20th International Conference on Modelling and Applied Simulation
18th International Multidisciplinary Modeling & Simulation Multiconference

ISSN 2724-0037 ISBN 978-88-85741-61-4 © 2021 The Authors.
doi: 10.46354/i3m.2021.mas.007

Exploiting Machine Learning and Industry 4.0 traceability technologies to re-engineering the seasoning process of traditional Parma's Ham.

Davide Mezzogori[1*], Francesco Zammori[1]

[1]Università degli Studi di Parma, Dipartimento di Ingegneria e Architettura, Via G.P. Usberti 181/A, Parma, 43100, Parma

[*]Corresponding author. Email address: davide.mezzogori@unipr.it

Abstract

The work presents a Machine Learning approach for predicting the quality of the curing process of Parma ham, combined with a study of business process re-engineering, based on RFID and Deep Learning technologies for automatic recognition and tracking of the hams along the curing process. Quality management has proven to be crucial for efficient and effective processes, even more so for the food industry, both for commercial and regulatory purposes. This is even more evident in artisanal-based processes, such as the one concerning traditional Prosciutto di Parma seasoning. The work proposes and compares a Feed-Forward Neural Network and a Random Forest for predicting the distribution of the number of hams by commercial quality class of a given aging lot. Such a prediction, based on origin, process, and curing data, can provide early indications of process output, enabling strategic commercial competitive advantages. The importance of the genetic component in the determination of the final quality is also evaluated, as it is considered one of the most influential external variables. Moreover, following the AS-IS description of the current process, a redesign is proposed, to enable data collection and tracking of individual ham in order to propose a future precision prediction system that would allow even finer control of the process.

Keywords: Machine Learning; Deep Learning; Business Process Re-engineering; Object Detection; Rfid

1. Introduction

Quality assessment, prediction and control have always been central to intelligent and efficient process management (Zu et al. 2008). The relevance is even greater in the food industry, due to strict regulations and to the role it plays, commercially, with respect to the consumer perception (Dora et al., 2013).

For some market segments or niches, it is even possible to classify the same product in terms of different commercial quality classes, sold through different sales channels and at different prices. Typical examples are that of cheese, vinegar and spirits (generally classified in terms of years of seasoning), chocolate, coffee and wine (in terms of ingredients mix and their origin), and many others. The management objective should be that to satisfy the different existing trade channels, each with its own quality requirements, as efficiently and effectively as possible. Hence, it is a question of optimizing the allocation of available resources (the different quantities available per quality level) in relation to market demand. Quality, as we know, is determined, by the quality of the raw materials, and by the management of the transformation process (Evans and Lindsay, 2002). Standard procedures for quality and process control are therefore essential. Anyhow, there may be also some uncontrollable, and typically exogenous, variables that could influence the output. Therefore, it is highly desirable to identify and quantify such variables, so that the expected quality level (of the output) could be determined as early as possible. It follows that the development of ad-hoc and advanced quality prediction systems offer an obvious competitive advantage to companies belonging to these sectors. The present work belongs to this stream of research, as it describes the feasibility study of a Machine Learning based approach for quality prediction, implemented in a firm for the processing and seasoning of the 'Prosciutto di Parma', an excellence of the Italian ham.

The processing and drying process of Prosciutto di Parma, whose origins date back to the artisan world, is regulated and prescribed by the consortium of Prosciutto di Parma, which has largely industrialized and standardized the process. An important specificity of the considered process must be highlighted, in relation to other artisanal production processes. For example, for wines and spirits production, the final qualitative result is largely pre-determined by the known characteristics of the raw material and the ageing process. Conversely, in the case of the Prosciutto di Parma, although the process is defined and unambiguous, the final quality is heavily determined by two exogenous factors: i) small variations (i.e. temperature, humidity, salting, etc.) in the first steps of the aging process when the pork's legs are still fresh and ii) characteristics of the raw material (e.g. pH, percentage of muscular mass, etc.), mainly linked to genetics and to the breeding environment.

In line with the work by Goyace et al. (2001), one of the first to discuss the use of artificial intelligence to assess subjective quality measurement in the food industry, the main objective of this study is to assess both the feasibility and the opportunity to develop an automated framework to assess the end quality of Prosciutto di Parma legs. Specifically, the framework will exploit an automated traceability system combined with a Machine Learning approach, based on data related to the manufacturing and aging and to the input material, measuring several quantities associated with the factors mentioned above. It is important to note that this is where the innovative proposal of this work lies. Indeed, the technologies here considered are usually employed for traceability for regulatory purposes in the agri-food field (Dabbene et al. 2014). Instead, we apply them to assess the feasibility of a quality prediction framework, based on process and material data.

Furthermore, a second goal will be to estimate the impact and the predominance of genetic factors on the overall quality of the ham. According to the experts on the field, in fact, genetic shows the greatest effect; however, this assumption has not yet been scientifically verified.

Finally, as a third and last objective, a Business Process Re-engineering (BPR) study, employing Industry 4.0 traceability technologies, is described, with the aim to enhance the existing traceability framework, so to better track each individual product, in contrast of the as-is approach, designed around production batches. Such approach should yield even more predictive power to a Machine Learning base approach for quality prediction, which would then be able to target the single product, thus outputting even more reliable predictions.

The rest of the paper is organized as follows: section 2 describes the problem settings in more details, as well as the existing process framework. Section 3 describes the available data gathered during the manufacturing process, as well as the Machine Learning methods implemented. Section 4 describes the results obtained, while Section 5 presents the BPR. Finally, Section 6 presents conclusions and future works.

2. Process description: the As Is state

In this section we describe the current (or As Is) manufacturing process and the aging workflow. The mapping of the As Is state is therefore instrumental to the description of the existing data sources that have been exploited to define the forecasting system of the commercial quality classes of the finished product.

The process begins with the receipt of fresh (uncured) pork legs from the slaughterhouse. Legs are shipped in large batches. and may come from several breeding farms. Each farm must be located in a specific geographical area in order to comply with the regulations governing the Parma Ham Consortium's supply chain. For each batch, the number of pork legs received from each farm is documented and recorded, in compliance with food safety and legislative regulations. Knowing the farm where the raw meat comes from makes it possible to indirectly obtain information regarding its genetic characteristics, because each farm constitutes a constant and predictable genetic pool. This information is considered valuable, as ham experts believe that the genetic (i.e., the strain of origin) of the pigs contributes largely to the final quality level of the ham.

Next, the production cycle of hams alternates transformation processes and resting periods. The process is mostly standardized and regulated (by the Consortium). However, for secrecy reasons, detailed processing data will not be fully disclosed in this study. Anyhow, this does not result in any alteration of the provided results. Very briefly, the cycle includes:

1. Two salting cycles interspersed with the same number of rest periods of 5 and 20 days in cold rooms (with horizontal positioning of the legs),
2. Desalting and resting for 30 days,
3. Grooming (the ham is trimmed and hung vertically) and rest for 60 days,
4. A cycle of tempering, washing (to remove mold), storage in the drying room for 6 days, then in the post-drying room for 30 days,

5. Pre-curing in a special warehouse for 30 days,
6. Final curing in another site with different temperature and humidity conditions for a total of up to 24 months.

More in detail, once received, the raw meat is immediately processed, before entering the first drying phase. The processing consists of the following steps:

1. The meat is cleaned by six workers, which removes unaesthetic and unnecessary parts that would hinder the aging process. Cleaning is also intended to confer the same level of aesthetic quality to all legs.
2. Downstream of the trimming station, a sorting machine, is used to weigh and brand each leg. Weighing is required both by national and international regulations (i.e. underweight or overweight legs are not further processed), as well as to measure the distribution of the number of legs per class of weight.
3. Next, legs are placed on a conveyor belt, and passe through the massaging machine, where rollers compress the legs at a pressure of about 1 bar so as to slightly break the muscular fibers. This is needed to guarantee a good absorption of the salt that will be added in the following phases.
4. The process continues at the pre-salting machine, which carries out three different activities: i) it spreads wet salt all over the leg; ii) it combs the leg with steel brushes; iii) it removes salt from the lean part of the leg, by blowing a strong jet of air from a nozzle.
5. Finally, the leg is processed manually by an expert operator, who is in charge of the main salting procedure, i.e. to manually salt the lean part of the ham, in such a way that the combination of pre-salting and salting activity distributes a salt weight percentage of 1.6% on the leg with a tolerance of 0.2%.

The final step ends with an automated check of the total added quantity of salt. To be compliant with quality standards, the percentage of added salt should fall within a predefined tolerance range. Nonetheless, at present, these limits are just indicative and non-mandatory. In case of non-conformity, the process is not stopped, and an eventual reprocessing is at the discretion of the worker.

At the end of this process, the hind legs have to be placed on mobile shelves, which, through an overhead monorail system, will be positioned in the curing cell. These movements are carried out by a robot, i.e. a large anthropomorphic arm which has a large fork at the end and is used to move a maximum of 6 legs at a time. Legs enter in the automated robot-station and are placed, one by one, on a shelf. The first leg will be positioned upwards, while the second leg will be positioned downwards and so on, until a shelf of 6 legs is formed. The robot is then activated and loads the group of six legs (placed on a shelf) in a 9-levels rack. Hence, a maximum of 54 legs can be stored on each rack. All this process is carried on in an automated manner, without the intervention of any operator. Only when the rack is fully loaded, an operator takes it away, using a system of ceiling mounted trolleys, and places it in the first temperature-controlled storage room. This is where the hind legs rest at a controlled temperature between 0.5°C and 3.5°C. There are no special restrictions on humidity, which is usually around 70%.

After a week, the racks leave the temperature-controlled room, so that the legs can be further processed. Essentially, this processing step removes the salt previously applied to the legs and adds new and fresh salt. The salting process is totally equivalent to the one described before. The only notable difference concerns the salt weight percentage that, in this case, must be equal to 1.1%, with a tolerance of 0.2%. Once again, the robot stocks the salted legs on the racks and the racks will stay in the temperature-controlled room for about a month. After this period, legs are carefully cleaned (so that all the previously applied salt is removed) and moved to another controlled room, where both the temperature and the humidity are constantly monitored. Specifically, temperature must remain between 1°C and 4°C, and humidity between 58% and 60%.

After these resting periods in temperature-controlled rooms (for a total of 2 months), the legs are prepared for the aging process at room temperature. To this aim they are cleaned for the last time, vertically placed on stocking frames, and stored for another month in the final temperature-controlled room, between 15°C and 16°C. Then, after a sanitizing process, legs are stocked, at room temperature, in cellars, where the first seasoning will last for around 13 months. Hence the total time from leg reception to maturation is about 14 months.

It is at the end of this 14 months process that the hams are classified into commercial quality classes. To this aim, a highly qualified and experienced operator, named "*spillatore*", inspects each leg both olfactorily and visually and, mostly depending on organoleptic properties, determines their quality class. Other elements considered in the classification are the shape, the exterior aesthetic and the size of the hams. Anyhow, classification is purely subjective and based, exclusively, on the experience gained through years of craftsmanship. It is important to stress that, after this categorization, the aging process will continue, for further 30 months, but the quality level will no change anymore. Apart from rare phenomena (such as the creation of molds) which irreparably downgrade the product, the final seasoning cannot modify the quality class of the hams. In other words, the process variability is "limited" to the first 14 months.

It is therefore clear that it is of interest to develop a system for forecasting the final quality, both in operational and strategic terms, anticipating as early as possible the division into classes that takes place 14 months after receiving the fresh legs. Operationally, this would provide an early view of future availability, thus allowing commercial operations to take place accordingly. Strategically, it is of interest to understand which factors have the greatest impact in determining the perceived quality level. In particular, factors related to pig breeding are considered to have a greater influence than process variables. Additionally, the genetic traits that are expressive of each herd are considered to have a strong impact on the quality of the product.

3. Data and Methods

The company has three production plants where it carries out both the salting and storage cycles in a refrigerated environment and the curing of the hams. As above explained, until the 14-th month, when the *spillatore* classifies the legs into 4 commercial classes (based on their commercial quality, size and shape), the company the company does not know the exact number of legs that will be marketed in a given quality class.

This strongly binds the marketing strategies pursued by the company and makes uncertain the commercial planning phase. Therefore, it is of strategic importance to accurately estimate the number of legs that will fall into each class, well before the sensory classification phase takes place, or even when fresh legs (coming from the slaughterhouse) are received and selected.

The purpose of the present work is to evaluate whether it is possible to carry out this classification using multivariate and multi-output regression models fed with data related to the origin of the legs and/or with the productive and seasoning process. Indeed, previous works hint at the applicability of these methodologies to tackle quality monitoring for ham processing (García-Esteban et al., 2018), although in a different manner and with different purposes to the present work. In particular, to train regression models, the following will be considered:

- Data related to the origin of the thighs (farm of origin).
- Data collected during acceptance and reception.
- Data linked to salting and storage cycles in cold rooms.

The definition of these three different data sources will allow testing of alternative models based on different combinations of data. For example, regression models based solely on source data could be considered, rather than models based solely on process data. In this way it will be possible to test different hypotheses and understand which regressive variables are important for the final classification. In the light of these results, the company will be able to evaluate the opportunity to improve the internal traceability system, by equipping it with radiofrequency and/or image recognition systems.

With respect to the origin of the hams, as suggested by the company's experts, it is logical to assume that the farming of origin (i.e. the farm where the piglet was born and weaned) may correspond to peculiar genetic traits and that this may have a strong impact on the final quality of the ham. The origin data is known because, by law, the farm of origin is impressed as a mark (i.e. tattoo, see Figure 2 in Section 5) on each leg. However, since no tracking system at level of the single ham exists, this data is not recorded precisely. Instead, for a given batch of hams, the operating staff merely records the total number of hams coming from each origin farm. A proper collection of origin data is further complicated by the fact that not all farms are involved in each batch, and the number of distinct farms is rather high (n = 1900). For this reason, in order to reduce the number of farms, it was necessary to aggregate the different farms from which the legs come from. Aggregation, made in terms of geographical areas, followed two paths: (i) a detailed aggregation by Italian provinces and (ii) a less precise aggregation by regional areas. Briefly, with regard to the first grouping methodology, we would like to point out that the province where the farm of origin is located was obtained by considering the first two letters of the tattoo marked on each leg, that encode, exactly, this information. We also note that the farms can be grouped in 79 provinces of northern Italy and so, to further reduce this number, a Pareto analysis was carried out. By doing so, 14 main provinces, accounting for 90% in volume, were identified; all the remaining ones were grouped in a single cluster, labelled as "other provinces". As for the second, regional areas have been defined, by joining neighboring provinces until they cumulate, on average, cumulate 5% of the total amount of legs purchased. Note that, the obtained region may not necessarily correspond to Italian administrative regions.

The variables (for a total of 20) concerning data collected during acceptance and reception can be grouped in the following groups (specific variables are omitted for privacy reasons):

- Month and week of processing, to handle seasonality effects.
- Type of leg, that is, distinctive physical characteristics, which is important at commercial and production level. In fact, according to the different characteristics, the products will follow different commercial channels, while at the production level, they will undergo ad-hoc processing.
- Average PH level.
- Distribution of number of hams per weight class.

Concerning production data, the following variables were used (for a total of 13):

- Percentage of weight loss after each processing step
- Percentage of added salt, after the first and second steps.

Lastly, we note that all these variables are relative to a production batch (i.e., as already stated, the data-recording system does not track each individual ham). Specifically, in the acceptance phase data are averaged over a randomly drawn sample. Similarly, processing data (recorded through all the steps described in the previous section) are measured with respect to a specific rack, aptly named "*sample racks*". Things are further complicated by the fact that the sample racks measurements are associated to all the batches processed during the same interval of time (typically a week). So, data were pre-processed and reconciled in order to properly associate to each batch both its acceptance and processing data. For the same technical reasons, the classifications models will be developed to predict the "quality class proportion" for each batch of ham (i.e., how many hams of a batch for each quality class), and not the quality class of a specific leg. Due to the actual data collection process this is not yet feasible.

Table 1. Summary of variables describing a given batch of hams

Group of variables	Description	Kind	# of variables
Temporal	Month and week of processing	Numerical	2
Physical characteristics 2	Visual evaluation	Categorical	4
	Distribution w.r.t. weight (ranging from 12 kg to 17 kg)	Numerical	9
	PH level	Numerical	1
	Distinctive physical features (i.e. commercially relevant)	Categorical	4
Production data	Weight loss	Numerical	10
	Added salt	Numerical	3
Origin	Distribution w.r.t. province of origin	Numerical	15
	Distribution w.r.t. geographical region of origin	Numerical	15

Table I gives a summary of the variables employed, with a distinction between numerical and categorical variables, and the cardinality for each group of variables. We note, again, that for privacy reasons it was not possible to fully disclose each variable measured.

The Machine Learning models tested are the Random Forest (Friedman et al., 2001) method and a Feed-Forward Neural Network (Goodfellow et al., 2016). The models are implemented using Python 3.9, using scikit-learn and tensorflow libraries. Both models take as input 34 variables (32 continuous, 2 categorical) and outputs 4 response variable, representing the percentage of hams for each quality class, with respect to the total number of hams in the batch. Both models' hyperparameters were optimized by means of grid search. In detail, the Random Forest is configured as follows:

- Number of estimators: 1000.
- Maximum depth: 50.

The neural network is so designed:

- Two hidden layers of 128 and 64 neurons each.
- Activation: relu.
- Epochs: 1000.
- Optimizer: adam.

For better visualization, Figure 1 shows the flowchart depicting the overall method herein described.

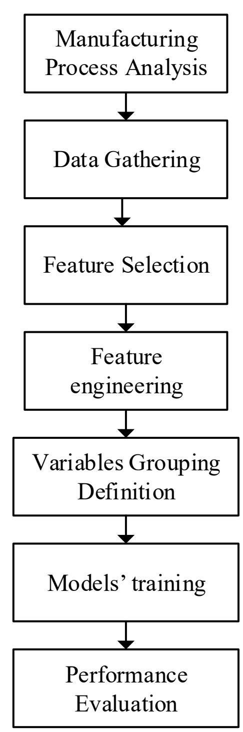

Figure 1. Methodology flowchart visualization

4. Results

The models, developed to assess the viability of a quality prediction model based on Machine Learning methods, are tested using three kind of data: i) data related to origin of raw material, ii) data collected during processing steps, iii) data collected in seasoning rooms. The following is the list of models, i.e. different combinations of source data. Each is meant to test the importance of different sources of data, including or excluding some of them, and testing different groupings of some of the variables.

1. The first model is built using only the second and third group (*schema 1)*. This is meant to be a benchmark.
2. To assess the importance of including data about the origin of the raw material, the second model includes all sources of data (*schema* 2). Origin data is grouped by province.
3. All variables are employed, origin data is grouped by region (*schema* 3).
4. No process data is included (*schema* 4).

The last model is meant to test the following hypothesis: to date, the process data are not measured with respect to individual hams but are averaged using the "sample racks" that are tracked along the process. This leads to a repetition of the same values on several records of the data set used for training models, which could result in a lower discriminatory power of the data with a substantial worsening of the final prediction.

The dataset, which spans three years, is split into a training set with 900 observations and a test set with 200 observations.

Finally, models' performances are measured in terms of Root Mean Square Error, and results are reported in Table II.

Table 2. Models' performance in terms of RMSE.

Schema	Neural Network	Random Forest
Schema 1	0.0873	0.0855
Schema 2	0.0995	0.0838
Schema 3	0.0914	0.0716
Schema 4	0.0833	0.07

As it can be seen, the Random Forest always outperforms the Neural Network, in all the schemas tested. Moreover, it is interesting that the lowest RMSE is achieved with respect to the fourth schema. Indeed, for each methodology, the fourth schema sensibly lowers the error, achieving the overall best. Thus, it could be stated that excluding the process data, which unfortunately are associated to more than one processing batch, helps to have a better prediction. Finally, focusing on the Random Forest, it can be seen that the inclusion of information related to the origin, breeding, and genetics of the hams, consistently helps to achieve lower errors, confirming the hypothesis that genetics and origin of the hams are important factors for the final commercial quality.

5. Business Process Re-engineering: an improved tracking system

As the results highlight, it is possible to predict for each batch, and with an acceptable low error, the percentage of legs that will be assigned to each commercial quality class. Moreover, data related to the origin of the hams proves to be very important to achieve good and reliable predictions. Conversely, process data revealed to be too much aggregated to be useful, as they are associated to more than one processing batch.

Both findings highlight the value of redesigning the tracking system, in order to be able to track each single ham during its processing. With such a system it would be possible to assess the quality class at the level of the single ham, making a much more precise and targeted prediction, instead of predicting the incidence by class on the total number of hams of a batch.

The key introduction would be to add an UHF passive RFID tag to each ham, a technology that have been already proved to be applicable in the food industry (Kelepouris et al., 2007). Specifically, as suggested by Wessel (2008), the tag should be tied to the upper and (bony) part of the leg at the beginning of the processing, right before any cleaning operation. If, instead, the tag was nailed, quality could be compromised by possible growing of molds. The use of an RFID tag would allow tracking each individual ham during all processing steps and would permit to associate to it individual and more precise data collected during the processing. In particular, it would be possible to link to each ham the following data:

- Weight at the beginning of the process.
- Percentage of added salt during the first and second phase of the process.
- Weight loss after each seasoning processes.

Note that changes to the process would be minimal. Indeed, all of these variables are already collected (for each sample rack) or could be collected during the process. Indeed, the weight of the ham can be measured with an industrial scale which is already installed along the processing line. All hams must pass through it at each step. By augmenting the scale with an RFID reader, it would be possible to precisely record the weight trend from the beginning to the end of the maturing process. The same applies to the recording of salt added during the first and second phase. A scale is already available, and it used to measure the amount of salt applied to each processed ham. However, the data is not recorded but randomly controlled by the line-operator as an indicator of the process' state and/or associated, at an aggregate level, to the "*sample racks*" now employed. Instead, with an accurate and precise RFID tracking system, the data could be linked to each ham, during each phase.

Moreover, the RFID technology could also be applied to track each rack, using active tags, equipped with temperature and humidity sensors. In this way, leveraging the anthropomorphic robot, and augmenting it with RFID capabilities, it would be possible to link each ham to the rack in which it is stored. The onboard active tag sensors of the rack would be then used to monitor temperature and humidity data recorded in the seasoning rooms. Clearly, by operating in this way, values of recorded humidity and temperature will be associated to each rack and not to each ham. That is, the same values will be shared by all the hams placed on the same rack. Using sensing RFIDs for each leg would be too costly and this solution would also be problematic due to the sever processing conditions undergone by the legs during the salting and massage steps. However, it is reasonable to think that such individual measurements are not needed, and that and averaged values is more than enough. Even more practically, the onboard sensors could prove to be useful operatively, triggering alarms when anomalies are detected, such as extreme fluctuations of temperature or humidity for a given position in the room.

Installing and leveraging an integrated RFID system, both for each individual ham, as well as for the racks used in the seasoning rooms, would ideally boost the data collected and, thus, the forecasting power of the proposed methodologies to predict quality levels prior to the human classification. However, this would only enable the data collection system to address specifically to each ham data gathered during the processing system. As demonstrated in previous section, knowledge of the origin, breeding, and genetics of the hams have proven to be very important to empower the forecasting tools. Fortunately, such information is available on each ham, encoded in the tattoo applied at the breeding farm, as previously described. Getting the tattoo read and decoded by the line-operators is feasible, but time consuming and certainly error prone. Therefore, developing an automatic system for the detection and recognition of the tattoo is a fundamental prerequisite to fully automate the data gathering and traceability system. In light of this, a Neural Network based system for the automated character recognition of the tattoo has been designed. To test the feasibility of such tool, 1900 photos of tattoo hams have been taken. The tattoos encode in six characters the following information: i) the breeding farm, and ii) the month of birth of the pig. The first two and last three characters encode the breeding farm, while the third one encodes the month of birth. Figure 2 displays two examples: the first shows an easily readable tattoo, while the second showcase an example of a tattoo which is merely readable by the human eye.

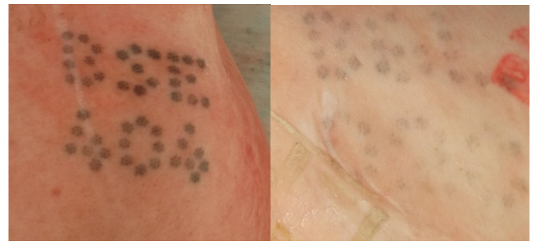

Figure 2. Two example of ham tattoos: a readable one (left) and difficulty to read one (right)

Automatic reading of these tattoos can be related to OCR (Optical Character Recognition) problems. Indeed, the problem can be cast in terms of multi-class object detection. To tackle such problem, we implemented a YOLO neural network (Redmond and Farhadi, 2018), suitable for real-time object detection systems in industrial environments. The YOLO model is a deep neural network composed of 106 fully convolutional layers. To train it, a large amount of data must be collected. However, it was possible to gather only 1900 photos of tattoos. Being too little of a dataset, we resorted to apply a powerful technique, which can be used in absence of large training sets. Through transfer learning (Pan and Yang, 2009), it is possible to leverage already trained deep neural network, and to fine tuning them to the task at hand, by essentially freezing all the weights associated to all layers, except the last. Only the last layer, the output layer, is retrained on the dataset at hand. Since only the last layer weights must be trained, the amount of data needed is far lower than to train the entire network. Indeed, we applied transfer learning on the YOLO deep neural network trained on the COCO dataset (Lin et al., 2014), altering the final layer, so to be able to recognize the characters of the breeding tattoos. The dataset has been manually labeled, tagging each tattoo character, and associating to it the corresponding class. The image set collected has been split on a train and test sets of 1600 and 300 images, respectively. Preliminary tests show very promising results: 260 (86%) tattoos are correctly and fully detected and recognized, and the remaining have only one character not detected and thus not recognized. However, heuristics may be designed to complete the missing character, based on the rest of the correctly recognized tattoos of a batch.

6. Conclusions

The article describes the typical aging process of Parma ham, and proposes an innovative forecasting model, to predict the expected commercial quality level of a batch of seasoning legs. The predictive framework, implemented by comparing Neural Network and Random Forest, is fed by process data, seasoning data, and data related to the breeding farms from which fresh pork's legs come from. Predictive models are trained, and their accuracy is tested against data collected over a three-year period. Obtained results are very encouraging and certify the possibility to implement a robust quality prediction system based on data that are currently available and, especially, data related to the genetic (strain of origin) of the pigs. This also confirms the intuition of experts of the fields, who have always considered genetics to be essential for the final quality of the ham. Results also demonstrate that the forecasting framework can be implemented with no or marginal changes to the current productive process. All required data are available and hence, with a very little effort, in the next future this system will play a pivotal role for the company. It is expected that the company will leverage on it for commercial purposes, especially to promise certain quality levels and to negotiate the price, well ahead of delivery, with certain benefits in terms of both profit and revenues. Anyhow, the analysis offers interesting ideas also for a possible redesign of the data collection system. As stated, the minimum amount of data is already available, but further improvements could be achieved moving from the current batch traceability system to a single-leg traceability system. To this aim, automatic recognition technologies should be used, and the paper proposes a combination of RFID and Deep Learning. The latter technology, validated through a preliminary experimental campaign, would be used for image recognition, for the automatic acquisition of the important unstructured information located on the hams. Finally, we note that, although this work makes use of several standard methodologies, it presents different points of innovation. First, Machine Learning methodologies are used for the first time in this specific market niche. Second, the focus is on the forecast of the so-called commercial quality (i.e., the 'sensorial' quality perceived by the end customer) and not, as usual in technical literature, on quality intended as compliance to technical standards. Finally, a BPR study leverages state-of-art neural network, which enables even more precise commercial quality prediction framework. Future improvements of the study could concern the analysis of pilot projects for the implementation of business process re-engineering described here, together with the prototyping of the system of automatic recognition of information imprinted on hams in an industrial context.

References

Dabbene, F., Gay, P., & Tortia, C. (2014). Traceability issues in food supply chain management: A review. Biosystems engineering, 120, 65-80.

Dora, M., Kumar, M., Van Goubergen, D., Molnar, A., & Gellynck, X. (2013). Food quality management system: Reviewing assessment strategies and a feasibility study for European food small and medium-sized enterprises. Food control, 31(2), 607-616.

Evans, J. R., & Lindsay, W. M. (2002). The management and control of quality (Vol. 5, pp. 115-128). Cincinnati, OH: South-western.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.

García-Esteban, J. A., Curto, B., Moreno, V., & Martín, B. (2018). Quality Control Monitoring by using Artificial Neural Networks for an Iberian Ham Industry. In *ICINCO (2)* (pp. 638-645).

Goyache, F., Bahamonde, A., Alonso, J., López, S., Del Coz, J. J., Quevedo, J. R., ... & Díez, J. (2001). The usefulness of artificial intelligence techniques to assess subjective quality of products in the food industry. *Trends in Food Science & Technology*, *12*(10), 370-381.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT press.

Kelepouris, T., Pramatari, K., & Doukidis, G. (2007). RFID-enabled traceability in the food supply chain. Industrial Management & data systems.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.

Wessel R., San Daniele Prosciutto to Wear RFID Seal of Authenticity, https://www.rfidjournal.com/san-daniele-prosciutto-to-wear-rfid-seal-of-authenticity

Zu, X., Fredendall, L. D., & Douglas, T. J. (2008). The evolving theory of quality management: the role of Six Sigma. Journal of operations Management, 26(5), 630-650.