t) \Rightarrow y(t) = \frac{B(z)}{A(z)}u(t) + \frac{1}{A(z)}e(t) = G(z)u(t) + H(z)e(t) \tag{7}$$

where:

$$G(z) = \frac{B(z)}{A(z)}, H(z) = \frac{1}{A(z)} \tag{8}$$

The blocks scheme is:

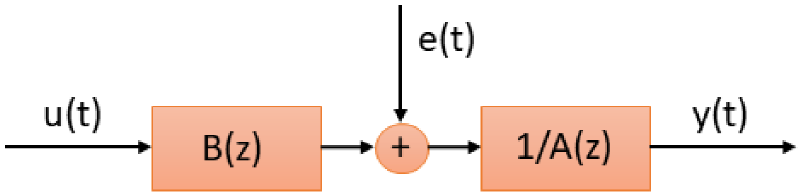

Figure 25: Example of an ARX model block diagram

It is possible to see how the noise pass through the term $1/A(z)$, the meaning is that the noise acts on the state of the system. The input is known as exogenous variable, then the model contains the **autoregressive (AR)** $A(z)$ and the **exogenous (X)** $B(z)$ parts. The integers n_a and n_b are the orders of these two parts of the ARX model[6].

2.4 ARMAX model structure

The input-output relationship of the 'AutoRegressive MovingAvarage eXogenus' model is a difference linear equation:

$$\begin{gathered} y(t) + a_1 y(t-1) + a_2 y(t-2) + ... + a_{n_a} y(t-n_a) = \\ b_1 u(t-1) + ... + b_{n_b} u(t-n_b) + e(t) + c_1 e(t-1) + ... + c_{n_c} e(t-n_c) \end{gathered} \tag{9}$$

where the white-noise $e(t)$ enters as a linear combination of $n_c + 1$ samples.
By introducing the polynomials:

$$\begin{aligned} A(z) &= 1 + a_1 z^{-1} + ... + a_{n_a} z^{-n_a} \\ B(z) &= b_1 z^{-1} + ... + b_{n_b} z^{-n_b} \\ C(z) &= 1 + c_1 z^{-1} + ... + c_{n_c} z^{-n_c} \end{aligned} \tag{10}$$

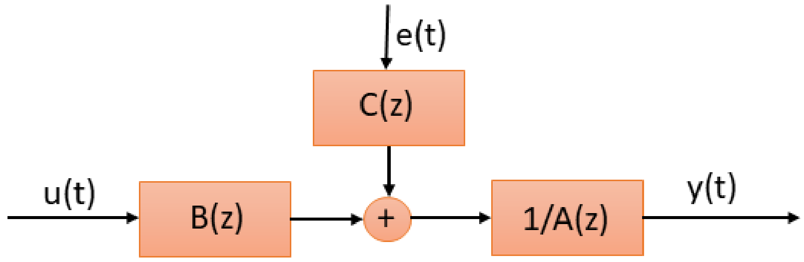

Figure 26: Example of an ARMAX model block diagram

The input output relationship can be written as:

$$\begin{aligned} A(z)y(t) &= B(z)u(t) + C(z)e(t) \\ y(t) &= \frac{B(z)}{A(z)}u(t) + \frac{C(z)}{A(z)}u(t) = G(z)u(t) + H(z)e(t) \end{aligned} \tag{11}$$

where: $G(z) = B(z)/A(z)$ and $H(z) = C(z)/A(z)$.
The **auto-regressive (AR)** part is included in the term $A(z)y(t)$, the **exogenous (X)** part in $B(z)u(t)$ and the **moving average (MA)** part in $C(z)e(t)$ (which is a colored noise instead of the white one).
The integers n_a,n_b,n_c are the orders of these three parts of the ARMAX model (ARMAX(n_a, n_b, n_c))[6].

2.5 OE model structure

The relationship between input and undisturbed output is a linear difference equation:

$$w(t) + f_1 w(t-1) + ... + f_{n_f} w(t - nf) = b_1 u(t-1) + ... + b_{n_b} u(t - n_b) \tag{12}$$

and the model output is corrupted by white measurement noise:

$$y(t) = w(t) + e(t) \tag{13}$$

By introducing the polynomials:

$$F(z) = 1 + f_1 z^{-1} + ... + f_{n_f} z{-}n_f \tag{14}$$

$$B(z) = b_1 z^{-1} + b_2 z^{-2} + ... + b_{n_b} z^{-n_b} \tag{15}$$

The above input-undisturbed output relationship can be written as:

$$
\begin{aligned}
F(z)w(t) = B(z)u(t) \Rightarrow \\
y(t) = w(t)+e(t) = \frac{B(z)}{F(z)}u(t) + e(t) = \\
= G(z)u(t) + e(t)
\end{aligned} \tag{16}
$$

where $G(z) = B(z)/F(z)$.
The integers n_b and n_f are the orders of the OE model, denoted as OE(n_b,n_f)[6].

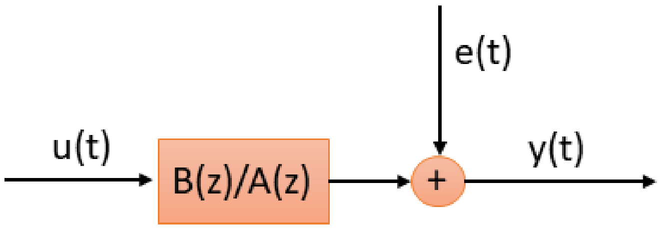

Figure 27: Example of an OE model block diagram

In the ARX and ARMAX model, the poly A(z) is the denominator of every component, this is often an uncomfortable situation (too much restrictive).
To relax this hypothesis it is possible to use the OE structure in order to have a better simulator of the real plant.
On the other hand, if a one step predictor of the system is needed, the ARX structure gives better results.

2.6 Data analysis

The data measured during the short circuit welding point has been used to search the polynomial model.
As a matter of fact, this kind of welding point are taken from the company before and after the dressing of the electrodes to take track of resistance changes during the dressing.
In particular the short circuit welding point considered are the ones taken after the dressing.
This to try to identify the most ideal condition possible, without the uncertainty introduced by the electrodes worn out and the metal sheets interposed.
The data have been filtered with a low pass filter, in order to reduce the noise and the mean value has been removed.

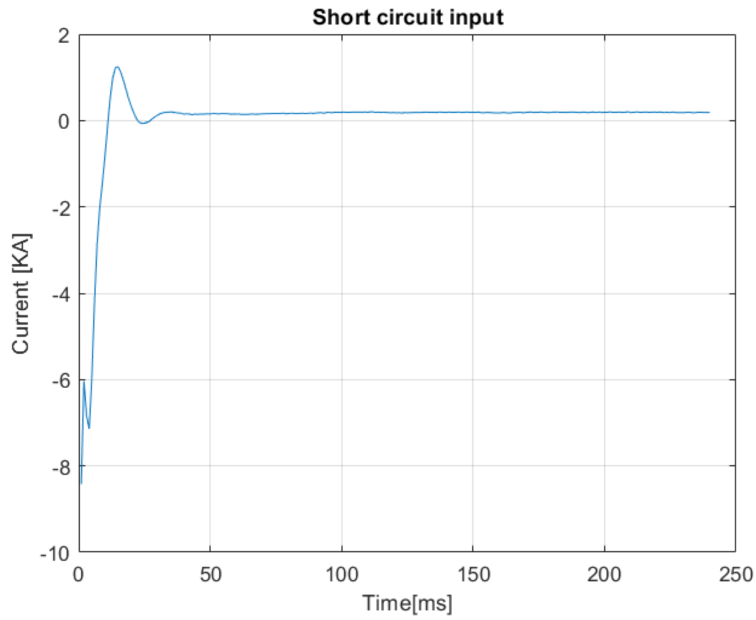

Figure 28: Example input in a short circuit welding point

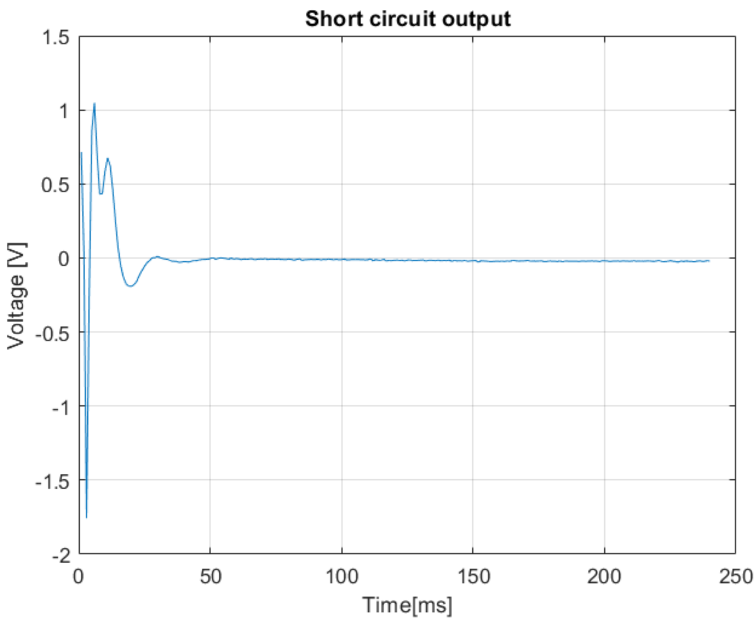

Figure 29: Example output in a short circuit welding point

It is possible to see that the welding point last $240ms$.
The first $40ms$ are considered "blanking time" because they are really noisy and, possibly, they don't bring helpful informations. For this reason, in the evaluation of the model, they will be not considered.

2.7 Polynomial models

With the selected data, different kind of ARX, ARMAX and OE models have been tried. The M set of possible models has been chosen always with a possible delay from 1 to 5, and every single term with an order from 1 to 5 too.

Model	Order	Delay
ARX	$n_a = n_b = 1$ to 5	n_k =1 to 5
ARMAX	$n_a = n_b = n_c = 1$ to 5	n_k =1 to 5
OE	$n_b = n_f = 1$ to 5	n_k =1 to 5

Table 3: ARX, ARMAX and OE orders and delays

n_k is called delay because is the first useful time instant of the input so that the equation, for example in the ARX model, become:

$$y(t) = -a_1 y(t-1) - ... - a_{n_a} y(t-n_a) + b_1 u(t-n_k) + ... + b_{n_b} u(t-n_b-nk+1) + e(t) \quad (17)$$

and the same for ARMAX and OE models.
The best models are firstly evaluated according to 99% auto-correlation region.

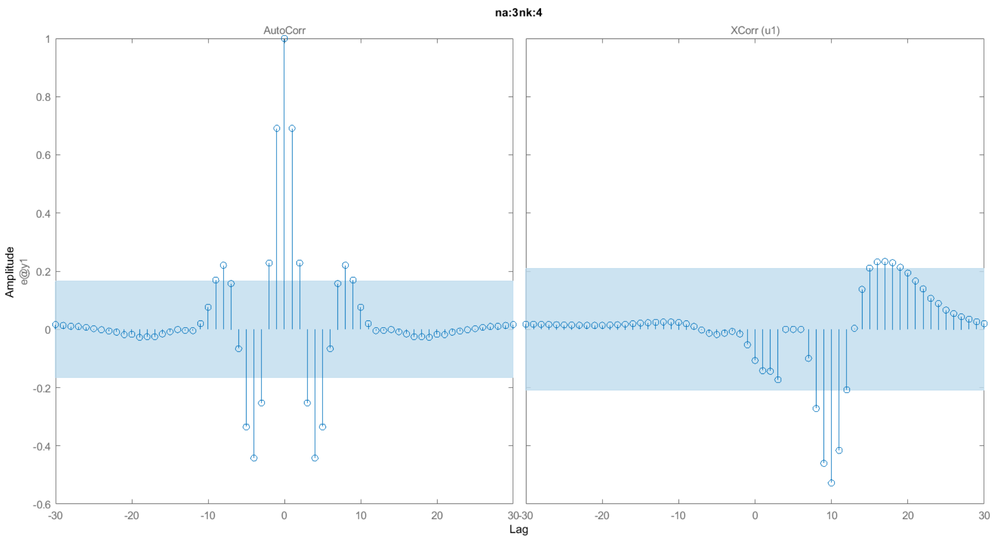

Figure 30: Auto-correlation and cross-correlation analysis

The choice of the best models is made evaluating the number of residues out of the 99% confidence zone in the right half plane (due to plot symmetry) of the auto-correlation.
In the Figure 30, for examples, it is possible to visualize the auto-correlation and cross-correlation function of the arx model with $n_a = 3$, $n_b = 3$ and $n_k = 4$. In the research of the polynomial model has been chosen a number of maximum residues out of the 99%

confidence region equal or less than 3. For this reason the presented example do not pass the auto-correlation analysis and its RMSE will not be considered in the final evaluation. Furthermore, 2 cases were considered:

- current input and voltage output
- current and squared current inputs and voltage output

As example, below are reported some of the best ARX models with double input and their auto-correlation analysis:

n_a	n_k	Resid
1	1	1
3	1	1
2	5	1
5	4	1
4	4	2
5	3	1
4	3	1
3	2	2
5	5	1

Table 4: ARX residues analysis results

Resid stands for the number of residues outside the 99% confidence region, while n_a is the order reference and n_k is the delay reference .
All the model reported are good choice to polynomial identification, the best between this model will be then selected with RMSE analysis.
The MSE is defined as :

$$\mathbf{MSE} = \frac{1}{N - N_0} \sum_{N_0+1}^{N} (y(t) - \hat{y}(t, \theta))^2 \tag{18}$$

The RMSE is defined as :

$$\mathbf{RMSE} = \sqrt{MSE} \tag{19}$$

Where $\hat{y}(t, \theta)$ is the output valued with the estimated model. The Root Mean Square Error has not a global meaning, it is just an index of the fitting of the curve. The RMSE represents the square root of the second sample moment of the differences between predicted values and observed values or the quadratic mean of these differences.The RMSE serves to aggregate the magnitudes of the errors in predictions for various data points into a single measure of predictive power. In the following plot is possible to see how, basically, increasing the order, the RMSE tends to decrease. The goal is to choose the lower RMSE between the models that passed the residual analysis.

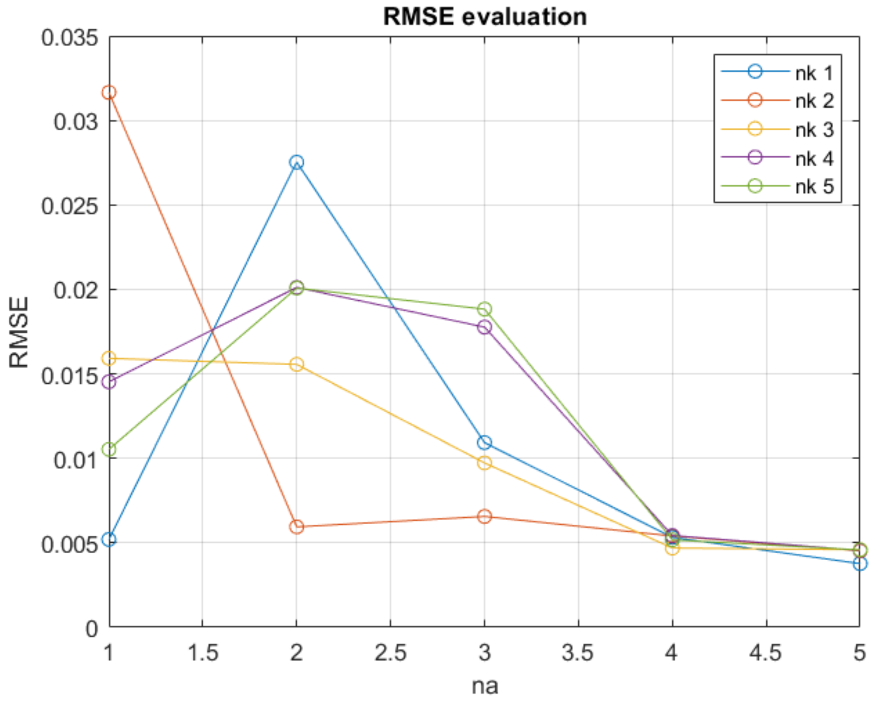

Figure 31: RMSE analysis of ARX model

It is possible to see that the ARX model with $n_a = 5$ and $n_k = 5$ has the lower possible RMSE (0.0026). For this reason the ARX(5,5,5) (ARX(n_a,n_b,n_k)) is the best possible model in the ARX set.
At this point the real output and the estimated one is compared due to the Best Fit.

$$\textbf{Best Fit} = 1 - \sqrt{\frac{MSE}{\frac{1}{N-N_0}\sum_{t=N_0+1}^{N}(y(t)-\hat{y})^2}} \tag{20}$$

It can be easily converted in percentage.
At the end of this process, the result is a polynomial model able to produce the same output of the system if it is fed with the same input. This model is just a mathematical representation and not necessarily has physical meaning.

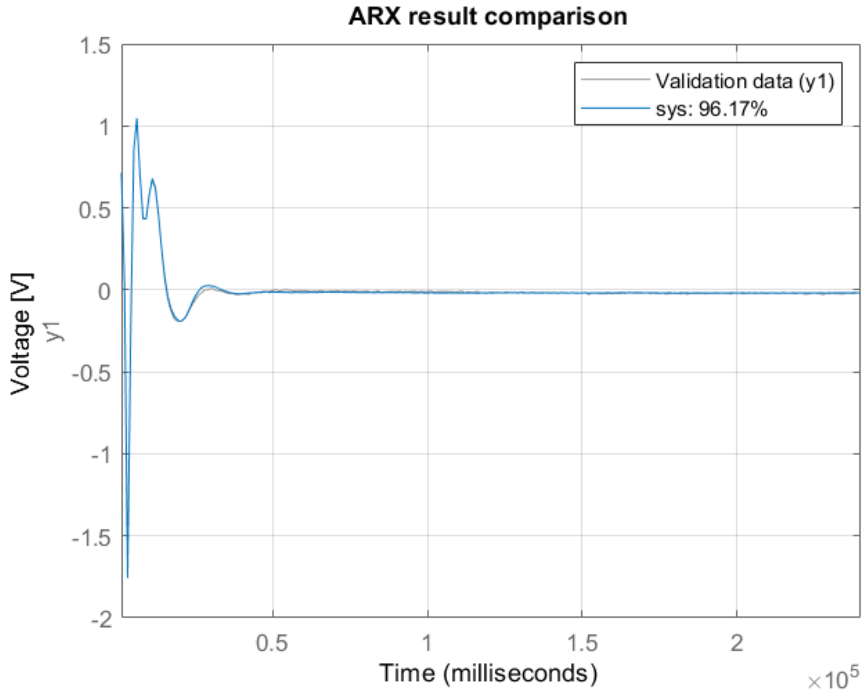

Figure 32: Measured and estimated output comparison

The same procedure can be iterated for the ARMAX and the OE model.
The results are:

Model	C	C and C^2
ARX	$n_a = n_b = 5$ and $nk = 2$	$n_a = n_b = 5$ and $nk = 5$
ARMAX	$n_a = n_b = n_c = 4$ and $n_k = 3$	$n_a = n_b = n_c = 5$ and $n_k = 3$
OE	$n_b = n_f = 4$ and $n_k = 3$	$n_b = n_f = 4$ and $n_k = 3$

Table 5: ARX, ARMAX and OE orders and delays for both input conditions

One thing to specify is that, in the situation with current and its squared value as input, the coefficient relatives to the inputs have to be doubled.
The Multiple Input Single Output (MISO) system has to be considered as the superposition of two Single Input Single Output (SISO) systems.
To report an example, the ARMAX MISO system will be of the type ARMAX$(n_a, n_{b1}, n_{b2}, n_c, n_{k1}, n_{k2})$ where n_{b1} and n_{k1} are related to the first input (C) while n_{b2} and n_{k2} to the second one (C^2). As simplification, has always been considered $n_{b1} = n_{b2}$ and $n_{k1} = n_{k2}$.
The obtained results are encouraging.
The data set was previously divided in 70% for the estimation of the model and a 30% for the validation.
The results during validation show a floating fitting between 80% and 95%.

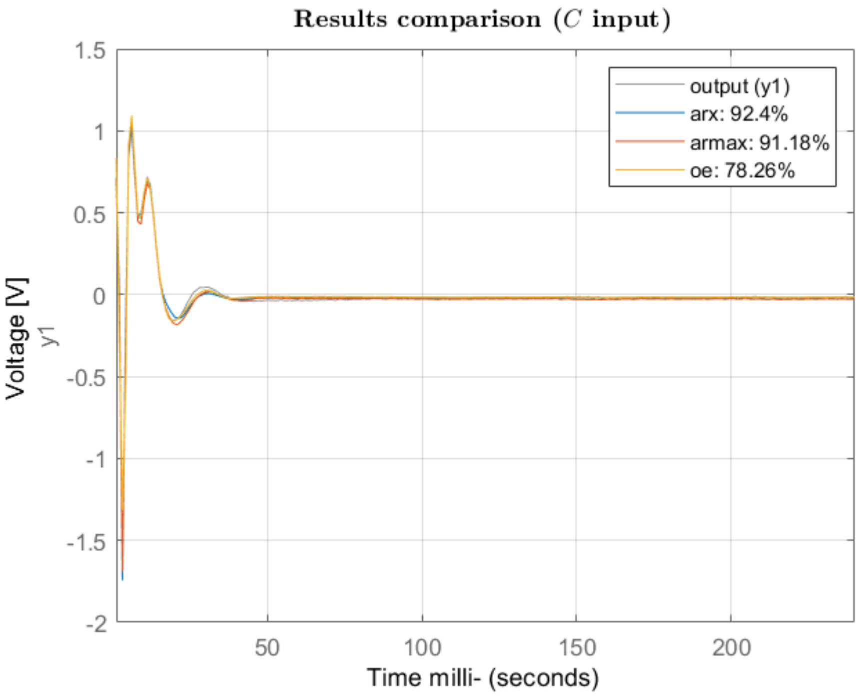

Figure 33: Measured and estimated output (single input models)

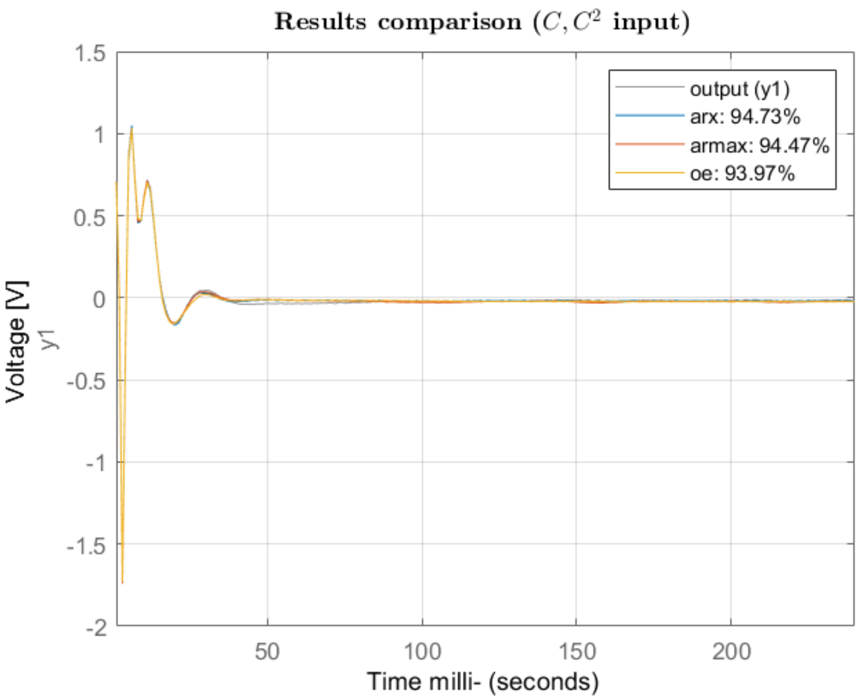

Figure 34: Measured and estimated output (double input models)

It is possible to see that, on average, the results obtained with the double input (C, C^2)

are better than the single input cases.
This shows a possible physical relationship between the output voltage and the power of the system ($P = I^2 R$).
Another attempt was also made, considering as input, also the cube value of the current. In this case the fitting percentage did not increase significantly as the C^2 case.
A possible meaning of this phenomenon is that physically has no meaning the cube of the current.

3 Physical model

From the physical point of view, the main role in the welding process is played by the Joule effect. This phenomenon is defined as the process by which the passage of an electric current through a conductor produces heat. The power of heating generated by an electrical conductor is proportional to the product of its resistance and the square of the current. When two metal sheets are put in touch and compressed, the highest electric resistance is right at the contact point. So, during the welding process, the higher power of heating is released in this point and here the metal starts melting.

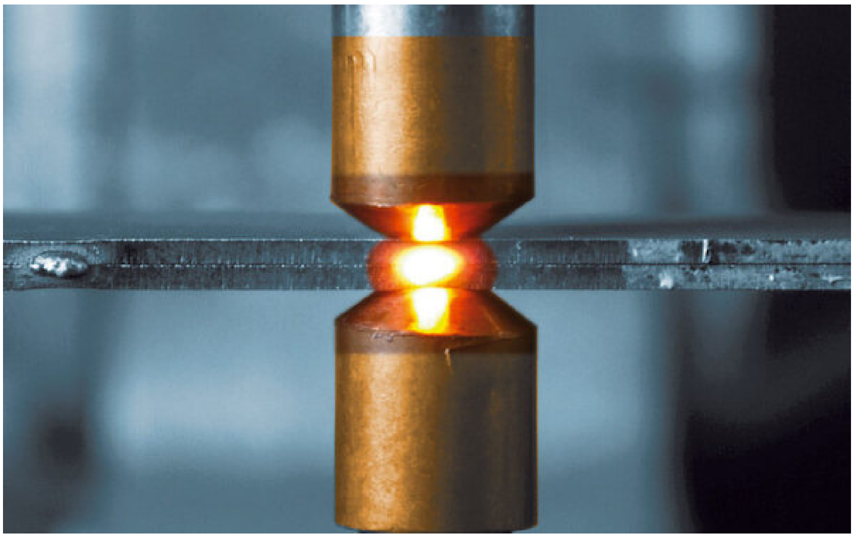

Figure 35: Weld core of metal sheets during a welding process

It is important that the welding process last long enough to allow an effective mixing of the melted metal of the sheets. In fact, a brief release of current can lead to a 'gluing' of the sheets rather than a welding. On the other hand, a longer one lead to a possible expansion of the welding core out of the contact zone with the electrodes, causing splash or a less controlled welding. Those reasons make explicit the importance of the power involved during the process. The ISI-Welding Company evaluates the power as an index of the quality of the point and applies also a control based on the released power during the welding process: if the energy is too low, it makes the process last longer; if it is too high, it makes it stops prematurely.
The dependence of the welding process by the power was also suggested by the previous studies on the polynomial models, since these models perform better with the information on the square of the current (directly proportional to the power). In general, the knowledge on the spot welding states that the resistance spot welding has to be seen as an **electric**, **mechanic** and **thermal** phenomenon at the same time[8].

- **Electrical conductivity:** current, necessary to generate a certain quantity of heat, increases with increasing conductivity.

- **Thermal conductivity:** current necessary to compensate the heat dissipated by conduction increases with increasing thermal conductivity.
- **Thermal expansion coefficient:** higher current value and shorter time are needed to avoid the expulsion of the melted core and guarantee sufficient heat at the same time[9].

In the last point it emerges that electrical quantities are also greatly influenced from the metal fusion and its consequent cooling. For those reasons, physically, there is the necessity to investigate the dependency between the current flowing into the circuit, energy variation and resistance variation. Unfortunately power and resistance data are not measured and it is hard to identify a precise data-driven model.
Our system can be classified as a **grey-box**: this is an approach that combines a partial theoretical structure with data to complete the model.
It is known that the voltage depends on the current, its derivative for inductive phenomena and its integer value for capacitive phenomena. Moreover, it depends also on the power and its derivative (to take into account the thermal effects), on the resistance and its derivative (to take into account the thermal expansion and its change of state).
Once the theoretical basis have been laid and a large dataset is made available, a data-driven methodology has to be chosen to implement the physical model. The method of Least Squares is a standard approach in regression analysis to approximate the solution of overdetermined systems (sets of equations in which there are more equations than unknowns) by minimizing the sum of the squares of the residuals made in the results of every single equation. A residual is defined as the difference between the actual value of the dependent variable and the value predicted by the model:

$$r_i = y_i - f(x_i, \theta) \tag{21}$$

The most important application is in data fitting. Least-squares problems can belong to two categories: linear and nonlinear least squares, depending on the linearity of the residuals. The linear least-squares problem occurs in statistical regression analysis, the nonlinear problem is usually solved by iterative refinement in which at each iteration the system is approximated by a linear one, and the calculation is similar in both cases.

The only input available is the current. A first basic model can be designed considering the voltage as dependent only on the current and its square value, for the proportionality with the power. This equation represent the mathematical model:

$$V(t) = \theta_1 I(t) + \theta_2 I(t)^2 \tag{22}$$

This equation can be arranged in matrix form:

$$y = X\theta \tag{23}$$

Where y is an $n - by - 1$ vector of responses (voltage values), X is the $n - by - m$ design matrix for the model (current and its square), θ is the $m - by - 1$ vector of parameters to identify. The least square method makes use of the pseudo-inverse of a matrix :

$$A^{+} = A^{T}(A^{T}A)^{-1} \tag{24}$$

This formula is a generalization of the inverse of a matrix in case this is not squared. Taking up the matrix equation of the model, with some mathematical tricks and using the pseudo-inverse the following steps can be performed:

$$\begin{gathered} X^{T}X\theta = X^{T}y \\ \downarrow \\ \theta = (X^{T}X)^{-1}X^{T}y \end{gathered} \tag{25}$$

In this way it is possible to compute θ from the collected inputs-outputs.
The first chosen dataset is the one with short-circuit point data. This is the simplest available,it is possible to obtain the first useful information from this model and at a later stage other complex models that take into account further factors can be implemented and checked. The matrix of parameters θ is therefore computed using the input-output set of data selected, and a test on a point is performed:

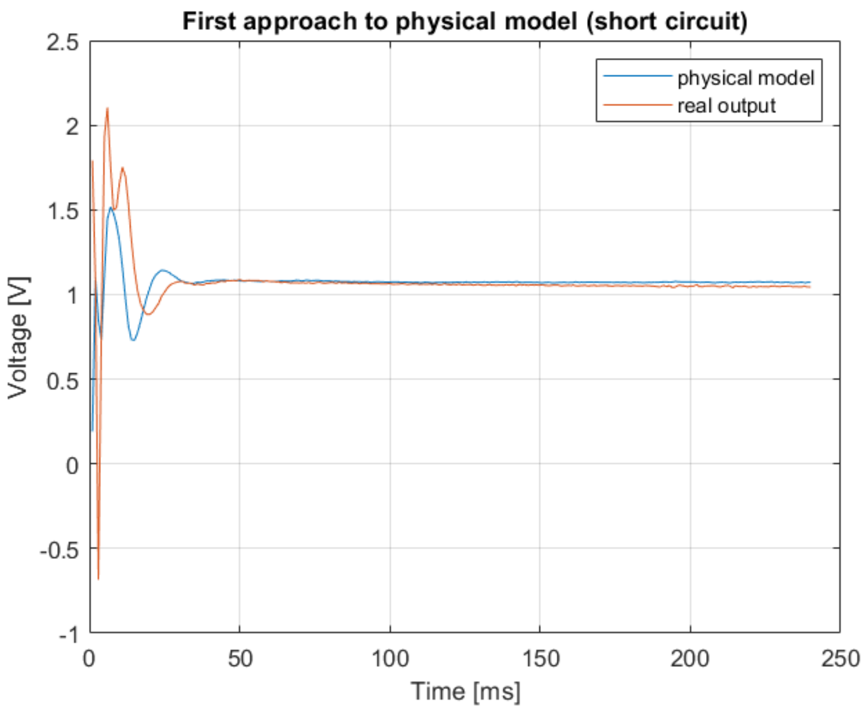

Figure 36: Comparison between estimated and real model (short circuit case)

Results show that the estimated model is not very able to approximate the real one for the entire duration of the signal.
Another attempt with this simple model is performed using a dataset of real welding spots. The new parameters are computed and another test is executed:

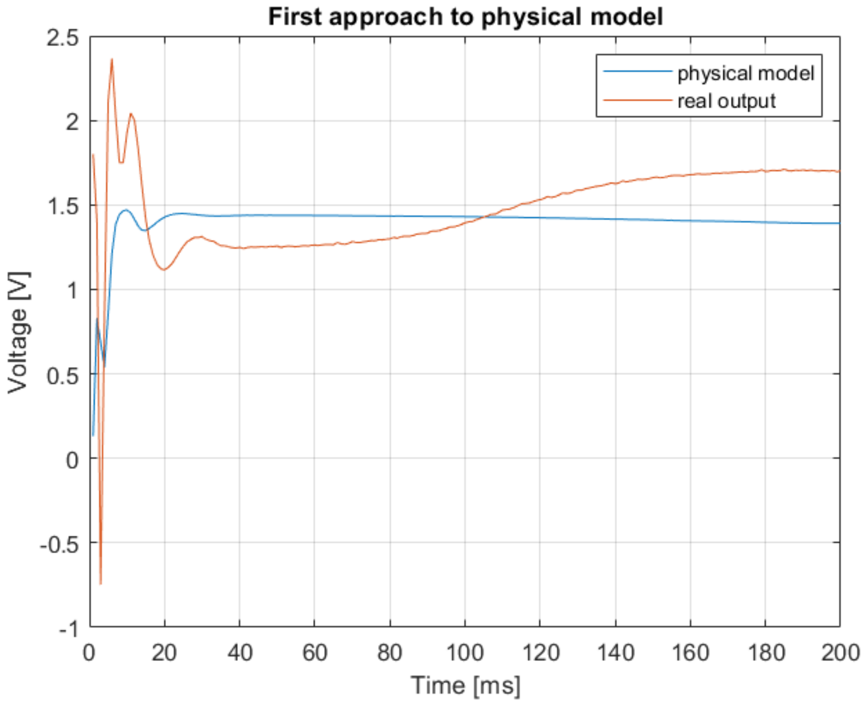

Figure 37: Comparison between estimated and real model (real welding case)

As expected, in this case the estimation is not acceptable. There is a simple reason: a lot of factors have not been intentionally considered, such as thermal and conductivity phenomena, the change of resistance due to the melting of the metal, capacitive and inductive phenomena that can arise in an electric circuit. All these factors in some way depends on the current. The following equation collects various shapes of the current which can be able to take in account these phenomena that are not directly measured. According to the Least Square method, every variable has been associated to a θ parameter.

$$V(t) = \theta_1 I(t) + \theta_2 I(t)^2 + \theta_3 \frac{dI(t)}{dt} \frac{1}{I(t)^2} + \theta_4 \frac{dI(t)}{dt} + \theta_5 \int I(t)dt \tag{26}$$

Moving over to the matrix, the equation does not change with respect to the previous one, what changes are the X and θ dimensions and the computational load will be higher. Of course real welding points data are the best choice for this kind of model built on assumptions that implies the presence of metal sheets during the welding process. Subsequently, θ parameters have been computed and analyzed. The parameter associated to the integral of the current is very small and can be neglected. This can mean that there are no capacitive phenomena involved in the welding process. As last step, a test is executed:

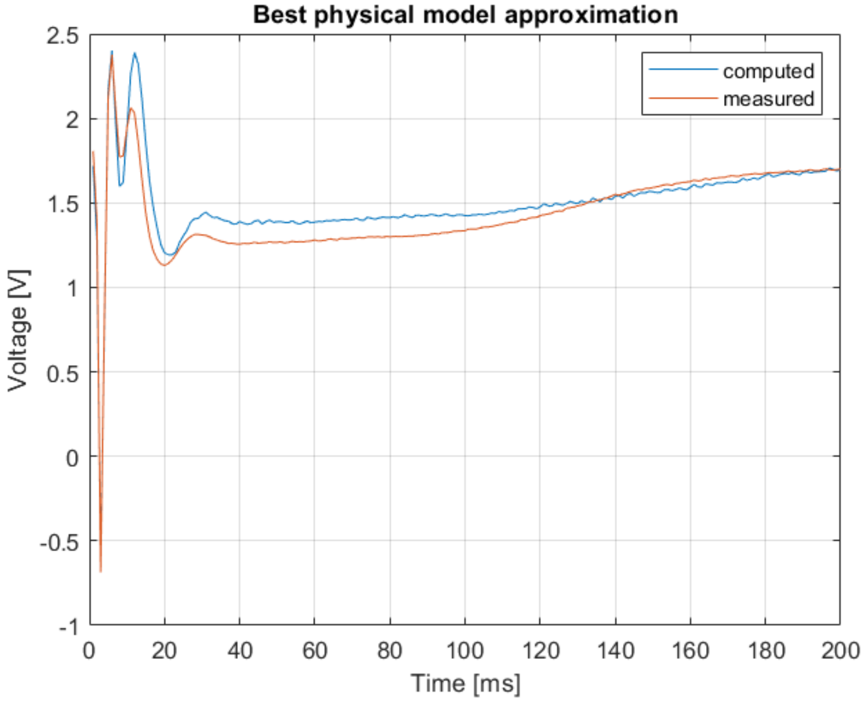

Figure 38: Data obtained by the physical model compared with real one

Results shows that the estimated model is able to roughly approximate the real output. Better performance can be achieved by measuring other important variables (that can be compression force, temperature, etc) and including them in the model.

4 The mean challenge

The main aim of this work is to find a reliable predictive maintenance system for electrode dressing. Actually an electrode is dressed cyclically after 100 welding points, unless an human operator schedules it before the due date (for example if a critical condition occurs). The dressing is a milling procedure to reshape the electrode and to give it newly the electrical qualities lost due to worn out.
For this reason it is a critical point to define in advance when the dressing is needed and when it is not, this procedure can save money and time.

4.1 Work chain and dressing

A production line has several machines in series, each of them will have a series of different welding programs that will always be done in the same order (for convenience we will call the points programmed on a single machine cycle). The programs are different welding points with different duration and different current and voltage requirements. Each machine keeps count of the points made because, without further needing, after 100 welding points the dressing is programmed. During the various welding points, the electrodes pollute themselves with the molten metal of the sheets, changing shape and technical quality, this phenomenon do not allow to have control over the quality of the stitch. The dressing is a procedure to restore the electrodes original shape by milling, in order to bring it back, as much as possible, to the nominal conditions. The dressing is also applied as a precaution to new electrodes to make sure that they have the necessary shape for the welding points foreseen by the machine. A shorted welding spot is performed before and after each dressing, i.e. without any plate to be welded between the electrodes. These points are taken before and after the dressing to track the changes of the electrodes resistance, the fact that they are short-circuited is intended to eliminate uncertainties and the various disturbances introduced by the metal sheets. Another consideration is that, if the machine operator notices a particular worn out in electrodes (and the quality of the welding point is no more acceptable), he can arrange a dressing before reaching the 100th welding point. Of course, the short-circuited points will also be scored in this case before and after dressing. Finally, even if the dressing is foreseen every 100 welding points, if the 100th point happens in the middle of the cycle, the machine will finish the remaining programs first and then the process necessary for dressing will begin.

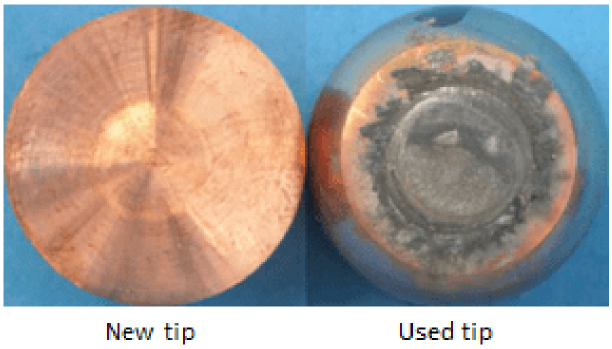

Figure 39: Example of used and new electrode

4.2 Short circuit point

After a first phase of data filtering and normalization, the first attempt is made using the short circuit points[10]. As it has been previously explained, the short circuit points have been taken in order to keep track of changes in electrode resistance during welding points.

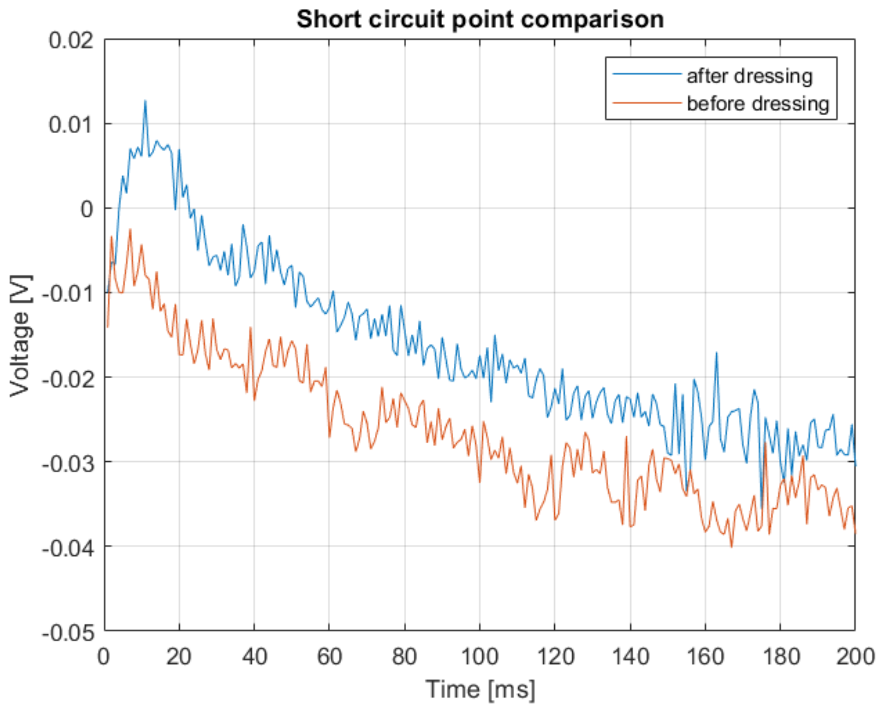

Figure 40: Comparison between the short circuit points before and after dressing

Every time a short circuit point is taken before and after dressing. As it possible to see in the figure, after the dressing the voltage is lower (in modulus) then in the point before the dressing.
This means that, possibly, there might be an increase of electrodes resistance welding by welding.

For this reason, a first kind of machine learning algorithm has been tried: a simple 1 layer neural network, to test if this kind of situation can be easily recognized.
This is an hard simplification of the problem but can give important tips.
This predictive maintenance problem has been treated as a classification problem where the welding machine under analysis can be seen as '**safe 0**' or '**unsafe 1**'.
The results are really encouraging because, in test phase, the algorithm is able to recognize the 100% of the situations.

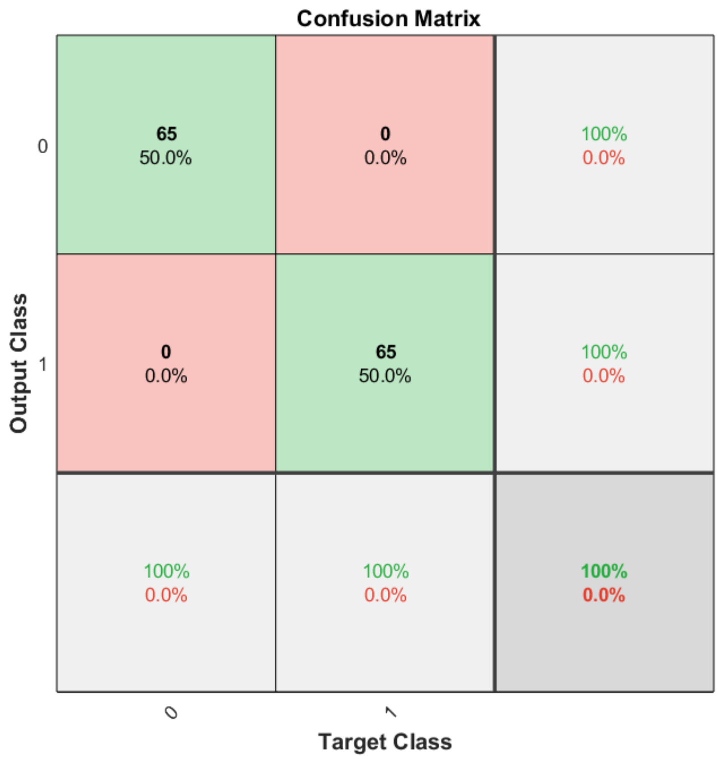

Figure 41: Confusion plot of the overall results (train,validation,test)

A possible limit of this test is the relative small data set, only 65 points available, but it is just a first step to clarify how classification problems and machine learning algorithms work and their potential.

4.3 Second step

The main problem that this thesis proposes to solve is to find a predictive maintenance system able to predict the dressing of electrodes while the machine normally operates.
Unfortunately a great part of the uncertainty is introduced by the metal sheets that must be welded together.

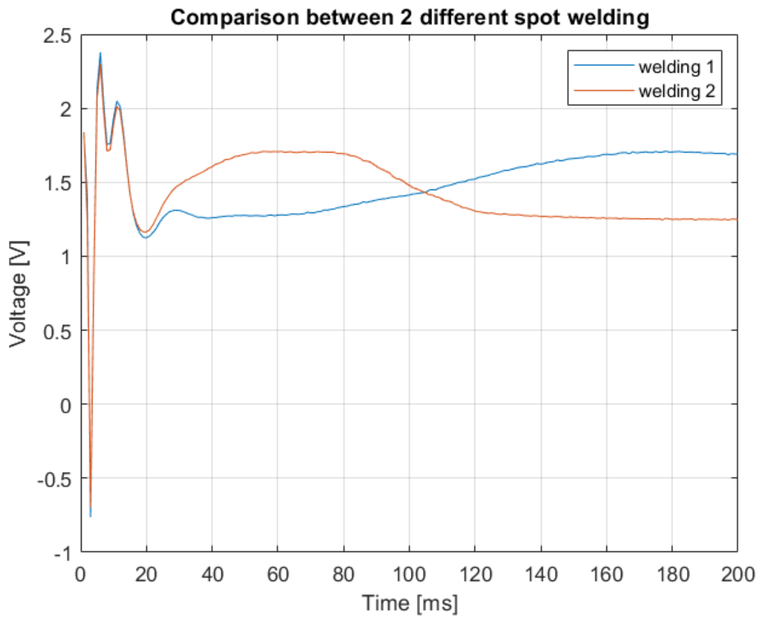

Figure 42: Comparison between 2 different spot welding

In the figure are represented the data measured in two different welding point carried out by the same machine performing the same program: practically they are 2 identical welding points, where in theory, also the metal sheet interposed between the electrodes is the same.
As it is possible to see, the results are completely different and that is why is so difficult to carry out a study on the parameters that can identify an electrode at the end of its life cycle with the measured data.
For this reason the use of neural networks was considered more appropriate.
In this second attempt, different algorithms of neural network are tried in order to identify the critical situation before the dressing.
The input are the measured data millisecond by millisecond without the first $40ms$ (blanking time).
For the classification problem is considered as **unsafe (1)** the 100th welding point before the electrode dressing, moreover the training is done just on one program of one machine. That is not to be considered as simplification of the problem because, as explained before, even if the electrodes is to be changed the machine will end first the cycle of planned programs and then it will approach the dressing procedure. For this reason it is possible to identify the critical condition in the last program of the cycle sending right after the welding guns to revive the electrodes.
This is the strategy that, from now on, is applied: the machine learning algorithm are applied only at the last welding point of a cycle.
In future, in order to make this approach more reliable, the algorithms can be applied at the last 2 or 3 programs to be sure that if one fails, it is possible to rely on the others.
The best results with this approach are reached with neural network algorithms. The network used are simple 1 layer neural network with 10 or 100 neurons in order to study

how, increasing the complexity of the net, the results will change.

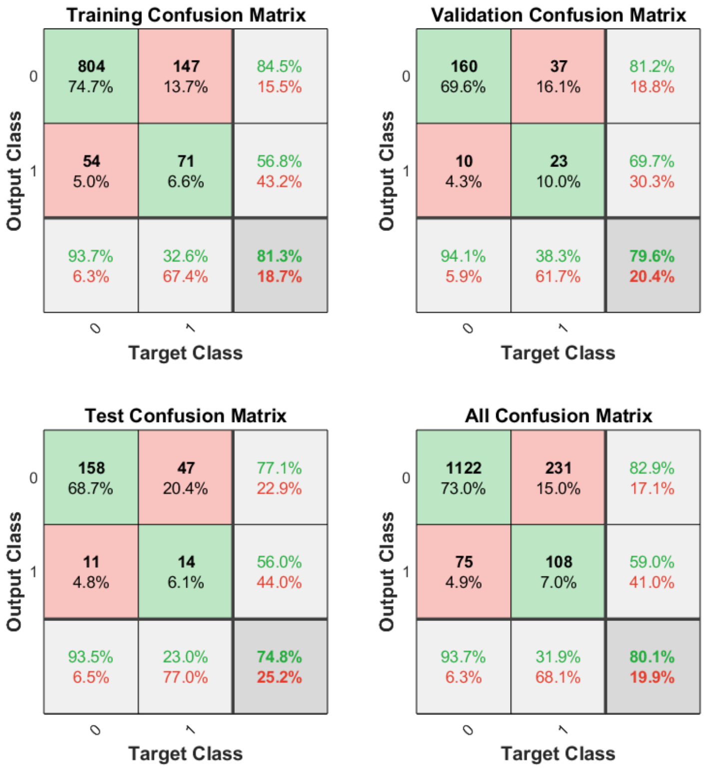

Figure 43: ANN results with 10 neurons

The results are encouraging but not good enough to entrust the maintenance planning to this system.
In a particular way, the identification of the unsafe situation (1) has a precision which fluctuates between 50 and 70 percent.
In the next image are represented the training, validation and test results for a neural network with 100 neurons in its hidden layer.

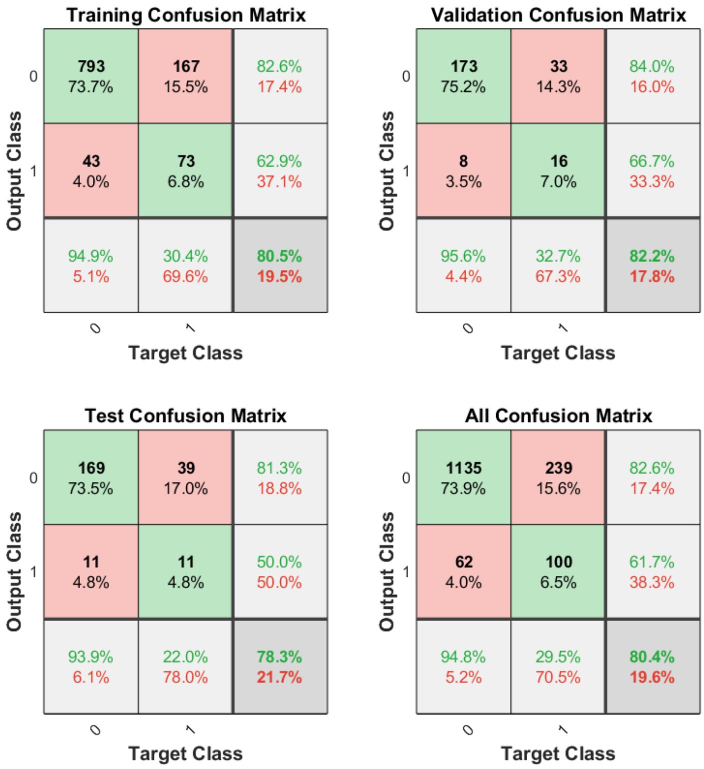

Figure 44: ANN results with 100 neurons

It is possible to see that increasing the complexity of the net the results are not better. Rather than further complicating the network, adding more layer or trying a long-short term memory approach, in the next steps will be analyzed another approach.

4.4 Data evidence

The next step arises from some observation that have been done studying the data sets. The only measured data are current and voltage (input/output), and the resistance can be computed as $R = \frac{V}{I}$. In the following image it is shown the trend of the computed resistance of 6 welding points (from a dressing to another).

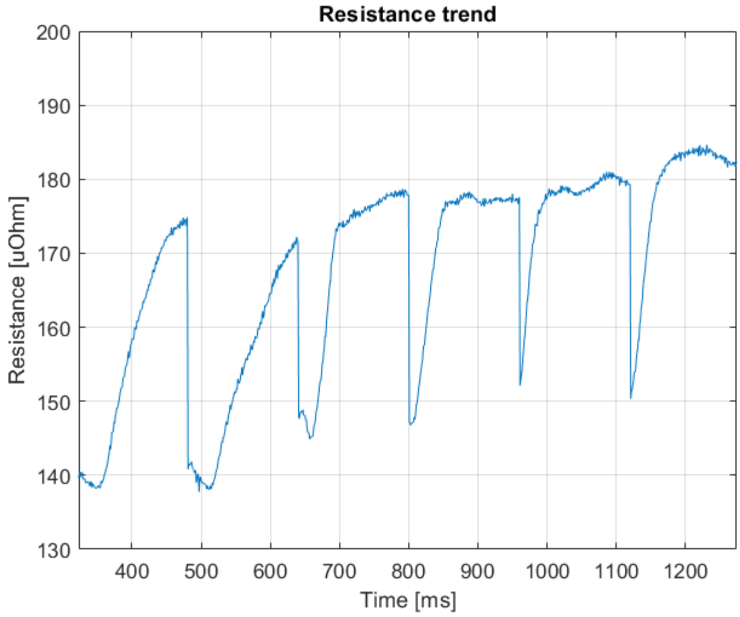

Figure 45: Resistance trend during different welding points

In the figure the welding points are represented without the first $40ms$ of blanking so that the time sequence is just symbolic. The points are surely consecutive in time and made from the same machine performing the same program but from one point to another there are several seconds of stop.
The first observation is that the value of the resistance increases with time and consecutive points.

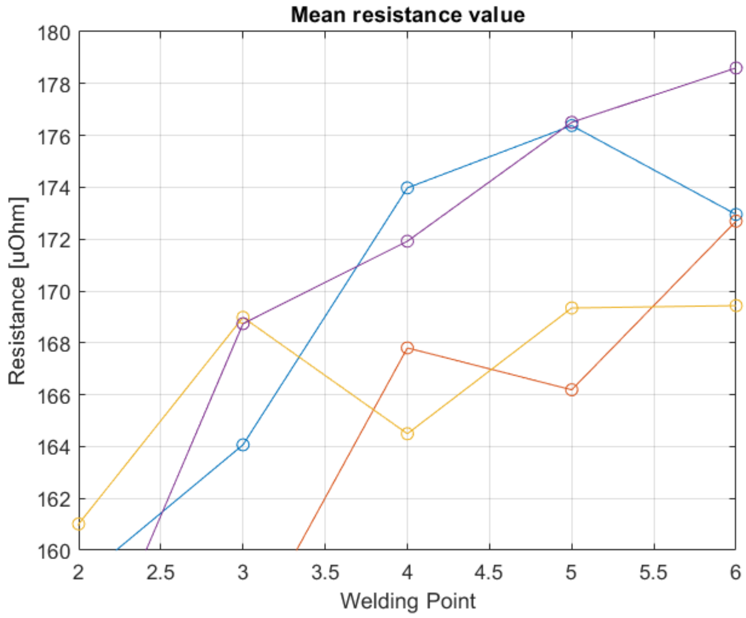

Figure 46: Resistance mean value in different cycle

In the Figure 46 are represented the mean value of the resistance of different welding points. Welding points with the same color belong to the same cycle.
It is possible to see how, generally, the mean value of the resistance tends to increase point after point.
Another observation, made on resistance trend, can be highlighted : in addiction to increasing the mean value, the resistance trend also changes the shape of the curve. In the resistance trend is possible to see how in the first point after the dressing, during the entire welding point the resistance keeps increasing its value. Going on with the points, the behavior of the curve becomes more similar to a transient with steady-state immediately after.

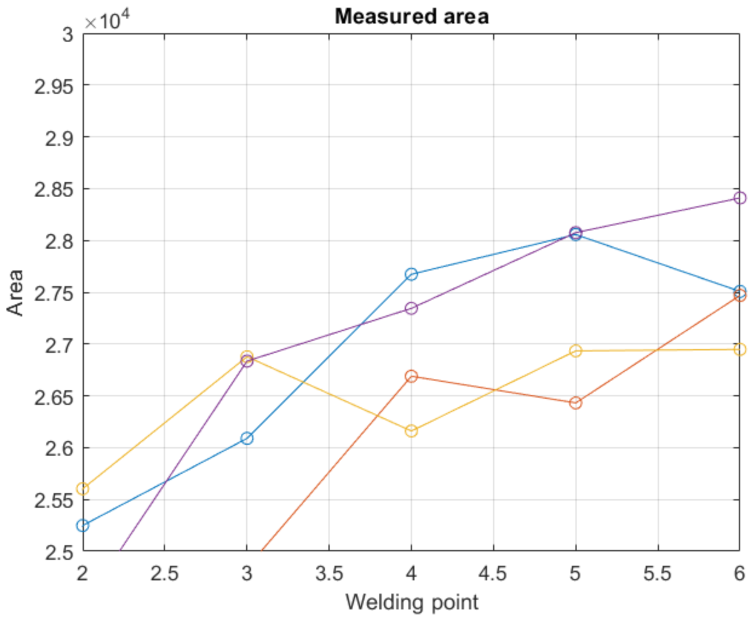

Figure 47: Resistance area in different cycle

The previous observation has been resumed in the computation of the area of every single curve.
It is possible to see that point after point the subtended area generally increases.
For this reason the next step, rather than giving to machine learning algorithm all the measured data, will try to give indicators that can highlight these changes.

4.4.1 Statistical descriptors

A descriptive statistic is a summary statistic that quantitatively describes or summarizes features from a collection of information, while descriptive statistics (in the mass noun sense) is the process of using and analyzing those statistics. Statistical descriptors are indices that can precisely and synthetically describe a set of data. They belongs to the field of the descriptive statistics and can be used both with continue and discrete variables. They can be divided into:

- position indices:
  - mode,
  - median,
  - mean.
- dispersion indices:
  - standard deviation,
  - variance.
- shape indices:
  - skewness,
  - kurtosis.

Position indices (also known as central trend measures) identify, in different ways, the central element of the distribution.
Dispersion indices evaluate how much data deviates from the central element of the distribution.
Finally, shape indices consider the shape of the distribution with respect to a normal (or Gaussian) distribution. In particular, skewness how much the distribution is asymmetric and kurtosis how it is flat. Mathematically, the skewness is defined as:

$$skewness = \frac{m_3}{m_3^{3/2}} \tag{27}$$

while kurtosis:

$$kurtosis = \frac{m_4}{m_2^2} \tag{28}$$

where m_k is the central moment of order k:

$$m_k = \sum_{i=1}^{n} (x_i - \mu)^k \tag{29}$$

x_i are the values of the dataset and μ is the mean value.

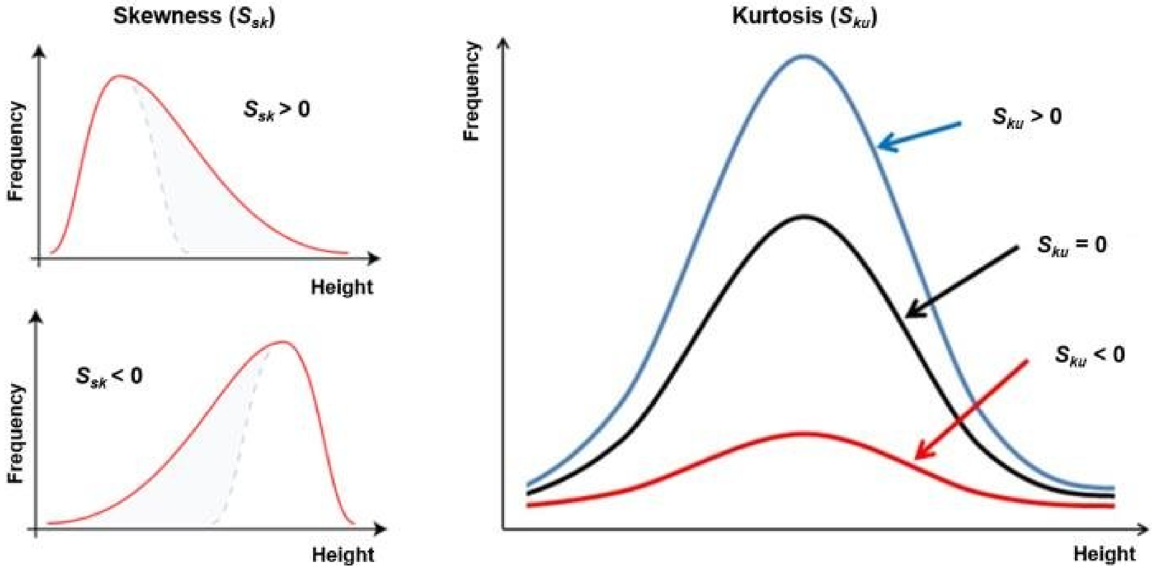

Figure 48: Skewness and kurtosis graphically explained

So, these indices are able to describe the distribution and can be used instead of taking in consideration all the measured values. A new neural network has been implemented with a new set of inputs. This inputs include:

- statistical descriptors: mean, mode, median, standard deviation, $3th$ order and $4th$ order central moment, skewness, kurtosis, max and min value of a welding point.
- other indices automatically computed from the company software: initial and final resistance values, instants of max and min values, rise and descent time, growth and degrowth rate.

4.5 Final solution

From now on the machine learning algorithm are trained on the statistical descriptors in order to give them a precise description of the welding rather then all the measured data. In this situation, for the algorithm is easier to get how the shape of the curve changes from a welding to another.
The statistical descriptor used are:

- mode
- median
- mean
- skewness
- kurtosis
- variance
- 3^{th} order moment
- 4^{th} order moment
- max value
- min value
- variance

Starting from the knowledge assumed during the modeling phase, it makes sense to compute these indexes for all the quantities included in the welding phenomenon as the resistance and its derivative or the energy. Moreover there are 4 more indexes that are automatically computed from the welding quality system of the ISI welding and are:

- min time
- max time
- growth rate
- degrowth rate

These indexes are computed on the voltage output and they also are used as descriptors of the behavior of the welding point. All the indexes seen until now will be computed after the first $40ms$ called blanking time. Several attempts will be done in order to try different configuration and combination of data to find which one gives the best results.
The best results are reached with the indexes computed for measured voltage, measured current, computed resistance and its derivative. Moreover, due to the fact that the algorithm better recognize a critical situation if has some information about the 'story' of the machine, the input will be the statistical description of the actual welding point and the previous one.
The first algorithm applied is a simple one layer neural network. The neural network is trained due to back propagation.
The term back propagation strictly refers only to the algorithm for computing the gradient, not how the gradient is used, in particular in the case of 1 layer neural network the algorithm applied is the delta rule.
The delta rule is a gradient descent learning rule for updating the weights of the inputs to artificial neurons in a single-layer neural network.
For a neuron j with activation function $g(x)$, the delta rule for j's ith weight w_{ij} is given by:

$$\Delta w_{ij} = \alpha(t_j - y_j)g'(h_j)x_i \tag{30}$$

where

- α is a small constant called learning rate
- $g(x)$ is the neuron's activation function
- g' is the derivative of g
- t_j is the target output
- h_j is the weighted sum of neuron's inputs
- y_j is the actual output
- x_i is the ith output

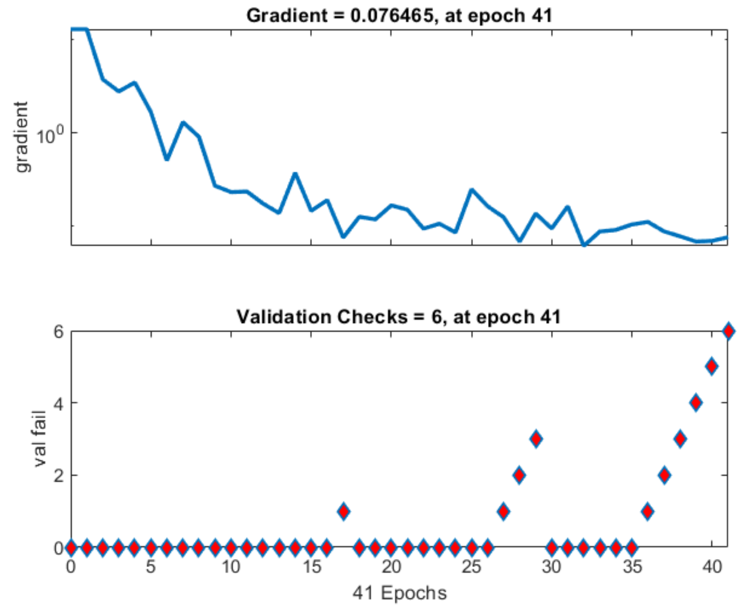

Figure 49: 1 layer neural network gradient

In the Figure 49, the gradient trend and the validation for each epoch are proposed.
An epoch, in terms of neural-based training, is one full pass through the data-set.
The gradient descent algorithm will stop if, after that for 6 consecutive checks, the gradient stops decreasing.
In the second image of Figure 49 it is possible to see that from the epoch 35 from 41 the gradient stops to decrease and the algorithm stops to look for a smaller one.

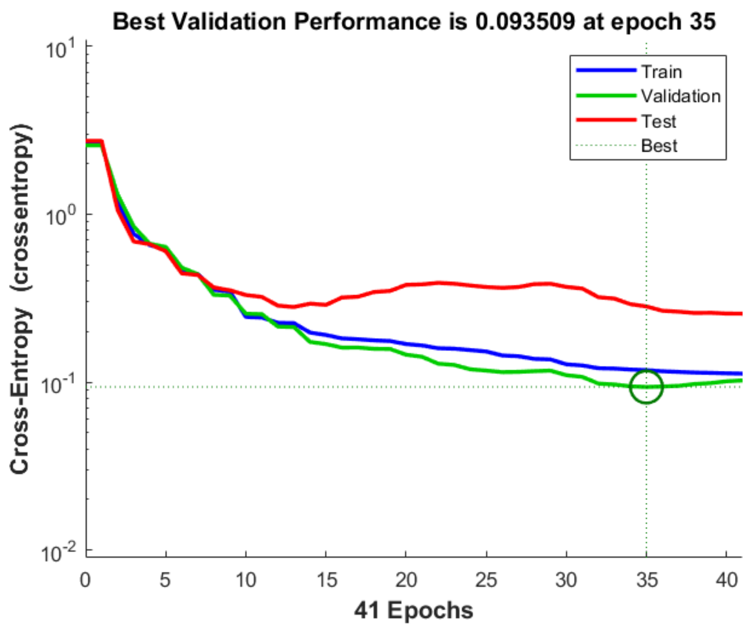

Figure 50: Validation performance

In the Figure 50 it is shown a validation of the performance using the Cross-Entropy. Cross-entropy loss measures the performance of a classification model whose output is a probability value between 0 and 1. Cross-entropy loss increases as the predicted probability diverges from the actual label. So predicting a probability of .012 when the actual observation label is 1 would be bad and result in a high loss value.
In binary classification, where the number of classes M equals 2, cross-entropy can be calculated as:

$$H = -(y\log(p) + (1-y)\log(1-p)) \tag{31}$$

Where y is the class label and p the predicted probability.

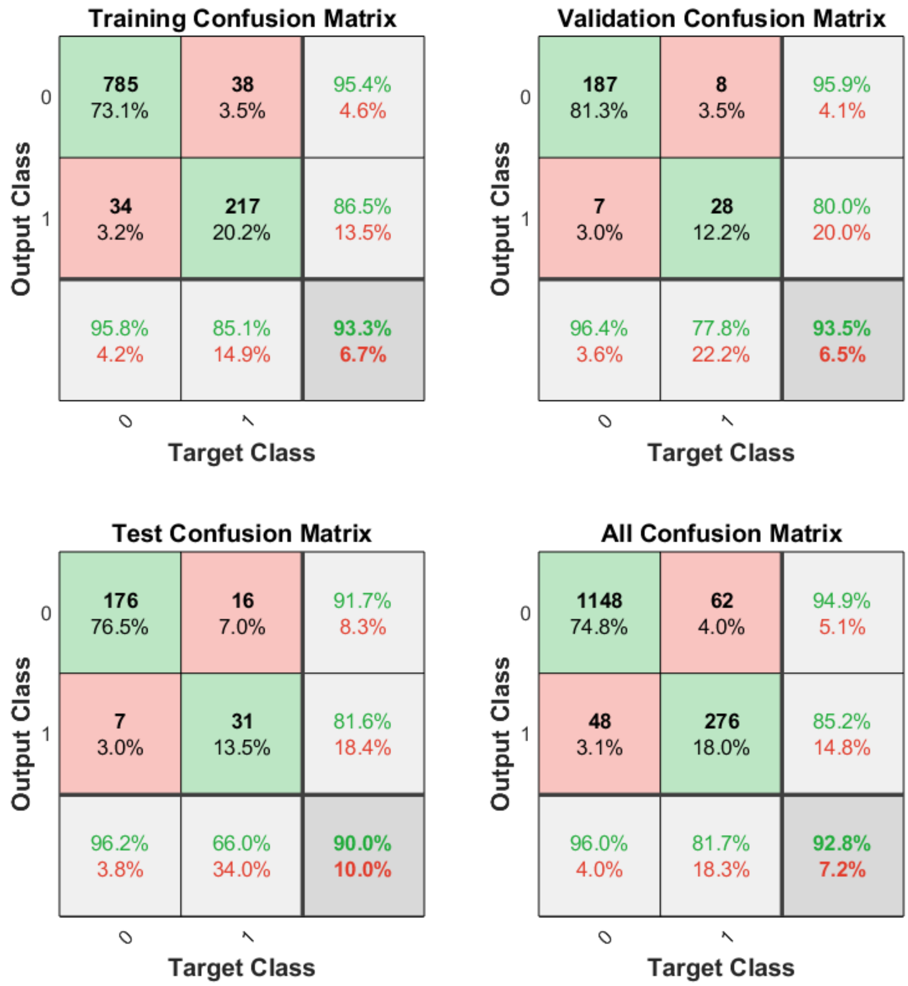

Figure 51: 1 layer neural network results

As always the 1 symbolizes that the electrodes need to be dressed. The results are really encouraging in testing phase.

Better results are reached with a multi-layer neural network. In this case the complexity is increased due to the presence of 3 layers of 1000 neurons and an LSTM layer.

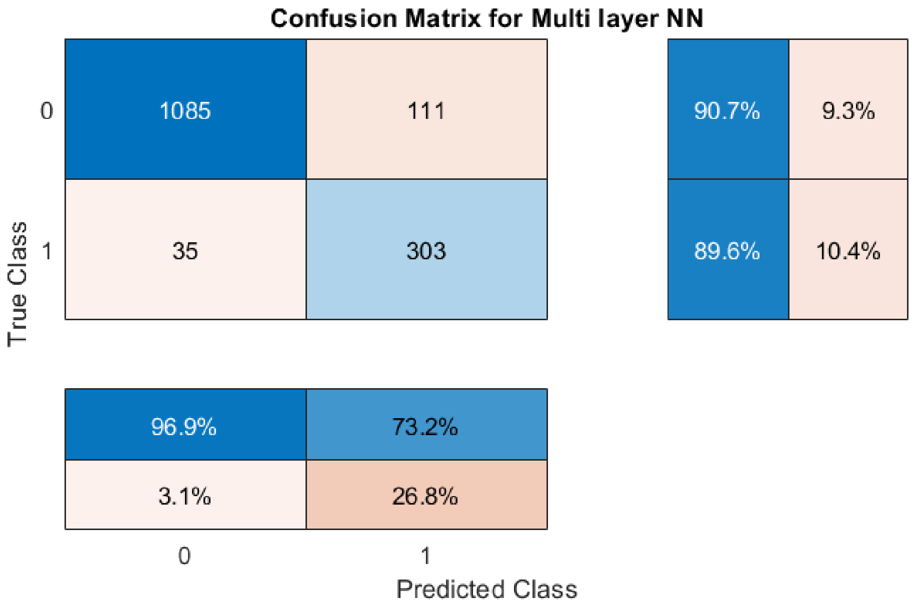

Figure 52: Multi layer neural network results

LSTM (Long Short Term Memory) networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series.
The last algorithm that gives good results is the Coarse Tree.

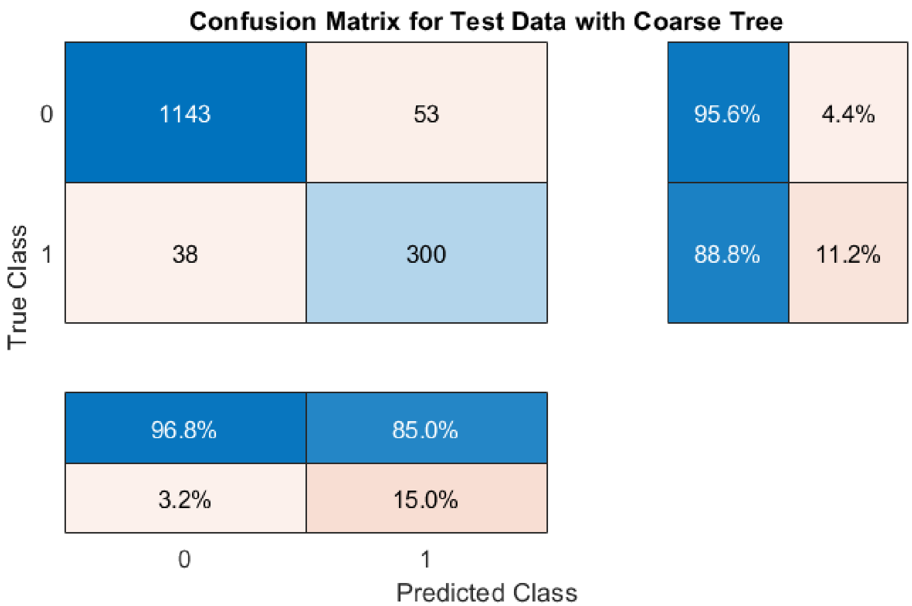

Figure 53: Coarse Tree results

4.6 Analysis of the results

The results until now show, regardless of the algorithm and its complexity, an uncertainty beyond which they cannot go. For this reason, analyzing the results an important finding emerged.
In the over 1500 points taken into consideration, sometimes the dressing do not occurs at the 100^{th}, as the planning foresees, sometimes it occurs at the 88^{th} or before.
The motivation is that, every welding gun has an human operator which acts as a garrison, if the human operator realizes that the machine is at its limit and from that moment on the machine enters a critical work situation,he can wait for the end of the current cycle and then arrange for an early dressing.
In particular, in the 1500 used to test the algorithm this happens 42 times.
In this occasion the label used as classification output was '0' as a safe situation, but in reality it is a '1' as the unsafe situation.
Stricly analyzing these points, it emerged that the algorithms can recognize these situations with different success rates.

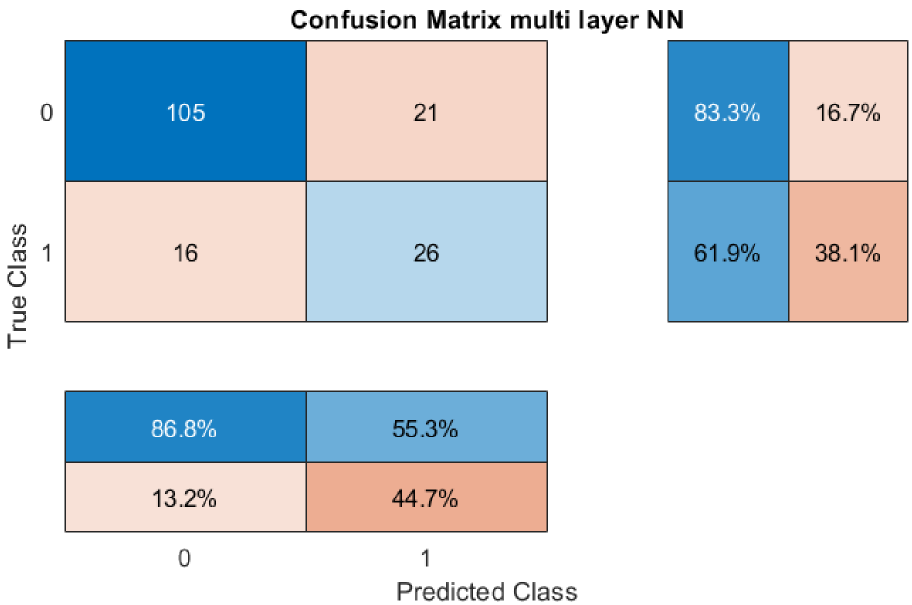

Figure 54: Multi-layer architecture results on false positive

In this test the 42 false safe situation were mixed up with 126 real safe cases. Is it possible to see that the multi-layer architecture is able to detect the possible unsafe situation with a precision of the 62%. Obviously this means that in the Figure 52 the percentage of the correctly detected case is higher then the reported percentage.
The second study case is the 1-layer architecture. This algorithm, until now, has a slightly lower success rate among proven methods. Now introducing the 'false 0' the situation change.

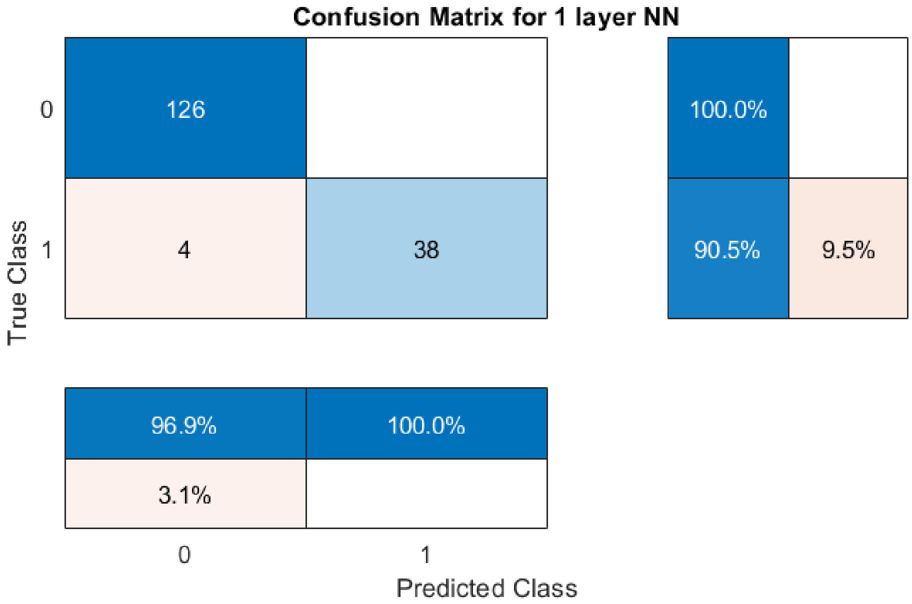

Figure 55: 1-layer architecture results on false positive

The 1-layer neural network is able to find the false positive situations with a success rate of the 90.5%.
The last algorithm is the Coarse Tree.

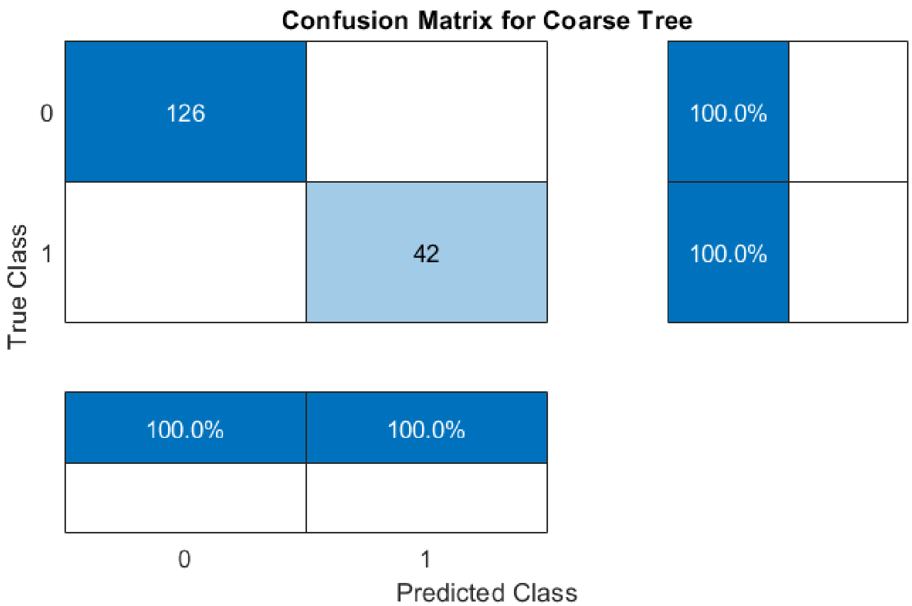

Figure 56: Coarse Tree architecture results on false positive

The coarse tree algorithm reached a success percentage of the 100% analyzing the false 0 situations.
The problem now is, are this results encouraging or not ?

The results obtained so far must be viewed critically, if the algorithms detected with a so high precision the false 0 situations, is it possible that the unsafe situation (1) cataloged as safe ones (0) were truly safe and the dressing were not needed ? Machine learning algorithms are absolutely not foolproof and to be reliable need tons and tons of data but, after the evidence emerged, it is possible to assert that the algorithm went far beyond from the expected value.
The question now is, which algorithms give the best results ?
Looking at the confusion plots seems to be Coarse Tree, but is it a certainty ? For this reason the desirable next step of this work it is to apply the algorithm on the real plant and to compare its prediction with opinion and experience of the human operator or to keep on training the algorithm with more detailed data sets.

4.7 Other tests

Probably the best solution to realize a reliable predictive maintenance system is to develop an algorithm and for each program. The reason behind this choice is that different welding guns or programs have really different requirements. For example can change the duration of the welding, the current necessary to melt the metal sheets or the thickness of the metal sheets itself. Nevertheless it can be interesting to try the algorithms on a machine and a program different from the one for which they were designed.

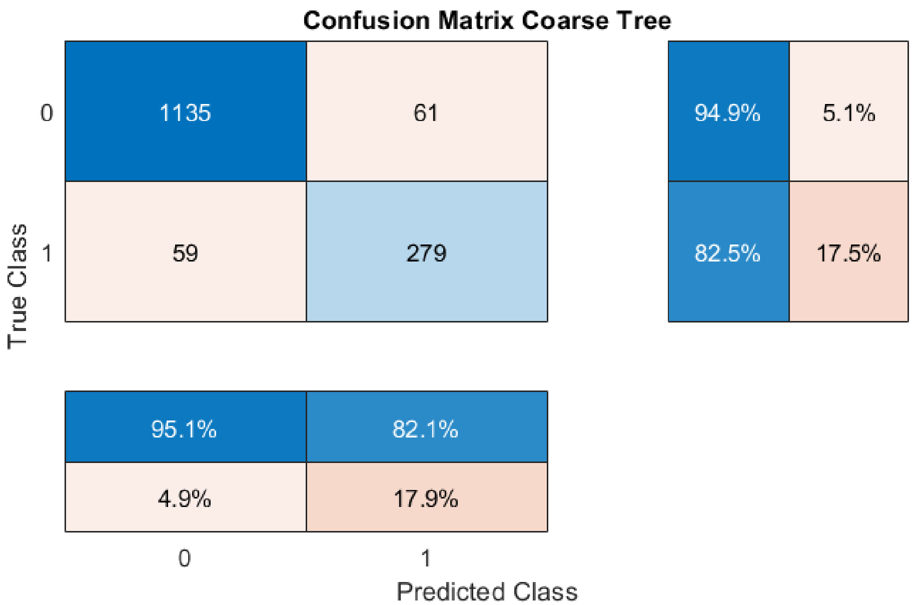

Figure 57: Coarse Tree architecture results

It is possible to see that the Coarse Tree keeps good results even if applied on another welding machine that perform a completely different point.
Moreover in this results there are not information about the early dressing.
The multi-layer algorithm if applied to another machine gives worst results.

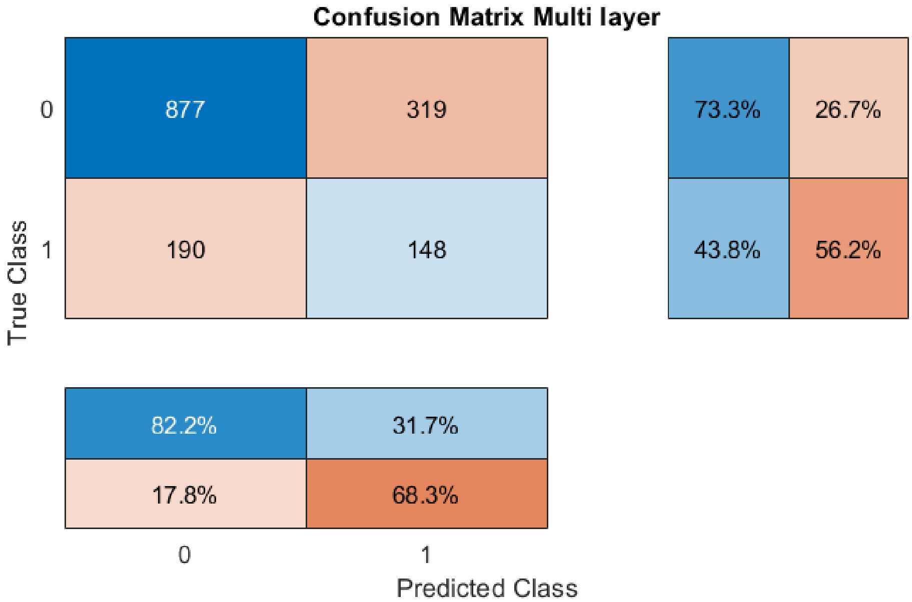

Figure 58: Multi layer architecture results

In my opinion this result does not mean that the algorithm is unreliable, because it is possible that different programs on different machines or even on the same one imply really different electrical quantities. For example if the metal sheets are thicker, it is possible that the mean resistance is higher as well as the voltage, and it is not easy to forecast how the other descriptors will change.
For this reason the easier solution is to train an algorithm for each program of each machine.
In this way, when at a certain welding in the cycle, the system provide the unsafe label, if the subsequent algorithms provide the same label, a dressing will be scheduled at the end of the cycle.
This method gives the possibility to have greater security given by the cross-checking of different algorithms.

5 Conclusion and future developments

The industry 4.0 arises from the ongoing industrial automation that in the last years meets new technologies to improve work conditions, to create new business model and to increase productivity and quality.
Large scale machine-to-machine communication and the internet of things (IoT) are integrated to make the machine smarter and smarter, able to self-monitor and able to analyze and diagnose issues without the need for human intervention.
For this reason it is not a long shot to say that predictive maintenance will become, in the next years, a technology possible to find in every industrial context.
In this work of thesis have been moved the first steps of a longer path.
To sum up, good results have been obtained under different aspects:

- **The physical model** obtained, considering that only current and voltage are measured, gives a great approximation of the real phenomenon but, most important things, the study of a physical model allows to have a deeper knowledge of the welding and it suggests which information can be useful to our machine learning algorithm to recognize critical situation. For example, the energy variation and the resistance variation are crucial quantities in the determination of the phenomenon.

- **The machine learning algorithms** gives really encouraging results, making it clear that a predictive maintenance system for electrodes dressing is possible.
Different algorithms gives different results but under current conditions it is not easy to define if there is an algorithm better than one other, the decision tree (Coarse Tree) reach the higher performance but there are still some uncertainty in the data set that need to be defined.
The main goal of the development of the predictive maintenance system has been the usage of the statistical descriptors. This approach admits to give to the algorithms a precise description of a welding point giving in input a very small sample of data, giving an advantage in computational terms.
Moreover a statistical description of a curve highlights the changes in shape and trend which, in the studied case, was exactly what the algorithm needed to define different situation but surely is a strategy that could find many applications in a lot of machine learning algorithm.

- **The critical point analysis** gives also the possibility to understand what can be done, directly on this work of thesis, to achieve better results. A crucial point during welding is the definition of the resistance at the ends of the electrodes. Actually it is computed starting from the measured current and voltage, for this reason the resistance is comprehensive of all the components in the circuit (from the transformer to the electrodes) while the resistance measured between the electrodes could be of greater interest for the algorithms because will track the change in resistance directly where needed.
If it was possible that, measuring the resistance at the ends of the electrodes it could be defined a model of the electrodes remaining useful life.
Another critical point is the heat produced on what there are no information.

> The heat produced during welding have a crucial role both in the dressing prediction and in the welding it self. After seeing the heavy role that energy plays in the welding phenomenon, data about the heating are possibly the key to improve the results.

At this point it is possible to do some considerations on the next step of this work and how possibly the predictive maintenance system can be improve.
In first place a new technology, to be considered reliable, have to pass a long testing phase. The results of this thesis are obtained using measured data but in simulation and surely the algorithms need to be tried on the real plant to know how they truly work.
Moreover, like it was explained in the dedicated chapter, in my opinion the predictive maintenance system can not be entrusted to just one algorithm but it is needed a net of different machine learning algorithms that act on different programs of the same machine. The reason is that machine learning is based on statistics and interfacing more algorithms among them, it means to have a more accurate prediction.
Then a more accurate physical study of the system can give important tips on the data that are more useful to the machine learning algorithms to identify the critical situations before they happens.
Another attempt that could be done is to make a frequency analysis of the welding points. It is possible that the behavior of welding points in safe and unsafe conditions have the power located in different frequency ranges.
This is another possible approach to define the remaining useful life of the electrodes.
Naturally it is desirable, for this kind of technologies, the usage of a Iot platform able to collect the data from all the available machine at the same time. In this way an human operator can be informed in real time about the condition of his machine and also he can be informed on how the predictive maintenance system is working in the rest of the plant.
It is important not to forget that the main resource is the human experience and for this reason an heuristic analysis, made continuously asking to human operator their opinion and what their experience recommends, could give important tips about the problem that the study and the theory can not give.
For this reason on the IoT platform can be equally registered the measured data and alert from the machine and personal notes form human operator made to transfer the experience from the worker to the one who have to develop the predictive maintenance system.
Moreover, to try to generalize the predictive maintenance of the dressing problem, could be useful to group the data coming from different programs of different machines performing welding points with similar requirements.
Anyway after the predictive maintenance system it is possible to pass to a proactive maintenance system.
Proactive maintenance is a preventive maintenance strategy that works to correct the root causes of failure and avoid breakdowns caused by underlying equipment conditions. The purpose of proactive maintenance is to see machine failures as something that can be anticipated and eliminated before they develop. Creating a proactive maintenance program helps organizations find hidden inefficiencies.

References

[1] *Story of the ISI-Welding Company*, `http://en.isi-gf.com/about.aspx?TypeId=1&FId=t1:1:1`

[2] Erik Baldoin, *Evoluzione delle pinze robot per saldatura a punti*, Politecnico di Torino, 2009

[3] Luciano Furlanetto, *Manuale di manutenzione degli impianti industriali*, Franco Angeli, 2003

[4] Nakajima, 1984

[5] L. Fedele, L. Furlanetto, D. Saccardi, *Progettare e gestire la manutenzione*, McGraw Hill, 2004

[6] G. De Nicolao, R. Scattolini, *Identificazione Parametrica*, Dipartimento di informatica e sistemistica dell'Università degli Studi di Pavia

[7] *Design Time Series NARX Feedback Neural Networks*, `https://it.mathworks.com/help/deeplearning/ug/design-time-series-narx-feedback-neural-networks.html`

[8] Ruiji Sun1, Wangling Yu, Haiyan H. Zhang, Manohar Das, Qingyou Han, *Mathematical Modeling of Resistance Spot Welding*,Journal of Multidisciplinary Engineering Science and Technology (JMEST)

[9] R. Adami, *La saldatura a resistenza*, Università degli Studi di Padova

[10] Luigi Fortuna, Salvatore Graziani, Alessandro Rizzo, Maria Gabriella Xibilia *Soft sensors for monitoring and control of industrial processes*, Springer Science & Business Media

[11] Andrew Ng, *Machine Learning Yearning*