



## Predictive Microbiology and Machine Learning by Optimization Productive Process: Metanalysis

**Verónica Yepes Medina\***

Bacteriologist, Specialist in Biotechnology and Master's Student in Biosciences, San Pedro, Colombia

**\*Corresponding Author:** Verónica Yepes Medina, Bacteriologist, Specialist in Biotechnology and Master's Student in Biosciences, San Pedro, Colombia.

**DOI:** 10.31080/ASMI.2023.06.1202

**Received:** December 21, 2022

**Published:** January 13, 2023

© All rights are reserved by **Verónica Yepes Medina.**

### Abstract

Unsafe food containing harmful bacteria, viruses, parasites or chemicals can cause more than 200 different illnesses, from diarrhea to cancer. Worldwide, an estimated 600 million (nearly 1 in 10 people) fall ill each year after eating contaminated food, resulting in 420.000 deaths and the loss of 33 million years of healthy life. Therefore, it is necessary to detect and respond to public health threats associated with unsafe food with enabling technologies or tools. Predictive microbiology is concerned with preventing, controlling or limiting the existence of microorganisms by mapping their potential responses to particular environmental conditions, such as temperature, pH, nutrients (protein and fat), water activity (aw) and others. And machine learning as a branch or artificial intelligence learns from these data, identifying patterns for decision making. Recent studies are based in the use of supervised machine learning models to predict the presence of a foodborne pathogenic microorganism at any stage of the production chain, the most commonly used models include Random Forest and support vector machine with rating metrics for accuracy and sensitivity >80%. The main evaluation metrics of the algorithms are: accuracy, F1 score, confusion matrix, sensitivity, specificity and area under the curve (ROC-AUC, Receiver-Operating-Characteristic). Studies have shown that Random Forest was the best model, exhibiting an accuracy of 95% and a F1 score of 98%. Here were evaluated seventeen (17) articles with library meta for of R studio version 4.2.1 and this information provides new opportunities to explore non-destructive models for rapid detection of microorganisms in the production chain.

**Keywords:** Predictive Microbiology; Machine Learning; Food Safety; Foodborne; Pathogens

### Abbreviations

AI: Artificial Intelligence; AUC: Area Under the Curve; AW: Water Activity; BERT: Bidirectional Encoder Representations from Transformers; CART: Classification and Regression Tree; HACCP: Hazard Analysis and Critical Control Point; KNN: K-Nearest Neighbors; ML: Machine Learning; NN: Neural Network; ROC: Receiver-Operating-Characteristic; STEC: Shiga toxin-producing *Escherichia coli*; SVM: Support Vector Machine

### Introduction

The "traditional" approach to food quality and safety management is based on end-product testing. However, end-product testing provides only very limited information about the safety status of a food. The food safety crisis of the 1990s (*Listeria*, *Salmonella*, *Escherichia coli*, *Campylobacter*, dioxins, antibiotics, acrylamide, etc.) highlighted the failure of this traditional approach. If a dangerous organism is found, it means something; but its absence in a limited number of samples is no guarantee of the safety of an entire production batch. Testing of the final product

is often done too late. Today, effective food safety and quality management systems must be based on prerequisite programs (good manufacturing practices and good hygienic practices), on the Hazard Analysis and Critical Control Point (HACCP) plan, as well as on quantitative tools, namely predictive microbiology and risk assessment approaches [1-3].

Accurate and appropriate testing for potential hazards in the food supply chain is key to ensuring consumer safety and food quality. Periodic monitoring and testing of ingredients can reveal fluctuations within the supply chain that may be an indicator of an ingredient's quality or potential hazard. Such quality is assessed through standardized chemical and microbiological composition testing to meet the requirements and specifications of government agencies around the world. For raw materials or finished products to meet these safety and quality limits, their composition generally must have a low microbiological load and be chemically identical in macro components such as carbohydrates, proteins, and fats. Methods in this segment must avoid false negative results that can endanger consumers [4].

A worrying aspect concerning food safety management is the continuous and indiscriminate recourse to microbiological analysis of final products, the production logic assumes that sampling only the final product is enough. It is common to find the following statements: "if I sample all the batches of food I have verified the process", "if I have a deviation in the process, I take more samples", "quality is an expense and not an investment and microbiological test take up to 8 days, since traditional methodologies are used or whoever performs the analysis has no knowledge of how microorganisms work individually and within an biological system" and "microbiology is a Petri dish with some colonies" [5]. As a consequence of the above thoughts, "death certificates", are issued because the product analyzed at the end of its process has no option to be adjusted or modified to meet the necessary requirements for its distribution; tons of food go to a return cellar or enter as "waste" to other links in the chain with millionaire losses. How to make processes more efficient? The answer lies in the combination of several worlds: microbial ecology, understood as the study of the interactions of microorganisms with their environment, with each other and with plant and animal species. Data mining or data exploration is a field of statistics and computer science concerned with the processing of data sets. It uses the methods of artificial

intelligence, machine learning and database systems. And statistics the science of collecting, analyzing, presenting and interpreting data [6]. It sounds complicated, expensive and not very feasible, but no; when we understand the way in which microorganisms develop and the conditions that favor or not their development, we could predict how they can behave within a system and make it scalable using data mining to have patterns over time with significant levels of accuracy that complement the work done in laboratories with traditional, alternative and/or molecular tests. Think about something, to reproduce any living being requires: nutrients, time, temperature and variables such as pH, water activity (aw) and oxygen, if you know these requirements you can design production lines where they are enhanced or controlled to favor or avoid the emergence of certain microbial strains [7]. In the production of yogurt with probiotics of the Bifidobacterium type, it is required that after the maturation tanks, the passage of the yogurt base to flavoring is done without the inclusion of oxygen, since this lactic acid bacteria are strictly anaerobic and the entry of oxygen into the system causes its death [8]. When meat dairy is elaborated, it is natural that the raw material contains *Salmonella* or *Listeria monocytogenes* if it is poultry or bovine origin respectively [9]. The preparation of meat by-products implies the validation of cooking times and temperature as well as cooling times and temperatures for packaging in order to eliminate these two pathogenic bacteria. Generally, the systems are designed so that the heat treatment reaches 75°C in the chamber and is maintained. This is because *Salmonella* achieves a logarithmic reduction to 70°C every one minute and this genus has a thermal resistance of 4.5°C and the temperature must be reduced to 4°C in less than 2 hours since it has optimal growth between 30 to 45°C [10]. Pathogenic bacteria are not always measured in the processes due to cost issues and index microorganisms are used. An increase in the *Escherichia coli* process line correlates with an increase in the probability of *Salmonella* appearance, evolutionarily *Escherichia coli* has a common ancestor with *Salmonella* and conserves pathogenicity genes with this genus, which makes it an excellent indicator for this pathogen [10-12].

Although the concept of predictive modeling was introduced starting in the 1920s with Bigelow's model describing the inactivation kinetics of microorganisms or, alternatively, the heat resistance of microorganisms, it took about 60 years to develop

models of microbial growth and death. Since then, predictive microbiology has focused on the development of primary deterministic models to predict microbial behavior in foods as a function of storage time (growth and survival) and treatment time (inactivation) with the help of secondary models describing the effect of intrinsic, extrinsic or processing factors on kinetic variables. Despite advances in this area, most models developed in the 1990s were based on deterministic approaches without taking into account the variability of factors affecting microbial response. Deterministic growth or inactivation models that produce "optimal" estimates are usually not optimal for successfully managing food safety. If, for instance, the downstream effects of unsatisfactory levels of a surviving pathogen in food after processing design, or if the growth limits of a foodborne pathogen at the single-cell level are not known, product formulation for safer food production would be a difficult issue. In addition, the use of a worst-case scenario approach to food processing leads to unrealistic estimates with a negative impact on food quality [13-15].

The current challenge is how to implement these technologies in the "farm-to-table" industrial chain, since the main limitation of culture-depend methods is that they only allow the study of culturable microorganisms, therefore, there is a risk of not having a view of the biodiversity of the microbial population present in the food [16]. However, the use of open data obtained through different omics and analyzed under models, algorithms or predictive technologies or machine learning allows the extraction of information encrypted from the microbiome and identify if there are pathogenic and/or altering microorganisms. This information provides new opportunities to explore the diversity and functionality of microorganisms in food, as well as in food processing and production environments [17-19].

If we assume that microbiome research is the analysis of the composition, function and biotic and abiotic interactions of microorganisms in complex communities [20]. Current microbiome analysis generally consists of two important components: upstream community profiles (what is the abundance of all microorganisms in each sample) and high-level downstream analysis (alpha/beta diversity analysis, differential abundance analysis). In recent years, evolving data analysis, visualization and machine learning methods have been gradually applied to the development of many software tools and web servers for microbiome data analysis covering these two components [21].

By taking all of the above what we get is the observation of environmental factors, integration of the data into statistical models and prediction of bacterial behavior in food that allow us the identification of hidden patterns, leading to implementation of data-driven interventions. The food industry needs appropriate quality and safety testing technologies to monitor and evaluate indicators in real time before the supply chain reaches the consumer, so that personnel at each link can make timely adjustments to ensure food safety and reduce economic losses [22].

### Predictive microbiology

Predictive microbiology has been defined as the knowledge of microbial responses to environmental conditions, synthesized in mathematical models, which allows an objective assessment of processing, distribution and storage operations on microbiological safety and food quality. Predictive microbiology, which was initially based on three fundamental principles: first, the growth, survival and inactivation of microorganisms are considered reproducible responses; second, microbial behavior is dictated primarily by a limited number of environmental factors; and finally, by quantifying the combined effect of these factors, it is possible to predict the response of microorganisms [23-25].

Predictive models are the most effective mathematical tools for estimating changes in microbial levels in food as the product moves through the farm-to-table chain, and have been recognized as important components in the food safety and quality risk assessment process [26,27]. Databases of microbial responses in food as affected by intrinsic and extrinsic food factors, such as temperature, pH, and water activity (aw), are very useful for the development and validation of predictive models and risk analysis. Between 2016 and 2021, the following predictive microbiology software resources and tools have been released (Table 1) [27].

The rapid progress of microbiology is inseparable from the simultaneous updating of observation methods or technologies. With the advent of the era of big data, the pressing problem for researchers has gradually evolved into how to quickly and efficiently filter/condense this exponentially increasing information to obtain general high-quality data, and how to transform massive microbiome data into easy-to-understand and easy-to-understand data. - to understand the data. visualize the data. Compared to traditional research, where insufficient data

Software	Write	Description	Available
BiOED	R package for optimal experimental design	Includes mathematical algorithms for analysis and optimal experimental design for optimal characterization of microbial inactivation	<a href="https://CRAN.R-project.org/package=biOED">https://CRAN.R-project.org/package=biOED</a>
Pathogens in food	Database of microbial responses in foodstuffs	It contains more than 6000 microbial behavior records and facilitates access and retrieval of data according to different selection criteria, including type of microorganisms, feed category and stage of the production chain	<a href="https://vcadavez.shinyapps.io/MeatProducts/">https://vcadavez.shinyapps.io/MeatProducts/</a>
Base de datos D	Microbial inactivation in food database	It provides scientists with information to build inactivation models. This database covers more than 5000 records of microbial inactivation parameters in food	<a href="https://foodmicrowur.shinyapps.io/Ddatabase/">https://foodmicrowur.shinyapps.io/Ddatabase/</a>
FSMR abierto	Model repository	Community-driven search engine for predictive microbiological models	<a href="https://knime.bfr-berlin/opensfmr/">https://knime.bfr-berlin/opensfmr/</a>
Com Base Premium	Forecasting tool and model repository	Web portal of predictive models and risk management tools to help improve food safety and quality. CB Premium models focus on how microorganisms grow and die in commercial foods	<a href="https://www.cbpremium.org/">https://www.cbpremium.org/</a>
DMRI predictive models for meat	Prediction tool	Includes predictive safety and shelf-life models for various microorganisms in different meat products, validated under static and dynamic conditions	<a href="http://dmripredict.dk/">http://dmripredict.dk/</a>
Listeria Cardenal Model	Prediction tool	This tool allows the simulation of a cardinal meta-analyzed parameter model of <i>L.monocytogenes</i> developed using data derived from growth experiments in 2009	<a href="https://vcadavez.shinyapps.io/Listeria-CardinalModel/">https://vcadavez.shinyapps.io/Listeria-CardinalModel/</a>
Praedicere possumus	Prediction tool	Include predictive models to assess the probability of growth and density of various pathogenic bacteria in food.	<a href="http://praedicere.uniud.it/">http://praedicere.uniud.it/</a>

**Table 1:** Prediction models between 2016 and 2021.

or technology leads to cognitive bias, low repeatability and long error times, the modern microbiology research process is likely to incorporate new technologies and big data methods to make it better and more accurate [28].

**Machine learning**

Machine learning (ML), first proposed by Arthur Samuel in 1959, a core subfield of artificial intelligence (AI), is a proven technique that generates momentum in the domain of microbiology. It has been used so far in computationally intensive problems, such as predicting drug targets and candidate vaccines, diagnosing microorganisms that cause infectious diseases, classifying drug resistance, predicting disease outbreaks and exploring microbial interactions. ML has been applied to all areas of microbiology research, including virology and parasitology [28-30].

Prior to ML, attempts to perform complex tasks such as letter and image recognition involved programming computers with rule-based systems. These systems tended to fail outside of test scenarios with unseen data. ML has shifted the paradigm from programming computers on how to perform a task to one in which computers learn how to perform the task without being programmed and has gained popularity in the era of big data in the scientific community. While mechanistic modeling relies primarily on simplified mathematical formulations to solve problems involving complex data sets, ML algorithms recognize patterns and/or underlying data structures and make predictions based on the patterns learned in those data sets [31]. ML is an extensive field of study that overlaps and inherits ideas from related fields such as statistics, computer science and AI, dedicated to explore to simulate or implement human learning behaviors, reorganize existing knowledge structures and continuously improve their performance to predict or classify other unknown data [22,30,32].

ML has two main learning modes, supervised or predictive to make future predictions from training data and unsupervised or descriptive, which is exploratory in nature with no training data, defined target or output. The goal of supervised ML is to find a model that best matches the input and output data. For instance, an image of a Petri dish containing a culture medium could be considered as “growth present” or “growth absent”. A computer develops an algorithm to classify each image and compares the accuracy of the algorithm to the “supervised truth”; typically, the initial results do not provide adequate performance and the algorithm will iteratively adjust its parameters by trial and error to optimize accuracy [33]. Another example of supervised ML is its effectiveness in attributing food sources to clinical cases of foodborne diseases such as listeriosis from sequencing data. Inferring genetic information from pathogen genotypes is often crucial biological inference. Attribution of the origin of *Listeria monocytogenes* infections allows food industry professionals, data managers, epidemiologists, microbiologist, and bioinformaticians to adapt their practices to prevent the spread of foodborne pathogens. It also enables more efficient use of resources to contain the survival and proliferation of pathogens at their source [34].

The use of ML algorithms in food safety since 2020 has increased in response to the complexity of quantifying some features and their effects on microbial population dynamics [35]. Table 2 discriminates the studies between 2020 and 2022 published in the food safety domain using ML.

Tanui, Karanth., *et al.* 2022. used ML algorithms to predict the presence of *Salmonella* in ground chicken meat. Using known genotypes as input features, the semi-supervised Random Forest showed the highest overall accuracy of 94% (95% confidence interval: 85%-99%) and a Kappa value of 82% and 87% prediction. The model predicted virulence-associated genes that could be used as features in predictive modeling efforts in the future, demonstrating its utility in different areas of food safety, including pathogen source identification, antibiotic resistance prediction and foodborne pathogen assessment.

Im., *et al.* 2021 developed several ML models to predict the pathogenicity of Shiga toxin-producing *Escherichia coli* (STEC) using whole genome sequencing data. The input dataset for

the ML models was revealed using distinct gene repertoires of positive (pathogenic) and negative (non-pathogenic) control groups in which each STEC isolate was designated based on source attribution, risk potential relative to isolate sources. Among the ML models examined, the support vector machine (SVM) model discriminated between the two groups most accurately. The SVM model successfully predicted the pathogenicity of isolates from major sources of STEC outbreaks, suggesting that the SVM model is a reliable and widely applicable method for assessing the pathogenic potential of STEC isolates compared to conventional methods.

Huang, *et al.* 2022 conducted a study to define the growth and non-growth limit of *Listeria monocytogenes* under the use of different preservatives used in the meat industry and pH as control factors. They developed a multifactorial logistic regression model to calculate the probability of growth and then define the growth limit using 2 thresholds. With threshold 1 ( $p = 0.0104$ ), the classification accuracy for growth events was 68,6% with a true positive rate of 77.6% and a true negative rate of 45.5% for growths obtained on PALCAM agar. At threshold 2 ( $p = 0.04$ ) the accuracy becomes 82.6% with a true positive rate of 95,5% and a true negative rate of 80%. This information can be used to formulate ready-to-eat-products that can prevent the growth of *Listeria monocytogenes* if stored at the optimum temperature.

Different algorithms have their own advantages and disadvantages, and there is no distinction between advantages and disadvantages. What needs to be done is to fully interpret the input data according to different demand scenarios, then build an appropriate model and constantly adjust it to achieve the best performance. Moreover, the belief that scientific problems will be solved only by using the most sophisticated and sophisticated models is not objective. Basically, computer technology simply helps people make decisions or automates the decision-making process and improves efficiency. Therefore, the choice of model should be the most suitable, not the most complex. There are four criteria to evaluate the advantages and disadvantages of machine learning algorithms [38] (1) Correctness, the most important criterion to evaluate the advantages and disadvantages of an algorithm (2) Robustness, that is, error tolerance, which represents the algorithm’s response and handling of unreadable data input.

Chance. (3) Readable, easy-to-understand algorithms mean less time-consuming debugging, modification and extension processes. (4) Actuality, that is, time complexity and space complexity, indicate the amount of computation and memory space required to run the algorithm, respectively [28].

Advances and applications of new experimental and computational methods drive the integration of microbial ecology research with cutting-edge technologies in computational biology and other fields. Simplify the scientific problem by predicting relationships between species or their environment (natural diversity, life cycles, interactions, and co-evolution) by altering prediction strategies based on species characteristics.

ML research is evolving rapidly, with rapidly changing architectures, algorithm combinations and computational strategies. The ultimate goal is not only to predict task accuracy, but also to discover the biological processes underlying scientific problems. The development of effective analytical tools, including ML, ensures data validity, proper annotation and open sharing and that most studies born from the intersection of microbiology and ML show result.

Application	Automated Learning Algorithm	Accuracy	Data type	Reference
Food Safety and Quality	Support Vector Machine (SVM)	100%	Image	[32]
	Nearest neighbors (KNN)	90%		
	Naive Bayes	80%		
Specific Food Safety Hazards. Biological Hazards.	Random Forest	99.3%	Open data (genomic)	[31]
	Logit boost (LB)	100%		
	Gradient boosting (GB)	99.3%		
	Support Vector Machine with Linear Kernel	98.7%		
	Support Vector Machine with Radial Kernel	97.4%		
Specific Food Safety Hazards. Biological Hazards.	Support Vector Machine (SVM)	84%	Genomic monitoring data	[36]
	Random Forest	84%		
	Digital Twins (DTs)	84%		
	Naive Bayes	84%		
Food Safety General	Random Forest	80,20%	Text	[39]
	Naive Bayes	85%		
	Logistic regression	79,70%		
	Gradient boosting (GB)	81,60%		
	Neural Network (NN)	80,30%		
	Support Vector Machine (SVM)	86,95		
	Adaboost (AB)	79.5%%		
	Bidirectional encoder representations from transformers (BERT)	85,50%		
Specific Food Safety Hazards. Biological Hazards.	Logistic regression multifactor	68,60%	Monitoring data	[37]
Food Safety General	Neural Network (NN)	82,50%	Monitoring data	[40]
	Kernel logistic regression			

Specific Food Safety Hazards. Biological Hazards.	Random Forest	78,30%	Open data (genomic)	[41]
	Logit boost (LB)	93,30%		
Food Safety and Quality	Adaboost(AB)	92,41%	Image	[42]
Specific Food Safety Hazards. Biological Hazards.	Support Vector Machine with Linear Kernel	89%	Open data (genomic)	[43]
	Random Forest	89%		
	Support Vector Machine with Radial Kernel	70%		
	Gradient boosting (GB)	88%		
	Logit boost (LB)	86%		
Specific Food Safety Hazards. Biological Hazards.	Logistic Regression	76%	Open data (genomic)	[44]
	Random Forest	72%		
	Support Vector Machine with Linear Kernel	70%%		
	Adaboost(AB)	72%		
Specific Food Safety Hazards. Biological Hazards.	Classification and Regression Tree (CART)	63%	Open data	[45]
	Random Forest	64%		
	Gradient boosting (GB)	69%		
	Adaboost (AB)	67%		
Food Safety General	Neural Network (NN)	96,25%	Image	[22]
	Support Vector Machine (SVM)	90%		
Water for agriculture. Biological Risk	Random Forest	83.2%	Monitoring Data	[46]
	Neural Network (NN)	86.1%		
	Support Vector Machine (SVM)	77.3%		
	Naive Bayes	78.7%		
Food safety. Biological Risk	Random Forest	72.2%	Open data and monitoring data	[34]
	Logit boost (LB)	73.2%		
	Stochastic gradient	70.1%		
	Support Vector Machine (SVM)	61.4%		

**Table 2:** Studies between 2020 and 2022 published in the food safety domain using ML.

**Materials and Methods**

As a statistical analysis of a large collection of published study results, meta-analysis aims to integrate and interpret findings to achieve comprehensive conclusions that individual studies alone would not clearly demonstrate. An electronic literature search was conducted in Web of Science and PubMed databases to find

articles and official reports published since 2018 that summarized the prediction of pathogenic microorganisms, mainly *Listeria monocytogenes* and *Salmonella* spp through supervised machine learning algorithms at any stage of food production globally. The search was performed systematically and aimed to find quality studies validated by the scientific community.

The literature reviews were performed using a syntax combining terms on the existence (prevalence, incidence, occurrence, quality, safety, food safety) of pathogens (*Salmonella* spp, *Listeria monocytogenes*, *Campylobacter*, *Escherichia coli* STEC) in food and their prediction through artificial intelligence (machine learning, data mining, models, algorithms, prediction), while other studies in the clinical domain, on insects and/or parasites, where products were artificially inoculated (inoculo, enriched) and published before 2018 were excluded. All terms were combined by correctly applying AND, OR and AND NOT logical connectors.

In the initial assessment, 323 publications were retrieved in the Web of Science database and 151 in the PubMed database, after the title review it was identified that the 151 PubMed articles were included in the 323 retrievals made in Web of Science, so we continued with this bibliography.

After assessing all the information in the retrieved publications, seventeen (17) primary studies published from 2019 to 2022 were deemed appropriate for inclusion as they met the following criteria: (1) presence of pathogenic microorganisms of public health importance; (2) present extractable data; (3) results on supervised machine learning algorithm scoring metrics: accuracy, F1 score, area under the curve (ROC-AUC), sensitivity, specificity

and/or confusion matrix; and (4) results at any stage of food production globally.

The initial treatment of the 17 articles for the execution of the meta-analysis was performed with the creation of a matrix in Microsoft Excel 2010 whose main variables were: Reference, algorithms, model qualification metrics accuracy, F1 score, sensitivity, specificity, area under the curve and Cohen's Kappa coefficient, in addition to the type of data and application in food safety

### Results and Discussion

A review of 17 articles associated with the use of machine learning with application in food safety was carried out in order to compare the information obtained in the literature, using the *metafor* program of the R software version 4.2.1 to perform the metanalysis.

### Metanalysis

The synthesis of the results of different previous studies (17) on the use of automated learning algorithms on the prediction of pathogenic microorganisms as a biological risk within food safety for the metrics: accuracy, sensitivity, specificity, ROC-AUC, Cohen's Kappa coefficient and F1 score.

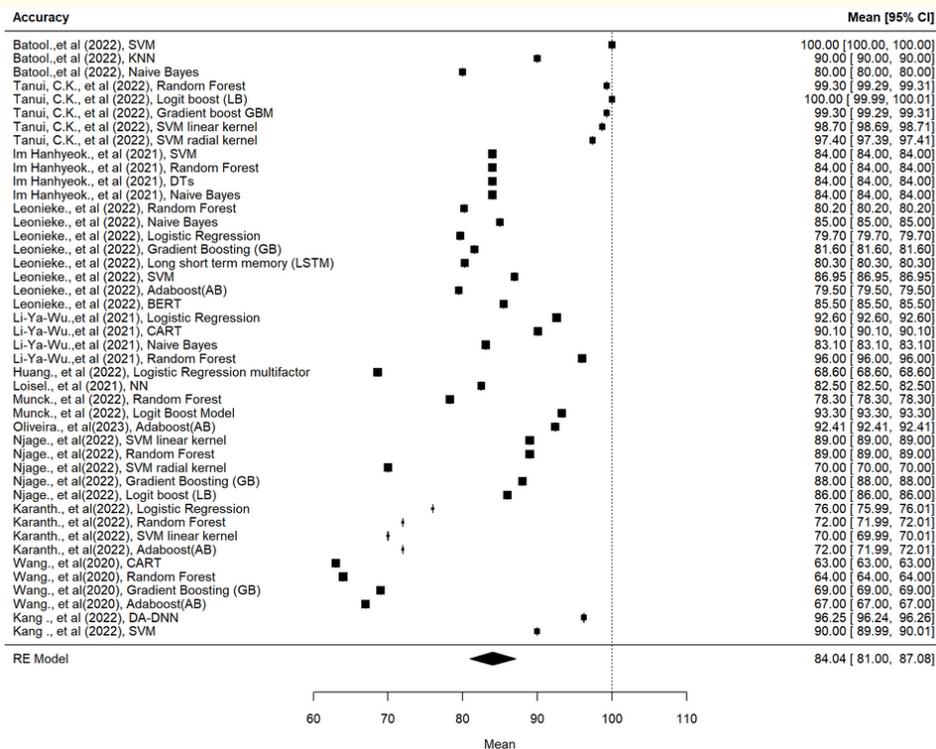


Figure 1: Results accuracy's metanalysis.

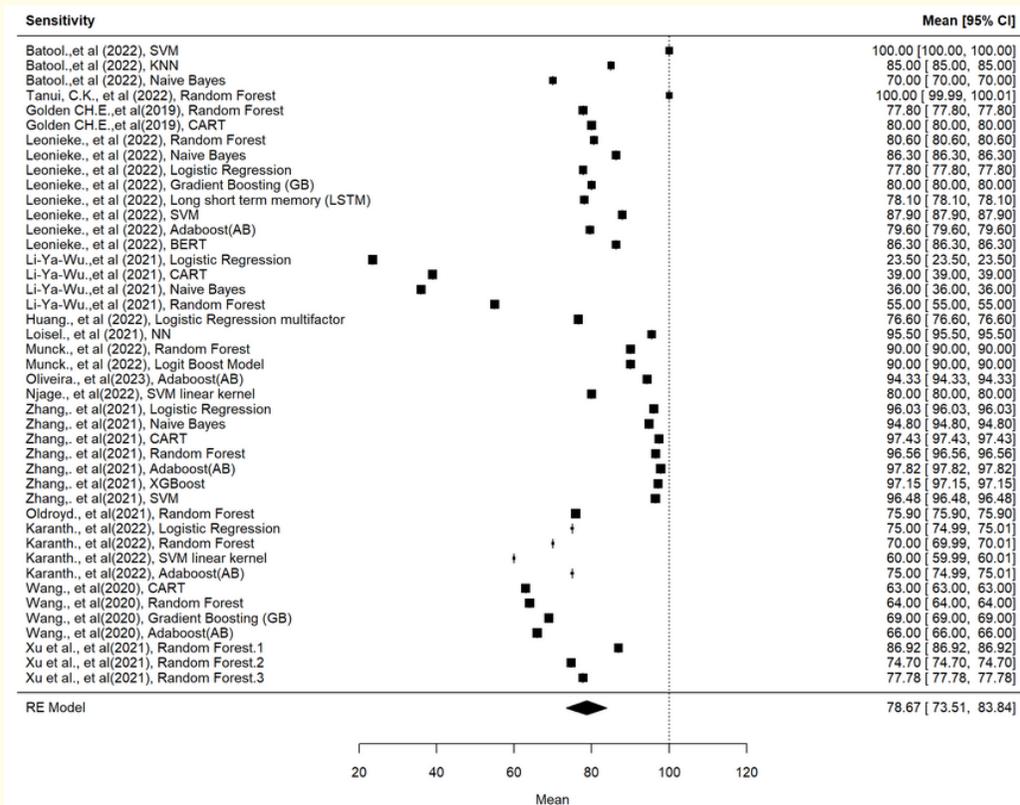


Figure 2: Sensitivity’s metanalysis.

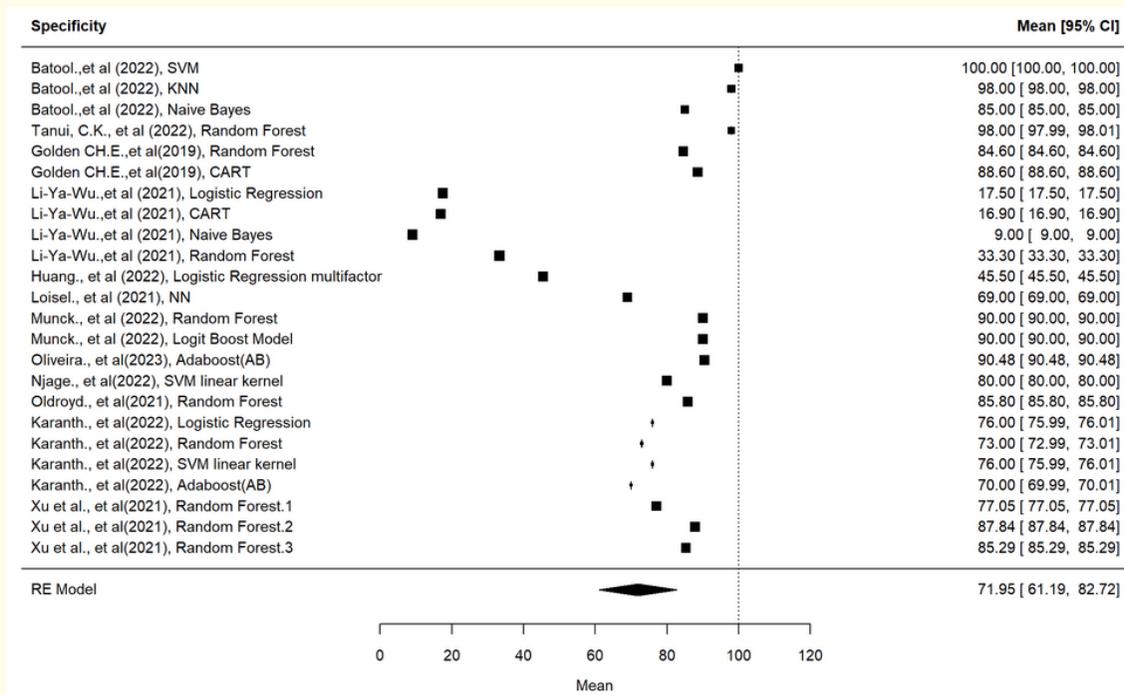


Figure 3: Specificity’s metanalysis.

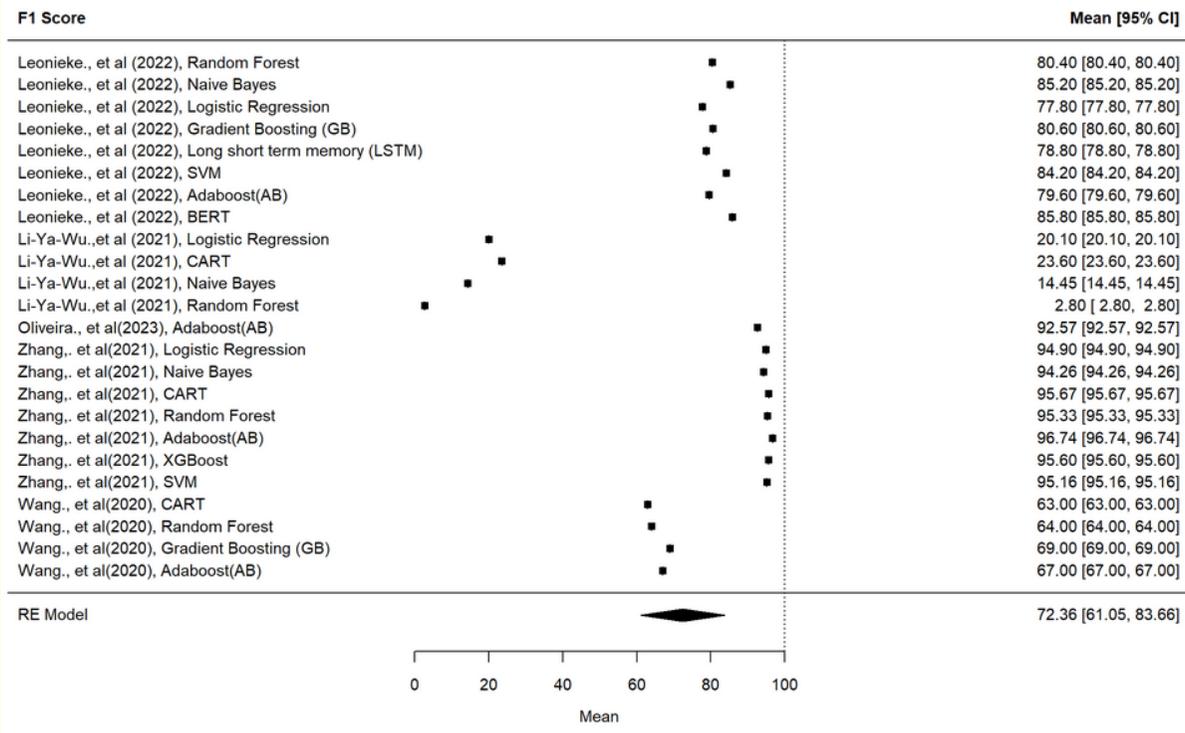


Figure 4: F1 score’s metanalysis.

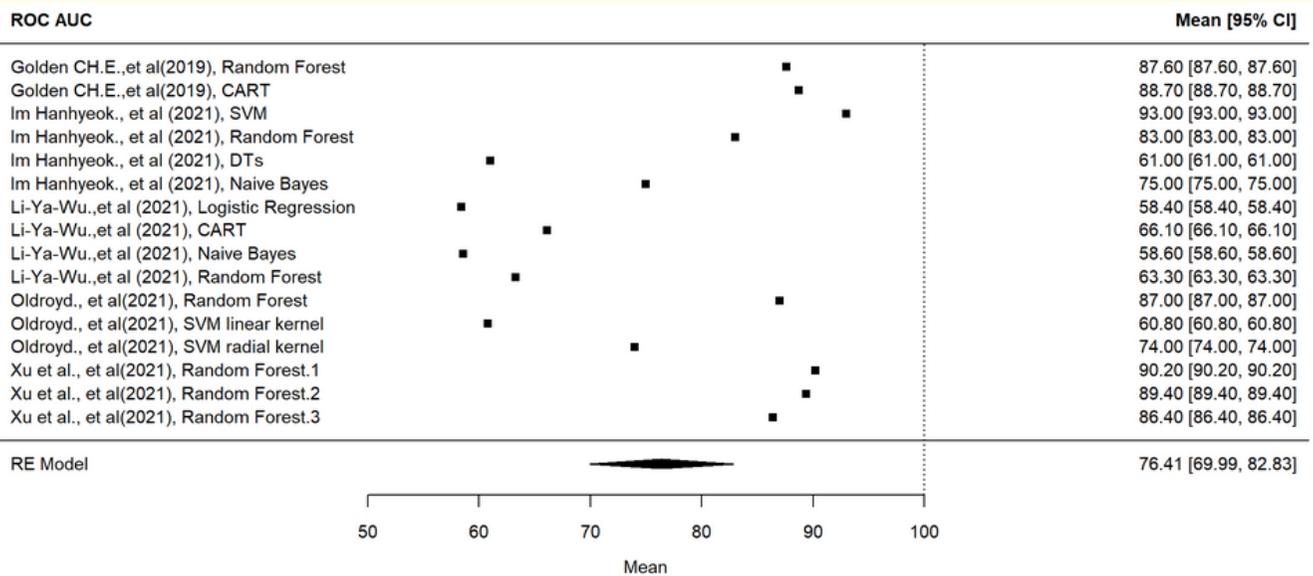


Figure 5: ROC-AUC metanalysis.

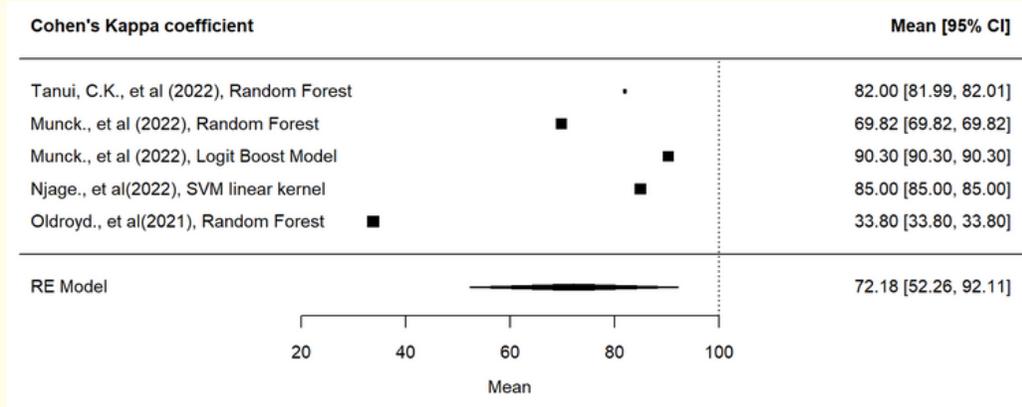


Figure 6: Cohen’s Kappa coefficient’s metanalysis.

Under a restricted maximum likelihood estimation from the “n” of each item and the mean of each metric for the different machine learning algorithms, the accuracy obtained for predicting the presence of a pathogenic of 78.67% with an interval between 73.51-83.84%. For specificity 71.95% with an interval between 61.19-82.72%. An estimate for the area under the curve a value of 76.41% (interval 69.99-82.83% and for the F1 score 72.35% (interval 61.0-83.66%).

### Conclusion

With these preliminary results, it was evident that machine learning algorithms are high impact tool in food safety and with greater data inclusion can deliver the best model to predict microbiological behaviors of pathogens of public health importance without the limitations of conventional culture-dependent microbiology. Since the coarse, overfitting occurs in the machine learning model.

We have an initial robust base to continue applying these models with more data to validate if the classification performed is consistent with these preliminary data and in the future have a model that can be presented to the food industry for its application in situ.

Supervised machine learning offers the opportunity to improve production processes, it does not eliminate conventional microbiology tests for process verification, it becomes an ally of food safety to predict behaviors at specific stages and obtain solutions in advance.

### Acknowledgments

Thanks to my daughter Maria José Peñate Yepes for her patience while I write what I am most passionate about in life, about microbiology. Thanks to my friend David Morales Rave for teaching me what he knows about programming and machine learning.

### Conflict of Interest

The author declare that she has no conflicts of interest.

### Bibliography

1. S N Garcia., *et al.* “A one health perspective on dairy production and dairy food safety”. *One Health* 7 (2019).
2. Committee on the Review of the Use of Scientific Criteria and Performance Standards for Safe Food. and Institute of Medicine (U.S.), Scientific criteria to ensure safe food. National Academies Press, (2003).
3. Z Liu., *et al.* “Food safety governance in China: From supervision to coregulation”. *Food Science and Nutrition* 7.12 (2019): 4127-4139.
4. K L Beck., *et al.* “Monitoring the microbiome for food safety and quality using deep shotgun sequencing”. *NPJ Science of Food* 5.1 (2021).
5. L Pérez-Lavalle., *et al.* “Microbiological criteria: In establishment principles food quality for their and application safety”. *The Italian Journal of Food Safety* 9.1 (2020).

6. I H Sarker. "Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective". *SN Computer Science* 2.5 (2021).
7. M E Hibbing, *et al.* "Bacterial competition: Surviving and thriving in the microbial jungle". *Nature Reviews Microbiology* 8.1 (2010): 15-25.
8. P M Fratamico, *et al.* "Foodborne pathogens : microbiology and molecular biology". Caister Academic Press (2005).
9. A Gonçalves-Tenório, *et al.* "Prevalence of pathogens in poultry meat: A meta-analysis of European published surveys". *Foods* 7.5 (2018).
10. L I Ma, *et al.* "Development of Thermal Surrogate Microorganisms in Ground Beef for In-Plant Critical Control Point Validation Studies". (2007).
11. N M Vega, *et al.* "Salmonella typhimurium intercepts Escherichia coli signaling to enhance antibiotic tolerance". *Proceedings of the National Academy of Sciences of the United States of America* 110.35 (2013): 14420-14425.
12. K Wessels, *et al.* "Salmonella in chicken meat: Consumption, outbreaks, characteristics, current control methods and the potential of bacteriophage use". *Foods* 10.8 (2021).
13. S Koseki. "Predictive Modeling for Estimation of Bacterial Behavior from Farm to Table". *Food Safety* 4.2 (2016): 33-44.
14. E Stavropoulou and E Bezirtzoglou. "Predictive modeling of microbial behavior in food". *Foods* 8.12. MDPI Multidisciplinary Digital Publishing Institute (2019).
15. TA Mcmeekin, *et al.* "Predictive microbiology: Past, present and future". in *Modelling Microorganisms in Food*, Elsevier Ltd, (2007): 7-21.
16. C Sabater, *et al.* "Novel methods of microbiome analysis in the food industry". *International Microbiology*. Springer Science and Business Media Deutschland GmbH 24.4 (2021): 593-605.
17. K Melnyk, *et al.* "GraphKKE: graph Kernel Koopman embedding for human microbiome analysis". *Applied Network Science* 5.1 (2020).
18. G Cammarota, *et al.* "Gut microbiome, big data and machine learning to promote precision medicine for cancer". *Nature Reviews Gastroenterology and Hepatology* 17.10 (2020): 635-648.
19. T W Whon, *et al.* "Omics in gut microbiome analysis". *Journal of Microbiology* 59.3 (2021): 292-297.
20. E Sakowski, *et al.* "Current State of and Future Opportunities for Prediction in Microbiome Research: Report from the Mid-Atlantic Microbiome Meet-up in Baltimore on 9 January 2019". *mSystems* 4.5 (2019).
21. Y Zhao, *et al.* "Animalcules: interactive microbiome analytics and visualization in R". *Microbiome* 9.1 (2021).
22. Z Kang, *et al.* "Advances in Machine Learning and Hyperspectral Imaging in the Food Supply Chain". *Food Engineering Reviews*. Springer (2022).
23. S Koseki, *et al.* "Recent advances in predictive microbiology: theory and application of conversion from population dynamics to individual cell heterogeneity during inactivation process". *Current Opinion in Food Science* 39 (2021): 60-67.
24. SC González, *et al.* "MicroHibro': A software tool for predictive microbiology and microbial risk assessment in foods". *International Journal of Food Microbiology* 290 (2019): 226-236.
25. E Chaix, *et al.* "Predictive Microbiology Coupled with Gas (O<sub>2</sub>/CO<sub>2</sub>) Transfer in Food/Packaging Systems: How to Develop an Efficient Decision Support Tool for Food Packaging Dimensioning". *Comprehensive Reviews in Food Science and Food Safety* 14.1 (2015): 1-21.
26. K P Koutsoumanis, *et al.* "Last developments in foodborne pathogens modeling". *Current Opinion in Food Science* 8 (2016): 89-98.
27. A Possas, *et al.* "New software solutions for microbiological food safety assessment and management". *Current Opinion in Food Science* 44 (2022).
28. Y Jiang, *et al.* "Machine Learning Advances in Microbiology: A Review of Methods and Applications". *Frontiers in Microbiology* 13 (2022).
29. C E Golden, *et al.* "Using Farm Practice Variables as Predictors of Listeria spp. Prevalence in Pastured Poultry Farms". *Frontiers in Sustainable Food Systems* 3 (2019).
30. S J Goodswen, *et al.* "Machine learning and applications in microbiology". *FEMS Microbiology Reviews* 45.5 (2021).
31. C K Tanui, *et al.* "Machine learning-based predictive modeling to identify genotypic traits associated with Salmonella enterica disease endpoints in isolates from ground chicken". *LWT* 154 (2022).

32. A Batool., *et al.* "An IoT and Machine Learning-Based Model to Monitor Perishable Food towards Improving Food Safety and Quality". *Journal of Food Quality* 2022 (2022).
33. E M Mateo., *et al.* "Comparative analysis of machine learning methods to predict growth of *f. Sporotrichioides* and production of t-2 and ht-2 toxins in treatments with ethylene-vinyl alcohol films containing pure components of essential oils". *Toxins* (Basel) 13.8 (2021).
34. C K Tanui., *et al.* "A Machine Learning Model for Food Source Attribution of *Listeria monocytogenes*". *Pathogens* 11.6 (2022).
35. S Hiura., *et al.* "Prediction of population behavior of *Listeria monocytogenes* in food using machine learning and a microbial growth and survival database". *Scientific Report* 11.1 (2021).
36. H Im., *et al.* "Pathogenic potential assessment of the Shiga toxin-producing *Escherichia coli* by a source attribution-considered machine learning model". (2021).
37. L Huang., *et al.* "Growth and No-Growth boundary of *Listeria monocytogenes* in beef - A logistic modeling". *Food Research International* 152 (2022).
38. J G Greener., *et al.* "A guide to machine learning for biologists". *Nature Reviews Molecular Cell Biology* 23.1". *Nature Research* 40-55 (2022).
39. L M van den Bulk., *et al.* "Automatic classification of literature in systematic reviews on food safety using machine learning". *Current Research in Food Science* 5 (2022): 84-95.
40. J Loisel., *et al.* "Cold chain break detection and analysis: Can machine learning help?". *Trends in Food Science and Technology* 112 (2021): 391-399.
41. N Munck., *et al.* "Application of Whole-Genome Sequences and Machine Learning in Source Attribution of *Salmonella Typhimurium*". *Risk Analysis* 40.9 (2020): 1693-1705.
42. AN de Oliveira., *et al.* "Tomato classification using mass spectrometry-machine learning technique: A food safety-enhancing platform". *Food Chemistry* 398 (2022).
43. P M K Njage., *et al.* "Quantitative microbial risk assessment based on whole genome sequencing data: Case of *listeria monocytogenes*". *Microorganisms* 8.11 (2020): 1-24.
44. S Karanth., *et al.* "Exploring the predictive capability of advanced machine learning in identifying severe disease phenotype in *Salmonella enterica*". *Food Research International* 151 (2022).
45. Z Lin., *et al.* "Machine learning approach for predicting single cell lag time of *Salmonella Enteritidis* after heat and chlorine treatment". *Food Research International* 156 (2022).
46. S Buyrukoğlu. "New hybrid data mining model for prediction of *Salmonella* presence in agricultural waters based on ensemble feature selection and machine learning algorithms". *Journal of Food Safety* 41.4 (2021).