Ann Ig. 2025 Jan-Feb; 37(1): 128-140 doi: 10.7416/ai.2024.2644. Epub 2024 Jul 11.

Machine learning *vs*. regression models to predict the risk of *Legionella* contamination in a hospital water network

Osvalda De Giglio[1], Fabrizio Fasano[1], Giusy Diella[1], Valentina Spagnuolo[1,2], Francesco Triggiano[1], Marco Lopuzzo[1,2], Francesca Apollonio[1], Carla Maria Leone[3], Maria Teresa Montagna[1]

Keywords: Legionella; Machine learning; Water network; Hospital; Artificial Intelligence
Parole chiave: Legionella; Machine Learning; Rete idrica; Ospedale; Intelligenza artificiale

Abstract

Introduction. *The periodic monitoring of Legionella in hospital water networks allows preventive measures to be taken to avoid the risk of legionellosis to patients and healthcare workers.*
Study design. *The aim of the study is to standardize a method for predicting the risk of Legionella contamination in the water supply of a hospital facility, by comparing Machine Learning, conventional and combined models.*
Methods. *During the period July 2021– October 2022, water sampling for Legionella detection was performed in the rooms of an Italian hospital pavilion (89.9% of the total number of rooms). Fifty-eight parameters regarding the structural and environmental characteristics of the water network were collected. Models were built on 70% of the dataset and tested on the remaining 30% to evaluate accuracy, sensitivity, and specificity.*
Results*. A total of 1,053 water samples were analyzed and 57 (5.4%) were positive for Legionella. Of the Machine Learning models tested, the most efficient had an input layer (56 neurons), hidden layer (30 neurons), and output layer (two neurons). Accuracy was 93.4%, sensitivity was 43.8%, and specificity was 96%. The regression model had an accuracy of 82.9%, sensitivity of 20.3%, and specificity of 97.3%. The combination of the models achieved an accuracy of 82.3%, sensitivity of 22.4%, and specificity of 98.4%. The most important parameters that influenced the model results were the type of water network (hot/cold), the replacement of filter valves, and atmospheric temperature. Among the models tested, Machine Learning obtained the best results in terms of accuracy and sensitivity.*
Conclusions. *Future studies are required to improve these predictive models by expanding the dataset using other parameters and other pavilions of the same hospital.*

[1] Interdisciplinary Department of Medicine, Hygiene Section, University of Bari Aldo Moro, Bari, Italy
[2] Department of Precision and Regenerative Medicine and Ionian Area (DiMePre-J), University of Bari Aldo Moro, Bari, Italy
[3] Azienda Ospedaliero Universitaria Policlinico di Bari, Hygiene Section, Bari, Italy

Annali di Igiene : Medicina Preventiva e di Comunità (Ann Ig)
ISSN 1120-9135 https://www.annali-igiene.it

Introduction

Legionella are Gram-negative bacteria that can colonize natural (e.g., rivers, lakes, and ponds) and artificial aquatic environments (e.g., drinking water systems, taps, faucets, showers, cooling towers, and fountains) (1). After individuals inhale contaminated aerosols, they can develop various clinical forms of legionellosis, such as a flu-like illness (Pontiac fever) or severe pneumonia known as Legionnaires' disease (LD) (2). The disease can be of community or nosocomial origin. In recent years, nosocomial legionellosis has attracted particular attention because of the complexity of hospital water systems and the vulnerability of hospitalized patients, which can lead to serious consequences with a high mortality rate (3).

The World Health Organization proposed the Water Safety Plan (WSP) in 2004 and revised it in subsequent years to both organize and systematize drinking water management practices and ensure the applicability of these practices to drinking water quality management (4,5). Additionally, according to the new European Drinking Water Directive (6) transposed in Italy on 18 February 2023 (7), *Legionella* is a microbiological parameter to be detected in the water supply of health and community facilities.

In recent years, LD cases have increased overall, probably because of the systematic surveillance developed in many countries and improved testing in microbiology laboratories (8-10). The Centers for Disease Control and Prevention (CDC) has estimated that 90% of outbreaks could be prevented through safe water management programs (11).

The ability to colonize various natural and artificial ecosystems makes the eradication of these microorganisms difficult (12). Several factors favor the proliferation of *Legionella*, including a water temperature between 20°C and 50°C (13) and the stagnation of water inside pipes (14). Moreover, *Legionella* can parasitize freshwater protozoa and persist in biofilm, thereby allowing for greater resistance to environmental factors and remediation treatments (12). Some authors (15) have highlighted the importance of chemical parameters (hardness, free chlorine concentration, pH, and trace element concentrations) and the material of water system pipes. For example, copper pipes reduce the risk of water colonization because of the natural antimicrobial effect of copper (16). More recently, competition with *Pseudomonas aeruginosa* has also been considered. Indeed, some researchers have reported that the presence of *P. aeruginosa* in the water supply is inversely correlated with the presence of *Legionella* (17).

Water is not free from microorganisms and poorly managed water networks can be particularly vulnerable to *Legionella* (18). In healthcare facilities, the management of construction activities is particularly complicated because of the complexity and variability of the buildings (age and size, time since the last renovation, number of floors, and number of rooms and water points/floors), which are often outdated and no longer suitable for current organizational and healthcare practices (19).

To date, conventional statistical methods and models for *Legionella* risk in water networks have been limited and often difficult to implement in practice (20-22).

In recent years, innovative artificial intelligence (AI) models, such as machine learning (ML)/deep learning (DL), have achieved tremendous success worldwide in various fields (23,24); however, there is still little scientific evidence on their application to risk caused by *Legionella* (25-27).

The term "artificial intelligence" was coined in the 1950s and describes a machine's capacity, particularly computer systems, to conduct operations that ordinarily require human intellect (e.g., visual perception, speech recognition, and decision-making) (28,29). ML is a branch of AI that uses algorithms to give machines the ability to learn from data (input) and improve over time without human help. DL is a subfield of ML and AI that uses artificial neural networks to simulate the cellular behavior of the human brain and learns from its experience. However, a massive volume of data needs to be provided at input (30).

The aim of the present study is to standardize a method for predicting the risk of *Legionella* contamination in the water supply of a hospital facility, by comparing ML, conventional models, and combined models.

Methods

Study design

The study was conducted in an Italian hospital, structured into several pavilions, which has implemented a WSP since October 2020. For this purpose, a systematic and organized water network monitoring process was planned, with associated differentiated maintenance interventions, derived from the analysis of the risk of water contamination by microorganisms, including *Legionella*.

For this study, a seven-floor pavilion (12,800 m^2) organized into two wings (north and south) was considered because of the plant scenario and the related maintenance interventions. One of the wings underwent a complete renovation of the network in the period March–June 2021, whereas the other wing did not undergo any extraordinary maintenance interventions. The pavilion had 396 rooms equipped with taps, showers, and bidets. The water network developed into five lines and 33 risers, characterized by a very varied structure in terms of installations (e.g., some sections were underground and others above ground, the presence of dead-end branches) and the characteristics of the water pipes (e.g., type of material, presence of filters, mixers). Technical data provided by the hospital's technical equipment and microbiological data were collected and analyzed to build predictive models of *Legionella* contamination in the water supply.

Legionella survey

Between July 2021 and October 2022, 356 of the 396 rooms (89.9%) present in the pavilion (99% confidence level, 2.2% confidence interval) were monitored for *Legionella* detection. A total of 1,053 water samples were analyzed (all samples for the wing of the pavilion under renovation were taken after the extraordinary maintenance intervention).

Water samples (1 L) were collected in sterile dark containers containing sodium thiosulphate pentahydrate (0.01%, w/v) to neutralize the chloride present in the water, transported to room temperature in isothermal bags, and analyzed within 24 hours according to current regulations (31,32). The water was filtered through a polycarbonate membrane with 0.2-μm pores and a diameter of 47 mm (Millipore Corporation, Bedford, MA, USA), and then suspended in 10 mL of the same water sample and vortexed. Subsequently, 200 μL of each sample was seeded on plates containing *Legionella* selective agar (GVPC, Biolife Italiana Srl, Milan, Italy) and incubated at 36°C ± 2°C for 7–10 days in a humid environment. Quantitative evaluation was expressed in colony-forming units/liter (cfu/L). Suspect colonies were subcultured on two *Legionella* BCYE agars (Biolife Italiana Srl, Milan, Italy) with and without L-cysteine. Colonies grown only on BCYE cysteine agar plates were considered to belong to the genus *Legionella* and were identified for confirmation in latex agglutination tests with polyvalent (Biolife Italiana Srl, Milan, Italy) and monovalent (Biogenetics Srl, Tokyo, Japan) antisera.

Water samples containing < 50 cfu/L were considered negative (hereafter referred to as 0 cfu/L).

Data collection

A total of 58 parameters relating to the structural and environmental characteristics of the water network and pavilion were studied, and are listed schematically below:

- structural parameters of the pavilion: floor and wings;
- infrastructural parameters of the water network up to the point of supply: length, location (underground and above ground), material of tube pipes (copper, steel, and multi-layer), diameters of tube pipes measured in mm (63.5, 50.8, 38.1, 31.75, 25.4, 19.05, 12.7, 10.16, 8.128, 6.604, 5.08, 4.572, 4.064, 3.556, 3.302, and 3.048), average pipe diameter, number of diameter changes along the water mains route, number of network lines, number of risers, type of network (hot/cold), presence of dead-end branches, and presence of corners along the water mains route (both total and partial);
- parameters of the water supply points in the rooms: number of water supply points used and not used, and type (tap, shower, and bidet);
- parameters of water network maintenance: total network renovation (Yes/No), days since renovation, replacement of filter valves (Yes/No) and days since the last replacement, replacement of aerator filters (Yes/No) and days since the last replacement, replacement of mixers (Yes/No) and days since the last replacement, replacement of shower heads (Yes/No) and days since the last replacement, replacement of flexible hoses (Yes/No) and days since the last replacement, disinfection of the network with sodium hypochlorite for two days (Yes/No) and days since the last disinfection, and presence of an absolute filter at the distribution point (Yes/No) and days since the installation of the absolute filter;
- water sampling parameters: water temperature at the time of sampling, pre- or post-flush sampling method, and detection (positive/negative) and load (cfu/L) of *Legionella*.
- climatic parameters: month of sampling, average air temperature on the day of sampling, and temperature range recorded on the day of sampling (33).

Statistical analysis

The development of the models involved the following steps:

• To make the descriptive parameters comparable with the alphanumeric parameters and include them in the analysis, the ordinal coding technique was used (34).

• The data obtained from the 1,053 water samples analyzed were pre-randomized.

• The independent parameters were normalized to a single comparable unit of measurement (range 0–1) using the following formula (35):

$$Xn= (Xnn-Min(X))/Max(X)-Min(X),$$

where:

Xn is the normalized value of each variable for record n
Xnn is the non-normalized value of each variable for record n
Max(X) is the maximum value of each variable
Min(X) is the minimum value of each variable.

The entire dataset was divided into two parts: 70% to train the model and 30% to evaluate the quality of the model (testing phase) (36). Furthermore, to test the robustness of the model another partition of the dataset was used: 50% for training and 50% for testing (37).

Development of predictive models

A useful glossary table consisting of commonly used terms in predictive modeling can be found in Table 1 (38-44).

Several ML/DL models were tested to predict the risk of *Legionella* contamination in the water network. All ML/DL models were developed considering the training dataset (70% and 50%). The number of hidden layers, number of neurons within each layer, and model training parameters (model training epochs, batch size, and validation split) were modified.

The purpose of these models was to understand which independent variables (n = 57) influence the dependent variable "*Legionella* detection (positive/ negative)".

Each ML/DL model was supervised and adapted to solve classification problems (prediction of the *Legionella* sampling results, positive or negative). A bias neuron was added to each layer of the models to increase their effectiveness. The activation function for each layer was ReLU (Rectified Linear Unit); for the last level, which was a classification problem, softmax was considered (45,46).

For each model, the confusion matrix was calculated on 30% and 50% of the test dataset, which allowed us to understand how correctly the model was able to predict the sampling results compared with the real data present in the dataset. Through these confusion matrices, it was possible to define both the accuracy of the model, and its sensitivity and specificity (47).

Additionally, the R package Variable Importance Plots (VIP), which is a permutation based VI scoring method, was used to evaluate which factors most influenced the dependent variables within each model (44).

Poisson regression model

To estimate which parameters can predict water contamination by *Legionella*, the inferential statistical model was tested on the dependent variable "*Legionella* load (cfu/L)" and the independent variables (n = 57) according to the methods used by other authors (48-50). The *Poisson* regression

Table 1 - Glossary summary of common terminology in predictive modeling.

Term	Definition
Neuron	The basic element of a neural network, which connects toother neurons through transmitting data to each other (38)
Neural network	It consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations (39)
Bias neuron	A weight parameter for an extra input whose activation is permanently set to +1 (40)
Hidden layer	In an artificial neural network, this is defined as the layer between the input and output layers, where the result of their action cannot be directly observed (38)
epochs of learning	Each repeated entry of the full set of training patterns (40)
batch size	Hyperparameter of deep learning that controls the number of the training samples that are "fed" into the neural network before internal model parameters are updated (41)
validation split	A set of data used to test the performance of the network during training, but not used for modifying the weights of the network (40)
Rectified Linear Input (ReLU)	The activation function most frequently used followed by SoftMax for classification problems (42, 43)
permutation based VI scoring method	It is a method to measure variable importance scores for the predictors in a model (44).

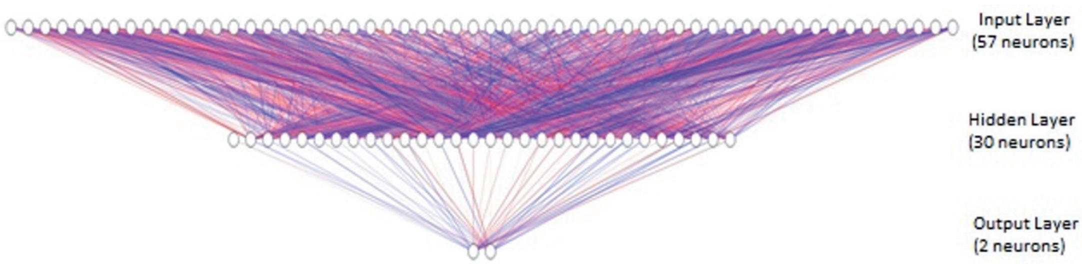

Figure 1 - Machine learning model architecture for Legionella results.

model was developed (70% and 50% of the dataset) and tested (30% and 50% of the dataset) on the *Legionella* load detected in the samples (0 cfu/L in the case of a negative result). Subsequently, only those parameters/risk factors with a p-value < 0.05 were considered to be statistically significant and included in the final model.

Predictions from the final *Poisson* regression model for each parameter analyzed in this study were used to calculate an overall risk score for *Legionella* positive outcomes using the following formula (50):

$$e^{(\alpha+\beta 1X1+ \beta 2X2+...+ \beta kXk)},$$

where:

α = intercept of the model;
β = coefficient of the regression model for each risk factor;
X = score of each risk factor.

ROC curve

The ROC curve was applied to the maintenance interventions of the water systems that were particularly relevant for the purposes of the forecasting models to determine the cut-off of days within which to perform subsequent maintenance interventions. R version 3.6.1 was used to perform all statistical tests.

Results

Of the 1,053 water samples analyzed for *Legionella*, 57 (5.4%) tested positive, of which 49 (86%) were for *Legionella pneumophila* (Lpn) serogroup (sg) 1, seven (12.3%) for Lpn sg 6, and one (1.7%) for Lpn sg 1+6. Regarding the detected concentration, 42 (73.7%) samples had a load < 1,000 cfu/L, 14 (24.6%) between 1,000 and 10,000 cfu/L, and one (1.7%) > 10,000 cfu/L.

When randomized and divided between the training (70%, 737/1,053) and testing (30%, 316/1,053) datasets, the positive samples were fairly evenly distributed (5.2%, 38/737 vs 6%, 19/316 water samples respectively). Similarly, for the algorithms created and tested with a 50%-50% split between the training and testing datasets, the positive water samples were distributed with 5.7% (30/527) in the training and 5.1% in the testing dataset (27/526).

Machine learning model

Of all the models tested, the ML model with 70% of dataset for training and 30% for testing, proved to be the most efficient model for predicting *Legionella* sample results (positive/negative) (Figure 1) (51). It started with 57 input benchmarks (input layer) and had a single hidden layer of 30 neurons and an output layer of two neurons (one for positive sample results and one for negative sample results). All layers had a bias neuron to increase their effectiveness.

The model was trained on the 737 samples of the train dataset (70%) with the following parameters: epochs of learning = 200, batch size = 4, and validation split = 0.6. The trained model, checked on 316 samples of the test dataset (30%), yielded the results in the confusion matrix shown in Table 2.

Table 2 - Confusion matrix for the machine learning model for *Legionella* detection

	Predictive results: positive	Predictive results: negative	Total real results
Real results: positive	**7**	12	19
Real results: negative	9	**288**	297
Total predictive results	16	300	**316**

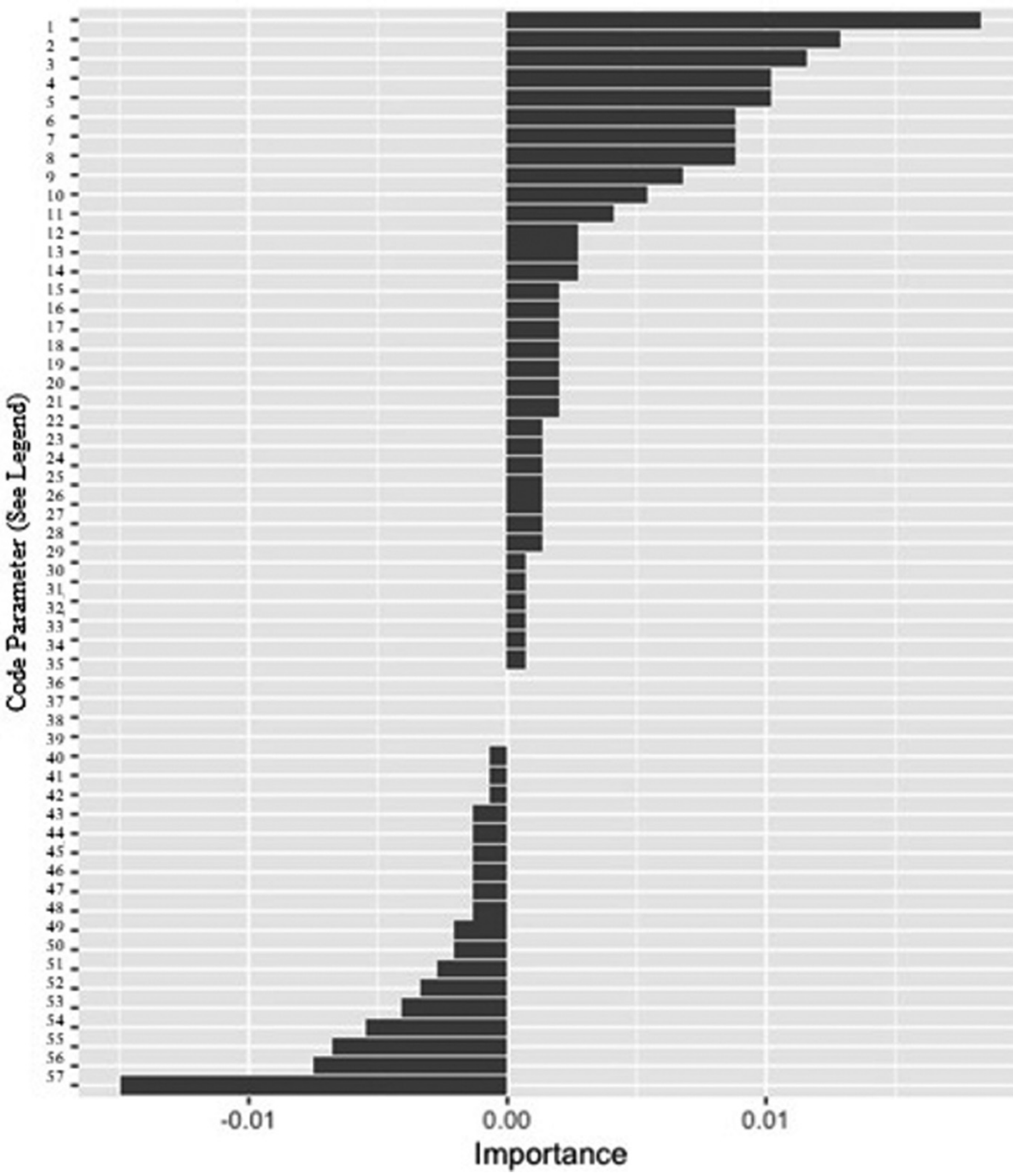

Figure 2 - Graphical representation of the importance of each parameter examined in the accuracy of the machine learning model.

Legend: 1. Type of water network (hot/cold), 2. water temperature at the time of sampling, 3. percentage of the water network with a 6.604 mm pipe diameter, 4. percentage of the underground water network, 5. number of days from the last filter valve replacement to the day of sampling, 6. aerator filter replacement (Yes/No), 7. number of days since the last disinfection of the water network with sodium hypochlorite for two days, 8. percentage of the water network with a 31.75 mm pipe diameter, 9. days since the last replacement of flexible hoses, 10. percentage of the above ground network, 11. replacement of flexible hoses (Yes/No), 12. length of the water network, 13. Total restructuring of the water network (Yes/No), 14. number of dead legs along the route, 15. total number of corners of the water network along the way , 16. replacement of filter valves (Yes/No), 17. days since the last shower head replacement to the day of sampling, 18. number of days since the total water network renovation, 19. mean atmospheric temperature on the day of sampling, 20. average tube pipe diameter, 21. percentage of the network with a 8.128 mm pipe diameter, 22. mixer replacement (Yes/No), 23. days since the last mixer replacement, 24. floor, 25. month of water sampling, 26. presence of an absolute filter at the point of use (Yes/No), 27. percentage of the water network with a 63.5 mm pipe diameter, 28. type of point of use (tap, shower, and bidet), 29. number of days since the last replacement of aerator filters to the day of sampling, 30. replacement of shower heads (Yes/No), 31. number of water delivery points in the room, 32. number of days since the installation of the absolute filter to the day of sampling, 33. percentage of the water network with a 12.7 mm pipe diameter, 34. percentage of the water network with a 3.556 mm pipe diameter, 35. number of unusable water delivery points in the room, 36. temperature range registered on the day of sampling, 37. percentage of the water network with a 4.572 mm pipe diameter, 38. percentage of the water system with 3.302 mm pipes, 39. percentage of the water system with 4.064 mm pipes, 40. sampling methods pre- or post-flushing, 41. percentage of the water system with 5.08 mm pipes, 42. wing, 43. percentage of the copper network, 44. percentage of the steel network, 45. number of water network risers, 46. percentage of the water mains with a 38.1 mm pipe diameter, 47. percentage of the water mains with a 3.048 mm pipe diameter, 48. number of diameter changes along the way, 49. percentage of multi-layer water pipes, 50. percentage of the water mains with a 19.05 mm pipe diameter, 51.number of corners along the water mains route (partial) 52. water network line, 53. percentage of the water system with 5.08 mm pipes, 54. percentage of the water system with 25.4 mm pipes, 55. percentage of the water system with 50.8 mm pipes, 56. percentage of the water system with 10.16 mm pipes, 57. disinfection of the network with sodium hypochlorite (continuous hyperchlorination) for two days (Yes/No).

Table 3 - Poisson regression model applied to the *Legionella* load (cfu/L).

	β	(eβ−1) = RR (%)	p-value
Intercept	-0.10230		< 0.0001*
Floor of the pavilion	0.13358	14.3	< 0.0001*
Hot/cold water network	0.10323	10.9	< 0.0001*
Mean atmospheric temperature on the day of sampling	0.09117	9.5	< 0.0001*
Atmospheric temperature range on the day of sampling	-0.12797	-12.0	< 0.0001*
Days since the last replacement of the filter valves	0.12782	13.6	< 0.0001*

*p < 0.05 statistically significant

Table 4 - Confusion matrix for the Poisson regression model.

	Predictive results: positive	Predictive results: negative	Total real Results
Real results: positive	**12**	7	19
Real results: negative	47	**250**	297
Total predictive results	59	257	**316**

Table 5 - Confusion matrix for the machine learning model + Poisson regression model.

	Predictive results: positive	Predictive results: negative	Total real Results
Real results: positive	**15**	4	19
Real results: negative	52	**245**	297
Total predictive results	67	249	**316**

The ML model had a prediction accuracy of 93.4% (295/316), with a sensitivity of 43.8% (7/16) and specificity of 96% (288/300).

The VIP package of R was applied to the model to determine which parameters had the greatest influence on the accuracy of the model. The results were shown in Figure 2.

From our data, it appears that some parameters, such as 1. type of water network, hot/cold, 2. water temperature at time of sampling, and 3. percentage of the water network with a 6.604 mm pipe diameter, had great importance in the model in a directly proportional sense (e.g., when the temperature of the water increased, the risk of *Legionella* contamination also increased). Other parameters, such as 55. percentage of the water system with 50.8 mm pipes, 56. percentage of the water system with 10.16 mm pipes, and 57. disinfection of the network with sodium hypochlorite for two days), were very important, but in an inversely proportional sense (e.g., as the percentage of pipes with a diameter of 10.16 mm or 50.8 mm that reached the point of supply increased, the risk of *Legionella* contamination decreased).

Poisson regression model

Table 3 shows the parameters that were statistically significant in influencing the best *Poisson regression* model applied to the *Legionella* load (training dataset 70%). Some parameters had a directly proportional influence, whereas others were inversely proportional (i.e., for each degree increase in the atmospheric temperature range on the day of sampling, the relative risk of *Legionella* contamination decreased by 12% in terms of the load).

Testing the model on the test dataset (30%) yielded the results shown in Table 4. The Poisson regression model had a prediction accuracy of 82.9% (262/316), with a sensitivity of 20.3% (12/59) and specificity of 97.3% (250/257).

The combination of the two models, where at least one of the two found positive predictions, obtained the results in Table 5. The combined model had a prediction accuracy of 82.3% (260/316), with a sensitivity of 22.4% (15/67) and specificity of 98.4% (245/249).

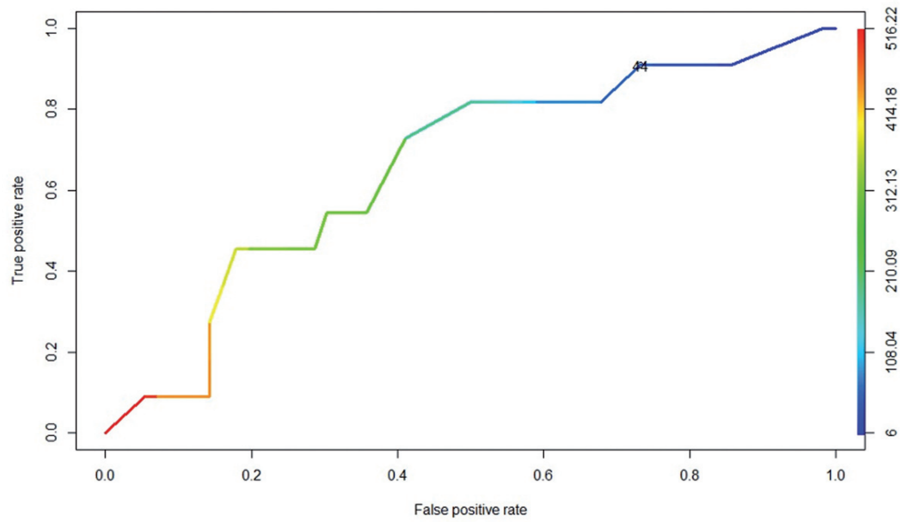

Figure 3 - ROC curves for the frequency of filter valve replacement and result of *Legionella* samples.

ROC curve

Considering the importance of the parameter “days since the last replacement of filter valves” by both models (Figure 2 and Table 2), the ROC curve (Figure 3) was applied to compare days from the last replacement with the result of *Legionella* in the sampling day.

The ROC curve showed that the ideal cut-off, beyond which 90% of the *Legionella* samples tested positive, was 44 days after the last maintenance intervention.

Discussion

Current AI tools are increasingly advancing, particularly ML and DL techniques, and have been applied in many areas of medicine, such as providing health information, making medical diagnoses, and predicting a patient’s risk of future complications (52). Our study is one of the few in the field that analyzes a large number of parameters (n = 57) to predict *Legionella* contamination of a water supply. It is also the first to combine two types of statistical models (ML and Poisson regression).

Among the models tested (ML, Poisson regression, and ML combined with Poisson regression), ML obtained the best results both in terms of predictive accuracy (93.4%) and sensitivity (43.8%). Regarding specificity, the combination of the two models provided the best results (98.4%). The application of these innovative models ensured a more correct approach than traditional models for monitoring the water network, a factor that should not be overlooked when discussing healthcare facilities and vulnerable patients (53). Our results showed that these predictive models could be useful to improve the quality of management in complex hospital organizations, which represent a high-risk environment for LD transmission due to, for example, old plumbing systems, dead-end branches, lack of use of tap water (13).

The analysis of the factors that influence the prediction models yielded interesting results. By comparing the parameters that most influenced the presence (ML model) and load (*Poisson* Regression) of *Legionella*, a coincidence was found for some parameters: “type of water network (hot/cold water)” and “days from the last filter valve replacement to the day of sampling”.

The “type of water network (hot/cold)” was the first factor in the order of importance according to the ML model and was statistically significant for the *Poisson* regression model. In particular, the *Poisson* regression analysis showed that the cold water network presented the risk of greater contamination than the hot water network. The role of the type of water network in influencing the presence of *Legionella* has been confirmed in the scientific literature (54) and it is not uncommon to find *Legionella* in cold water networks (>20°C) (47, 55-57). This aspect needs to be studied in depth because only hypotheses can explain these results at present: it is possible that the temperature of the cold water network does not reach temperatures < 20°C for various reasons (e.g., systems that are too superficial); that users obtain hot water more frequently, which reduces the flow of cold water; or that some buildings are closed in emergency cases, which causes the stagnation of the network and consequently greater contamination by *Legionella* (6,58). All these situations should be avoided in the management of the water network to reduce the risk of Legionella contamination.

The other parameter in common that most influenced the two models was the maintenance intervention for filter valve replacement (days from the last replacement to the day of sampling). Our results indicate that the valve filter becomes contaminated before the expiry date established by the manufacturer (90 days), therefore it is necessary to microbiologically monitor the filter to establish the duration of its validity. According to the *Poisson* regression model, each additional day that the filter valve was not replaced corresponded to a 13.6% increase in the risk of *Legionella* contamination of the water network. This confirmed the importance of establishing an appropriate maintenance program regarding filter valve replacement (59).

The atmospheric temperature parameter analyzed in the predictive models was also particularly interesting. According to some authors (48,60,61), the ML model highlighted that, as the average daily atmospheric temperature increased, the presence of *Legionella* in the water network also increased. According to other authors (62,63), the *Poisson* regression model revealed that the average atmospheric temperature had a trend directly proportional to the risk of increased *Legionella* load.

The ML model highlighted how the diameter of the water pipes also influences the presence of *Legionella*: overall, as the diameter increased, the presence of *Legionella* decreased. Other authors (64-68) have shown that the diameter of the pipes (including the water flow) can influence the formation of biofilm, which is widely considered to be the ideal habitat for the proliferation of *Legionella*. Our results could be influenced by some limitations of this study. For example, we did not consider the extent, presence of biofilms and/or other competing microorganisms such as *Pseudomonas aeruginosa*. Furthermore, we found that continuous disinfection with sodium hypochlorite was a parameter associated with the presence of *Legionella* because it was performed when a high microbial load was present in the water samples. Therefore, we believe that the application of these predictive models can be improved by expanding the number of parameters to be studied.

In addition, the lack of sensitivity could probably be due to the low number of positive samples, which do not allow the algorithm to adapt perfectly to the variation in risk of *Legionella* contamination at each individual water supply point. For this reason, we intend to extend the study to other pavilions of the same hospital to increase the dataset. This would improve the performance of the ML/DL models on the one hand and increase the test sensitivity on the other (69). Another method to increase the sensitivity of the model may be to eliminate the independent variables that least influence the development of the model. These "pruning" techniques have been developed recently and several authors have shown that they often lead to improved model performance (70,71).

Conclusions

In this study, we have shown that the application of artificial intelligence methods to aqueous matrices can improve the modeling of water contamination compared to classical statistical analysis.

Some recommendations arising from the main findings are summarized below:

- check and maintain the cold water temperature <20°C, because it can present a greater risk of *Legionella* contamination than the hot water network;
- check the expiry date of the valve filter, as it may become contaminated before the expiry date;
- monitor the water network, especially during the hottest periods, as the average atmospheric temperature favors the risk of *Legionella* contamination.

In accordance with the new European directive 2020/2184, predictive models would allow a rational choice for the control and prevention of water contamination (e.g. remediation systems) and a better management of the risk of waterborne diseases in terms of time and cost.

Author Contributions: Conceptualization, O.D.G, F.F. and M.T.M; methodology, O.D.G, G.D, F.F, F.T, V.S. and M.L.; data curation, formal analysis, O.D.G, G.D, F.F. and F.T.; writing - original draft, O.D.G, F.F, G.D. and M.T.M; writing - review & editing, M.L.,V.S., F.A., F.T. and C.M.L; supervision, O.D.G, F.F. and M.T.M.
All authors have read and agreed to the present version of the manuscript.

Funding: This research did not receive any external funding. It was supported by university funds managed by Osvalda De Giglio.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Edanz (https://www.edanz.com/ac)

for editing a draft of this manuscript.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Riassunto

Machine Learning vs. modelli di regressione per prevedere il rischio di contaminazione da *Legionella* in una rete idrica ospedaliera

Introduzione. Il monitoraggio periodico per rilevare la presenza di *Legionella* nelle reti idriche ospedaliere consente di adottare misure preventive per evitare il rischio di legionellosi in pazienti e operatori sanitari.

Disegno dello studio. Scopo dello studio è standardizzare un metodo per prevedere il rischio di contaminazione da *Legionella* nella rete idrica di una struttura ospedaliera, confrontando modelli di *Machine Learning* con modelli convenzionali e combinati.

Metodi. Nel periodo luglio 2021 – ottobre 2022 la ricerca di *Legionella* è stata effettuata in campioni di acqua prelevati in 356 stanze presenti in un padiglione ospedaliero italiano. Sono stati esaminati cinquantotto parametri riguardanti le caratteristiche strutturali e ambientali della rete idrica. I modelli sono stati costruiti sul 70% del dataset e testati sul restante 30% per valutare l'accuratezza, la sensibilità e la specificità.

Risultati. Sono stati analizzati 1.053 campioni di acqua, di cui 57 (5,4%) positivi per *Legionella*. Dei modelli *Machine Learning* testati, il più efficiente aveva uno strato di *input* (56 neuroni), uno strato nascosto (30 neuroni) e uno strato di *output* (due neuroni). L'accuratezza è risultata pari al 93,4%, la sensibilità al 43,8% e la specificità al 96%. Il modello di regressione ha rilevato un'accuratezza dell'82,9%, una sensibilità del 20,3% e una specificità del 97,3%. La combinazione dei modelli ha raggiunto un'accuratezza dell'82,3%, una sensibilità del 22,4% e una specificità del 98,4%. I parametri più importanti che hanno influenzato i risultati del modello sono stati il tipo di rete idrica (acqua calda/fredda), la sostituzione delle valvole dei filtri e la temperatura atmosferica. Tra i modelli testati, *Machine Learning* ha ottenuto i migliori risultati in termini di accuratezza e sensibilità.

Conclusioni. Sono necessari ulteriori studi per migliorare questi modelli predittivi, ampliando il *dataset* con l'inserimento di altri parametri e di altri padiglioni dello stesso ospedale.

References

1. Fields BS, Benson RF, Besser RE. Legionella and Legionnaires' disease: 25 years of investigation. Clin Microbiol Rev. 2002 Jul;**15**(3):506-26. doi: 10.1128/CMR.15.3.506-526.2002. PMID: 12097254.
2. Iliadi V, Staykova J, Iliadis S, Konstantinidou I, Sivykh P, Romanidou G, et al. Legionella pneumophila: The Journey from the Environment to the Blood. J Clin Med. 2022 Oct 18;**11**(20):6126. doi: 10.3390/jcm11206126. PMID: 36294446.
3. Samuelsson J, Payne Hallström L, Marrone G, Gomes Dias. J. Legionnaires' disease in the EU/EEA*: increasing trend from 2017 to 2019. Euro. Surveill. 2023, **28**(11), 2200114. doi: 10.2807/1560-7917.ES.2023.28.11.2200114. PMID: 36927719.
4. Guidelines for drinking-water quality: Third edition. Geneva: World Health Organization; 2004.
5. Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda. Geneva: World Health Organization; 2022. Available from: https://www.who.int/publications/i/item/9789240045064 [Last accessed: 2024 May 20].
6. Direttiva (E.U.), 2020/2184 del. Parlamento Europeo e del Consiglio del 16 dicembre 2020 Concernente la Qualità delle Acque Destinate al Consumo Umano. G.U. dell'Unione Europea L 435/1 del 23 dicembre 2020. Available from: http://data.europa.eu/eli/dir/2020/2184/oj [Last accessed: 2024 May 20].
7. Legislative Decree 18 February 2023 concerning the implementation of Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 concerning the quality of water intended for human consumption. Available from: https://www.gazzettaufficiale.it/eli/id/2023/03/06/23G00025/SG [Last accessed: 2024 May 20].
8. European Centre for Disease Prevention and Control (ECDC). Legionnaires' Disease: Annual Epidemiological Report for 2019. Annual Epidemiological Report on Communicable Diseases in Europe. Stockholm: ECDC; 2021.
9. Fischer FB, Saucy A, Vienneau D, Hattendorf J, Fanderl J, de Hoogh K, et al. Impacts of weather and air pollution on Legionnaires' disease in Switzerland: A national case-crossover study. Environ Res. 2023 Sep 15; **233**:116327. doi: 10.1016/j.envres.2023.116327. Epub 2023 Jun 22. PMID: 37354934.
10. Graham FF, Harte D, Zhang J, Fyfe C, Baker MG. Increased Incidence of Legionellosis after Improved Diagnostic Methods, New Zealand, 2000-2020. Emerg Infect Dis. 2023 Jun;**29**(6):1173-1182. doi: 10.3201/eid2906.221598. PMID: 37209673.
11. Centers for Disease Control and Prevention Legionnaires' Disease: Use Water Management Programs in Buildings to Help Prevent Outbreaks, 2016. Available from: https://www.cdc.gov/vitalsigns/legionnaires/index.html [Last accessed: 2024 May 20].
12. Kanarek P, Bogiel T, Breza-Boruta B. Legionellosis risk-an overview of Legionella spp. habitats in Europe. Environ Sci Pollut Res Int. 2022 Nov;**29**(51):76532-76542. doi: 10.1007/s11356-022-22950-9. Epub 2022 Sep 26. PMID: 36161570.
13. De Giglio O, Diella G, Lopuzzo M, Triggiano F, Calia C, Pousis C, et al. Management of Microbiological Contamination of the Water Network of a Newly Built Hospital Pavilion. Pathogens. 2021 Jan 16;**10**(1),75. doi: 10.3390/pathogens10010075.
14. Ghaznavi C, Ishikane M, Yoneoka D, Tanoue Y, Kawashima T, Eguchi A, et al. Effect of the COVID-19 pandemic

and state of emergency declarations on the relative incidence of legionellosis and invasive pneumococcal disease in Japan. J Infect Chemother. 2023 Jan;**29**(1), 90-4. doi: 10.1016/j.jiac.2022.08.016. Epub 2022 Sep 16. PMID: 36116719.

15. Borella P, Montagna MT, Stampi S, Stancanelli G, Romano-Spica V, Triassi M, et al. Legionella contamination in hot water of Italian hotels. Appl Environ Microbiol. 2005 Oct;**71**(10):5805-13. doi: 10.1128/AEM.71.10.5805-5813.2005. PMID: 16204491.
16. Kyritsi MA, Mouchtouri VA, Katsioulis A, Kostara E, Nakoulas V, Hatzinikou M, et al. Legionella Colonization of Hotel Water Systems in Touristic Places of Greece: Association with System Characteristics and Physicochemical Parameters. Int J Environ Res Public Health. 2018 Nov 30; **15**(12):2707. doi: https://doi.org/10.3390/ijerph15122707. PMID: 30513698.
17. D'Alò GL, Messina A, Mozzetti C, Cicciarella Modica D, De Filippis P. Competitive colonization of Legionella and Pseudomonas aeruginosa in water systems of residential facilities hosting closed communities Legionella versus Pseudomonas aeruginosa in water systems of residential facilities. Ig Sanita Pubbl. 2022 Mar-Apr; **79**(2):92-110.
18. De Giglio O, Diella G, Lopuzzo M, Triggiano F, Calia C, Pousis C, et al. Impact of lockdown on the microbiological status of the hospital water network during COVID-19 pandemic. Environ Res. 2020 Dec;**191**:110231. doi: 10.1016/j.envres.2020.110231. Epub 2020 Sep 23. PMID: 32976823.
19. Gamage SD, Jinadatha C, Coppin JD, Kralovic SM, Bender A, Ambrose M, et al. Factors That Affect Legionella Positivity in Healthcare Building Water Systems from a Large, National Environmental Surveillance Initiative. Environ Sci Technol. 2022 Aug 16;56(16):11363-11373. doi: 10.1021/acs.est.2c02194. Epub 2022 Aug 5. PMID: 35929739
20. Federigi I, De Giglio O, Diella G, Triggiano F, Apollonio F, D'Ambrosio M, et al. Quantitative Microbial Risk Assessment Applied to *Legionella* Contamination on Long-Distance Public Transport. Int J Environ Res Public Health. 2022 Feb 10;**19**(4):1960. doi: 10.3390/ijerph19041960. PMID: 35206148.
21. De Giglio O, Napoli C, Diella G, Fasano F, Lopuzzo M, Apollonio F, et al. Integrated approach for legionellosis risk analysis in touristic-recreational facilities. Environ Res. 2021 Nov; **202**:111649. doi: 10.1016/j.envres.2021.111649. Epub 2021 Jul 9. PMID: 34252427.
22. Nagy DJ, Dziewulski DM, Codru N, Lauper UL. Understanding the distribution of positive Legionella samples in healthcare-premise water systems: Using statistical analysis to determine a distribution for Legionella and to support sample size recommendations. Infect Control Hosp Epidemiol. 2021 Jan;**42**(1):63-68. doi: 10.1017/ice.2020.384. Epub 2020 Oct 8. PMID: 33028429.
23. Fasano F, Addante AS, Valenzano B, Scannicchio G. Variables Influencing per Capita Production, Separate Collection, and Costs of Municipal Solid Waste in the Apulia Region (Italy): An Experience of Deep Learning. Int J Environ Res Public Health. 2021 Jan 17;**18**(2):752. doi: 10.3390/ijerph18020752. PMID: 33477308.
24. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, et al. State-of-the-Art Survey on Deep Learning Theory and Architectures. Electronics. 2019; **8**(3):292. doi: https://doi.org/10.3390/electronics8030292.
25. Brunello A, Civilini M, De Martin S, Saccomanno M, Vitacolonna N. Machine learning-assisted environmental surveillance of *Legionella*: A retrospective observational study in Friuli-Venezia Giulia region of Italy in the period 2002–2019. Informatics in Medicine Unlocked.2022; **28**:100803. doi: https://doi.org/10.1016/j.imu.2021.100803.
26. Tata A, Marzoli F, Cordovana M, Zacometti C, Massaro A, Barco L, et al. A multi-center validation study on the discrimination of *Legionella pneumophila* sg.1, *Legionella pneumophila* sg. 2-15 and *Legionella* non-*pneumophila* isolates from water by FT-IR spectroscopy. Front Microbiol. 2023 Apr 13;**14**:1150942. doi: 10.3389/fmicb.2023.1150942. PMID: 37125166.
27. Sinčak P, Ondo J, Kaposztasova D, Virčkova M, Vranayova Z, Sabol J. Artificial intelligence in public health prevention of legionelosis in drinking water systems. Int J Environ Res Public Health. 2014 Aug 21;**11**(8):8597-611. doi: 10.3390/ijerph110808597. PMID: 25153475.
28. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. Global Edition; 2021.
29. Soori M, Arezoo B, Dastres R. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. Cognitive Robotics. 202; **3**:54-70. doi: https://doi.org/10.1016/j.cogr.2023.04.001.
30. Sharma N, Sharma R, Jindal N. Machine Learning and Deep Learning Applications-A Vision. Global Transitions Proceedings. 2021;**2**(1):24-28. doi:https://doi.org/10.1016/j.gltp.2021.01.004.
31. Guidelines for the Prevention and Control of Legionellosis, 2015. Available from: http://www.salute.gov.it/imgs/C_17_pubblicazioni_2362_allegato.pdf. [Last accessed: 2024 May 20].
32. ISO 11731:2017. Water Quality—Enumeration of Legionella; International Organization for Standardization: Geneva, Switzerland; 2017.
33. Civil Protection Department Apulia Region. Available from: https://protezionecivile.puglia.it/bollettini-meteorologici-regionali-mensili. [Last accessed: 2024 May 20].
34. Potdar K, Pardawala TS, Pai CDA. Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. Int. J. Comput Appl. 2017;**175**:7-9. doi: 10.5120/ijca2017915495.
35. Patro S, Sahu KK. Normalization: A Preprocessing Stage. IARJSET. 2015;**2**(3):20-22. doi: 10.5120/ijca2017915495.
36. Xu Y, Goodacre R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. J Anal Test. 2018;**2**(3):249-262. doi: 10.1007/s41664-018-0068-2. Epub 2018 Oct 29. PMID: 30842888.

37. Dobbin KK, Simon RM. Optimally splitting cases for training and testing high dimensional classifiers. BMC Med Genomics. 2011 Apr 8;**4**:31. doi: 10.1186/1755-8794-4-31. PMID: 21477282.
38. Kufel J, Bargieł-Łčczek K, Kocot S, Kodik M, Bartnikowska W, Janik M, et al. What Is Machine Learning, Artificial Neural Networks and Deep Learning?-Examples of Practical Applications in Medicine. Diagnostics (Basel). 2023 Aug 3;**13**(15):2582. doi: 10.3390/diagnostics13152582. PMID: 37568945.
39. Jürgen Schmidhuber. Deep learning in neural networks: An overview. Neural Networks. 2015;**61**:85-117. https://doi.org/10.1016/j.neunet.2014.09.003.
40. Stegemann J, Buenfeld N. A Glossary of Basic Neural Network Terminology for Regression Problems. Neural Comput. & Applic. 1999; **8**:290–6. https://doi.org/10.1007/s005210050034.
41. Xu C, Coen-Pirani P, Jiang X. Empirical Study of Overfitting in Deep Learning for Predicting Breast Cancer Metastasis. Cancers. 2023;**15**:1969. https://doi.org/10.3390/cancers15071969.
42. Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives. IEEE Transact Pattern Anal Machine Intell. 2013;**35**:1798-1828. doi: 10.1109/TPAMI.2013.50.
43. Deng L,Yu D. Deep Learning: Methods and Applications. Found. Trends Signal Process 2014;**7**:197-387. doi: 10.1561/2000000039.
44. Greenwell BM, Boehmke BC. Variable Importance Plots-An Introduction to the vip Package. R Journal 2020;**12(**1):343-366. https://doi.org/10.32614/RJ-2020-013.
45. Favorskaya MN, Andreev VV. The study of activation functions in deep learning for pedestrian detection and tracking. Int Arch Photogramm Remote Sens Spat Inf. Sci 2019; **XLII-2/W12**:53-9. doi: 10.5194/isprs-archives-XLII-2-W12-53-2019.
46. Eckle K, Shmidt-Hieber J. A comparison of deep networks with ReLU activation function and linear spline-type methods. Neural Netw**.** 2019;**110**:232–242. doi: 10.1016/j.neunet.2018.11.005.
47. Huang F, Zhang J, Zhou C, Wang Y, Huang J,Zhu L. A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. Landslides. 2020;**17**:217-229. doi: 10.1007/s10346-019-01274-9.
48. De Giglio O, Fasano F, Diella G, Lopuzzo M, Napoli C, Apollonio F, et al. Legionella and legionellosis in touristic-recreational facilities: Influence of climate factors and geostatistical analysis in Southern Italy (2001-2017). Environ Res. 2019;**178**:108721. doi: 10.1016/j.envres.2019.108721. Epub 2019 Sep 6. PMID: 31541805.
49. Conza L, Casati Pagani S, Gaia V. Influence of climate and geography on the occurrence of Legionella and amoebae in composting facilities. BMC Res Notes. 2014 Nov 24;**7**:831. doi: 10.1186/1756-0500-7-831. PMID: 25421541.
50. Cui Y, Kim DY, Zhu J. On the generalized poisson regression mixture model for mapping quantitative trait loci with count data. Genetics. 2006 Dec;**174**(4):2159-72. doi: 10.1534/genetics.106.061960. Epub 2006 Oct 8. PMID: 17028335.
51. Nguyen QH, Ly HB, Ho LS, Al-Ansari N, Le HV, Tran VQ, et al. Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. Mathematical Problems in Engineering. 2021:1-15. doi: 10.1155/2021/4832864.
52. Singh P, Singh N, Singh KK, Singh A. Chapter 5 - Diagnosing of disease using machine learning. In: Singh KK, Elhoseny M, Singh A, Elngar AA, Eds. Machine Learning and the Internet of Medical Things in Healthcare. Academic Press; 2021:89-111. doi: https://doi.org/10.1016/B978-0-12-821229-5.00003-3.
53. Wilson AM, Canter K, Abney SE, Gerba CP, Myers ER, Hanlin J, et al. An application for relating Legionella shower water monitoring results to estimated health outcomes. Water Res. 2022 Aug 1;**221**:118812. doi: 10.1016/j.watres.2022.118812. Epub 2022 Jul 3. PMID: 35816914.
54. Marchesi I, Paduano S, Frezza G, Sircana L, Vecchi E, Zuccarello P, et al. Safety and Effectiveness of Monochloramine Treatment for Disinfecting Hospital Water Networks. Int J Environ Res Public Health. 2020 Aug 22;**17**(17):6116. doi: 10.3390/ijerph17176116. PMID: 32842654.
55. Papadakis A, Keramarou M, Chochlakis D, Sandalakis V, Mouchtouri VA, Psaroulaki A. *Legionella* spp. Colonization in Water Systems of Hotels Linked with Travel-Associated Legionnaires' Disease. **Water**. 2021;**13**(16):2243. https://doi.org/10.3390/w13162243.
56. Arvand M, Jungkind K, Hack A. Contamination of the cold water distribution system of health care facilities by Legionella pneumophila: do we know the true dimension? Euro Surveill. 2011 Apr 21;**16**(16):19844. PMID: 21527132.
57. Stout JE, Yu VL, Muraca P. Isolation of Legionella pneumophila from the cold water of hospital ice machines: implications for origin and transmission of the organism. Infect Control. 1985;**6**(4):141-6. doi: 10.1017/s0195941700062937. PMID: 3886578.
58. Istituto Superiore di Sanità 2020. Rapporto COVID-19, n. 21/2020. Guida per la prevenzione della contaminazione da Legionella negli impianti idrici di strutture turistico recettive, e altri edifici ad uso civile e industriale non utilizzati durante la pandemia COVID-19.
59. Sheffer PJ, Stout JE, Wagener MM, Muder RR. Efficacy of new point-of-use water filter for preventing exposure to Legionella and waterborne bacteria. Am J Infect Control. 2005;**33**(5 Suppl 1):S20-5. doi: 10.1016/j.ajic.2005.03.012. PMID: 15940113.
60. Walker JT. The influence of climate change on waterborne disease and Legionella: a review. Perspect Public Health. 2018 Sep;**138**(5):282-286. doi: 10.1177/1757913918791198. PMID: 30156484.
61. Fragou K, Kokkinos P, Gogos C, Alamanos Y, Vantarakis A. Prevalence of Legionella spp. in water systems of hospitals and hotels in South Western Greece. Int J Environ Health Res. 2012;**22**(4):340-54. doi: 10.1080/09603123.2011.643229.

Epub 2011 Dec 12. PMID: 22149148.
62. Montagna MT, Brigida S, Fasano F, Leone CM, D'Ambrosio M, Spagnuolo V, et al. The role of air temperature in Legionella water contamination and legionellosis incidence rates in southern Italy (2018-2023). Ann Ig. 2023 Nov-Dec;**35**(6):631-640. doi: 10.7416/ai.2023.2578. Epub 2023 Sep 20. PMID: 37724578.
63. Dupke S, Buchholz U, Fastner J, Förster C, Frank C, Lewin A, et al. Impact of climate change on waterborne infections and intoxications. J Health Monit. 2023 Jun 1;**8**(Suppl 3):62-77. doi: 10.25646/11402. PMID: 37342430; PMCID: PMC10278370.
64. Pavissich JP, Aybar M, Martin KJ, Nerenberg R. A methodology to assess the effects of biofilm roughness on substrate fluxes using image analysis, substrate profiling, and mathematical modelling. Water Sci Technol. 2014;**69**(9):1932-41. doi: 10.2166/wst.2014.103. PMID: 24804670.
65. Tierra G, Pavissich JP, Nerenberg R, Xu Z, Alber MS. Multicomponent model of deformation and detachment of a biofilm under fluid flow. J R Soc Interface. 2015 May 6;**12**(106):20150045. doi: 10.1098/rsif.2015.0045. PMID: 25808342.
66. Liu J, Chen H, Yao L, Wei Z, Lou L, Shan Y, et al.The spatial distribution of pollutants in pipe-scale of large-diameter pipelines in a drinking water distribution system. J Hazard Mater. 2016 Nov 5; **317**:27-35. doi: 10.1016/j.jhazmat.2016.05.048. Epub 2016 May 17. PMID: 27244696.
67. Shen Y, Monroy GL, Derlon N, Janjaroen D, Huang C, Morgenroth E, et al. Role of biofilm roughness and hydrodynamic conditions in *Legionella pneumophila* adhesion to and detachment from simulated drinking water biofilms. Environ Sci Technol. 2015;**49**(7):4274-82. doi: 10.1021/es505842v. Epub 2015 Mar 11. PMID: 25699403.
68. Lin H, Zhu X, Wang Y, Yu X. Effect of sodium hypochlorite on typical biofilms formed in drinking water distribution systems. J Water Health. 2017;**15**(2):218-227. doi: 10.2166/wh.2017.141. PMID: 28362303.
69. Hordri NF, Samar A, Yuhaniz SS, Shamsuddin SM. A systematic literature review on features of deep learning in big data analytics. Int J Adv Soft Comput Appl. 2017;**9**(1):32-49.
70. Vadera S, Ameen S. Methods for Pruning Deep Neural Networks. IEEE Access. 2022;**10**:63280-63300.
71. Ma YD, Zhao ZC, Liu D, He Z, Zhou W. OCAP: On-device Class-Aware Pruning for personalized edge DNN models. J Syst Architect. 2023;**142**:102956.

Corresponding author: Prof. Osvalda De Giglio, Interdisciplinary Department of Medicine, University of Bari Aldo Moro, Piazza Giulio Cesare 11, 70124 Bari, Italy
e-mail: osvalda.degiglio@uniba.it