

# Predicting Fish-Mortality: Artificial Neural Networks vs Symbolic Regression

Theofanis Aravanis

Dept of Mechanical Engineering  
University of the Peloponnese  
Patras, Greece  
taravanis@upatras.gr

Aristeidis Ilias

Dept of Computer Engin. and Informatics  
University of Patras  
Patras, Greece  
aristeid@ceid.upatras.gr

Ioannis Hatzilygeroudis

Dept of Computer Engin. and Informatics  
University of Patras  
Patras, Greece  
ihatz@ceid.upatras.gr

Georgios Spiliopoulos

Fish Health & Welfare Manager  
Kefalonia Fisheries S.A.  
Kefalonia, Greece  
g.spiliopoulos@kefish.gr

**Abstract**—In this paper, we present a comparison between a collection of Artificial Neural Networks (ANNs) —specifically, Multilayer Perceptrons— and a collection of Symbolic Regression (SR) models, all developed for predicting fish-deaths due to infectious diseases in aquaculture. The implemented Machine-Learning models are able to identify patterns that indicate an increased risk of fish-deaths, and provide early warning-alerts to fish-farmers and other stakeholders, enabling them to take necessary measures to prevent or minimize losses of fishes. The models were trained using real-world data acquired from a large Greek fish-farming unit, and evaluated in a validation dataset (distinct from the training dataset) based on their Mean Absolute Error (MAE) performance. The study found that, for each disease considered, the corresponding ANN model outperformed the respective SR model in terms of MAE. However, the ANNs sacrificed interpretability for higher performance, while the developed SR models, despite lower performance, managed to produce transparent and understandable mathematical expressions for estimating fish-deaths. Overall, this study not only provides a valuable method for generating early warning-alerts of pathogenic circumstances to the interested stakeholders, improving the efficiency and sustainability of fish-farming operations, but it also sheds light on the strengths and limitations of the developed ANN and SR models.

**Index Terms**—Fish-Deaths Prediction, Aquaculture, Machine Learning, Artificial Neural Networks, Symbolic Regression

## I. INTRODUCTION

Fish farming plays a crucial role in the global food supply chain by providing a sustainable source of protein for human consumption [10]. However, this industry is plagued by various diseases that can cause significant losses in fish populations [13]. These losses not only impact the economic viability of fish-farming operations, but also have wider ecological implications [5]. One effective way to mitigate losses in fish populations is through the use of intelligent *Machine Learning* (ML) techniques that provide *early warning-alerts* of pathogenic circumstances to the interested stakeholders. Given the critical nature of this issue, it has been extensively studied in the literature; for a review of various ML techniques used

for the diagnosis of fish diseases in aquatic applications, the interested reader is referred to [7].

The present article contributes to the above-mentioned line of research. In particular, it presents a *comparison* between two well-established and powerful ML models —namely, *Artificial Neural Networks* (ANNs) on the one hand [6], and *Symbolic Regression* (SR) models on the other hand [8], [11]— that are utilized to predict deaths of fishes (sea basses), caused by *infectious diseases*. Each one of these ML models has its own strengths and weaknesses, which can be inferred by their key differences that are summarized in the following points:

- ANNs are *universal function approximators*, [2], that have a layered structure, where the input is processed through multiple layers of interconnected nodes called *neurons*. ANNs use a *distributed representation* of knowledge, as the knowledge is encoded into the *synaptic weights* of the connections between the neurons. SR models, on the other hand, use a *symbolic representation* of knowledge, in the sense that knowledge is encoded into *mathematical equations* or *symbolic rules* that represent the relationship between the input and output variables.
- ANNs are considered as *black-box* models, since it is difficult to understand how such models arrive at their decisions. This is not the case with SR models, which are considered as *white-box* models, in view of the fact that the relationships between the input and output variables are represented using symbolic representations, making it easy to understand how such a model arrives at a certain decision. This higher interpretability of SR models has made them a favourable tool used for extracting the underlying laws of physical systems from experimental data [8], [11].

The ML models developed herein were trained on a set of *real-world* data, collected from a large fish-farming unit based in Greece. Each ML model takes as input the corresponding

data of the last  $N$  days (that is, number of fish-deaths, water temperature, administered amount of food, administered medicated food, antibiotics and vaccination doses), and produces as an output the estimated number of fish-deaths, due to a specific infectious disease, on the  $j$ -th *forthcoming* day.

The obtained results point out that, for each disease under consideration, the ANN model *outperforms* the corresponding SR model, as the former model achieves lower Mean Absolute Error (MAE) than the latter model. The higher performance of ANNs in terms of MAE is being paid by their *non-interpretability*, in contrast to the developed SR models which, despite their lower performance, manage to obtain *transparent* and *comprehensible* mathematical expressions for estimating fish-deaths.

The objective of this study is twofold.

- Firstly, it aims to provide a comprehensive comparison of the performance of ANN and SR models used for the early detection of fish-deaths. This comparison shall yield valuable insights into the strengths and limitations of each technique, enabling the identification of the most suitable method for specific needs. It is worth noting that comparisons between ANN and SR models used for estimation are still relatively scarce in the literature; the reader is, indicatively, referred to [12] and [1] for such a comparison in the domains of Mechanical and Electrical Engineering, respectively. Comparisons between ANN and SR models used for estimating fish-deaths in aquaculture are, to the best of our knowledge, entirely absent.
- Secondly, by developing ML models trained on real-world data, this study offers useful tools to improve the overall sustainability and profitability of the fish-farming industry. The timely information provided to the interested stakeholders by these models enables an effective management of diseases, reducing the risk of fish-deaths and associated losses.

The remainder of this article is structured as follows: The next section explores the profile of the available data, on which the implemented ML models shall be trained. Section III introduces the architecture of the approach followed, and discusses the technical aspects of the developed ML models. Section IV presents and comments on the obtained results. The article closes with a brief conclusion section.

## II. DATA PROFILE

The available data were obtained from a fish-farming unit based in Greece, and cover the time-period from 01/06/2020 to 14/09/2022. The data include *daily* records concerning the following quantities:

- Number of fish-deaths due to a specific infectious disease, and total number of fish-deaths (due to all causes)
- Water temperature
- Administered amount of food

- Administered amount of medicated food
- Administered antibiotics doses
- Administered vaccination doses

The infectious diseases examined herein are those for which there was a sufficient amount of data available to effectively train the developed ML models; those diseases are the diseases caused by the bacteria *Pasteurella*, *Vibrio Harveyi* and *Myxobacteria*. Against this background, Figure 1 depicts the time-series of the daily fish-deaths due to the examined diseases, as well as the time-series of the daily total fish-deaths (due to all causes), as recorded in the available data, i.e., during the time-period from 01/06/2020 to 14/09/2022.<sup>1</sup> It is evident that fish-deaths due to *Pasteurella* are more frequent than fish-deaths due to *Vibrio Harveyi* and *Myxobacteria*, which suddenly appear in rare time-intervals. As far as the daily total fish-deaths are concerned, the two high peaks appearing in Figure 1 are noteworthy. We note also that the volume of the available data differs for each considered disease. In particular, the number of samples/records referring to fish-deaths due to *Pasteurella* is significantly greater than the number of samples/records referring to fish-deaths due to *Vibrio Harveyi* and *Myxobacteria*.

## III. SYSTEM'S ARCHITECTURE

In this section, the architecture of the followed approach is presented. As illustrated in Figure 2, for each infectious disease under consideration  $D$ , a trained ML model—either an Artificial Neural Network or a Symbolic Regression model—takes as input a dataset of the *last*  $N$  days of a particular fish-cage (i.e., data for the present day  $t_0$ , data for the day before today  $t_{-1}$ , data for the day before yesterday  $t_{-2}$ , and so forth), and produces as an output a real number, which represents the estimate for the fish-deaths, due to the disease  $D$ , on the  $j$ -th forthcoming day  $t_{+j}$ . The input data corresponding to a particular day  $i$  concern the real-world available data discussed in Section II; hence, the input data of the day  $i$  include the fish-deaths due to  $D$  on the day  $i$ , the water temperature on the day  $i$ , the amount of food given on the day  $i$ , the administered medicated food of the day  $i$ , as well as the antibiotics and vaccination doses administered on the day  $i$ . By combining the output of each trained ML model (which pertains to a specific disease), an estimate for the fish-deaths of the  $j$ -th forthcoming day, due to *all* considered infectious diseases, is produced. For our case study, it is assumed that each ML model receives as input the data of a fish-cage for the last 10

<sup>1</sup>The time-series of the daily fish-deaths due to the examined diseases, as depicted in Figure 1, present discontinuities, since the depicted fish-deaths concern *multiple* (rather than a single) fish-cages of the fish-farming unit. This means that such a time-series of Figure 1, firstly, represents all consecutive daily fish-deaths recorded in a fish-cage  $A$ , then represents all consecutive daily fish-deaths recorded in a fish-cage  $B$ , and so forth. We retain this illustration of time-series for a better presentation. On the other hand, the time-series of the daily total fish-deaths does not present such discontinuities, since the recorded total fish-deaths refer to *all* fish-cages of the fish-farming unit. Note, lastly, that the values of all time-series presented herein are normalized (roughly) between 0 and 1.

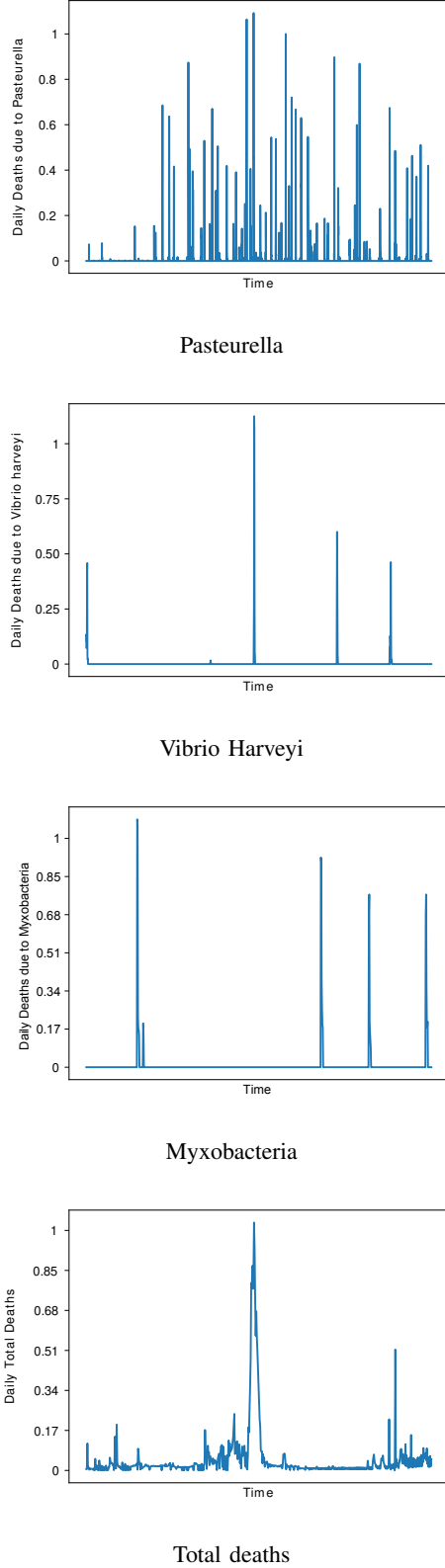


Fig. 1: Time-series of the daily fish-deaths due to the examined diseases, as well as of the daily total fish-deaths (due to all causes), from 01/06/2020 to 14/09/2022.

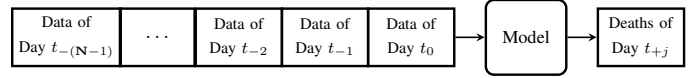


Fig. 2: Architecture diagram of a trained ML model.

days (thus,  $N = 10$ ), and produces as an output an estimate for the 3-rd forthcoming day (thus,  $j = 3$ ).<sup>2</sup>

It is noteworthy that the above approach for estimating fish-deaths can also be applied as a *binary classification* — a binary-classification approach would not capture the number of fish-deaths, but only whether fish-deaths are expected or not. This can be done as follows: In case the output of a trained ML model is a positive real number (or a positive real number greater than a certain pre-specified *threshold*), then we assume that fish-deaths, due to a disease  $D$ , are indeed expected on the  $j$ -th forthcoming day. In case the output of the ML model is zero (or a number close to zero), then we assume that no fish-deaths, due to a disease  $D$ , are expected on the  $j$ -th forthcoming day. In the context of this binary-classification approach, the generated output will now be a list of binary numbers (0 and 1), rather than a list of real numbers (as previously mentioned).

In addition to the ML models trained to estimate fish-deaths due to infectious diseases, we also develop an ANN and an SR model both trained for the estimation of the *total* number of fish-deaths (due to all causes). The input of each such model is the total number of fish-deaths recorded on each one of the *last* 4 days, and the output of the model is a real number representing the estimate for the total number of fish-deaths of the *next* day.

Having outlined the basic architecture of the adopted approach, we discuss in the remainder of this section the technical details of the developed ANN and SR models.

#### A. Artificial Neural Networks

Let us first discuss the architecture and parameters of the developed ANNs. All the ANNs considered have identical structure, so we shall focus on a single ANN. Accordingly, the alluded ANN is a conventional *Multilayer Perceptron* (MLP), which consists of *four layers*, namely, one input layer, two fully-connected (dense) hidden layers and one output layer (cf. Figure 3). The input layer of the network is a passive layer with no learnable parameters, that merely receives the data of each sample of the dataset. Each fully-connected hidden layer has 130 neurons (alias, units) and employs a *Rectified Linear Unit* (ReLU) activation function. The output layer contains a single *linear* neuron, which produces a real number that represents the *estimation* of fish-deaths due to a particular infectious disease. The ANN is compiled using a *Mean Absolute Error* (MAE) loss function, and the well-known *Adam* optimization algorithm, which is an extension of the *stochastic gradient descent* [3]. For tackling the *overfitting* problem, the *dropout* regularization technique was used where necessary.

<sup>2</sup>These specific numbers were chosen in consultation with the experts of the fish-farming unit.

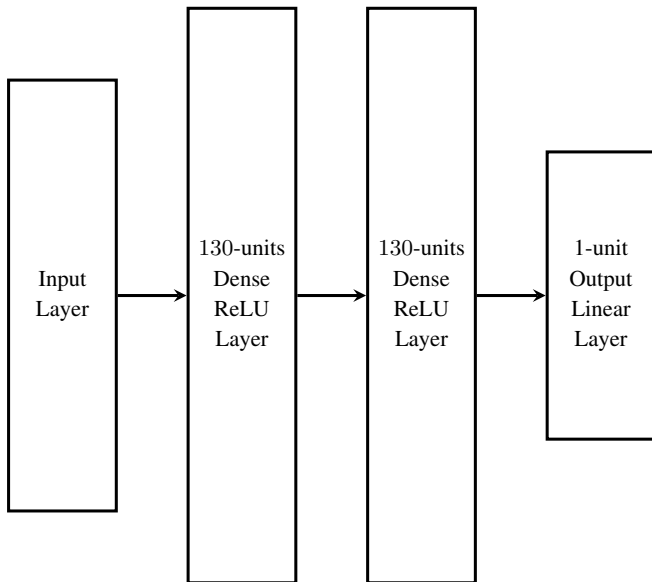


Fig. 3: Architecture of an ANN in charge of a particular disease.

By implementing dropout, selected neurons of the ANN are randomly ignored (“dropped out”) during the training-phase, and, as a consequence, no weight updates are applied to those ignored neurons [9].

### B. Symbolic Regression Models

*Symbolic Regression* (SR) is a method of *interpretable* Machine Learning that reduces data to mathematical equations. More precisely, SR is a type of regression analysis that searches the space of mathematical expressions (symbolic expressions) to find the model that *best fits* a given dataset, both in terms of *accuracy* and *simplicity*. Contrary to conventional regression techniques that seek to optimize parameters for a *predefined* structure of a regression model, SR does *not* impose prior assumptions, and instead infers both the structure of a regression model, as well as its parameters, directly from the data. This in turn entails that SR is *not* affected by human bias or gaps in the domain-knowledge, since it does not require a priori specification of a regression model (that is perhaps mathematically comprehensive from a human perspective). Yet, the disadvantage of SR is the fact that it takes considerable time to find a regression model that appropriately fits the dataset, as the corresponding search-space is huge.<sup>3</sup> Nevertheless, most SR algorithms prevent the combinatorial explosion by applying *evolutionary algorithms* that iteratively improve —over generations— the mathematical expression that best fits the available data.

Such an evolutionary algorithm, based on the fundamental concepts of Darwinian evolution, builds an initial population of naïve random formulas, by *randomly* combining mathematical

<sup>3</sup>As a matter of fact, the search-space in the context of SR is infinite, given that there is an infinite number of regression models that will perfectly fit a finite dataset.

building blocks, such as mathematical operators (i.e., +, −, \*, ÷), analytic functions (e.g., sqrt, cos, sin, exp, log), constants, and state variables. Each successive generation of formulas is, then, evolved from the preceding one, by selecting the *fittest* individuals from the population to undergo genetic operations. Eventually, the algorithm takes a series of totally random formulas, untrained and unaware of any given target function, and makes them breed, mutate and evolve their way towards a formula that best fits the data.

In our study, the development of the SR models was carried out using Python’s `gplearn` tool, which extends the `scikit-learn` machine-learning library to perform Genetic Programming for Symbolic Regression [4].<sup>4</sup> Table I lists the hyper-parameters setup for the `gplearn` tool. The explanation of each hyper-parameter is as follows: *Population Size* is the number of individuals (mathematical formulas) in each generation, *Generations* is the maximum number of generations, *Metric* is the measure of an individual’s fitness (herein is MAE), and *Stopping Criteria* expresses the fitness value at which the evolution-procedure terminates. *Function Set* contains the mathematical functions used when building and evolving generations. *Crossover Probability* controls the crossover method according to which genetic material between individuals is mixed. *Sub-tree Mutation Probability*, *Hoist Mutation Probability* and *Point Mutation Probability* control the respective mutation-operations. Lastly, the *Parsimony Coefficient* is a constant that penalizes large individuals by adjusting their fitness (MAE) to be less favourable for selection — this penalty helps in producing less computationally costly individuals which are, at the same time, more understandable.

Parameter	Value
Population Size	5000
Generations	20
Metric	Mean Absolute Error (MAE)
Stopping Criteria	0.01 (fish-deaths)
Function Set	add, sub, mul, div, sqrt, cos, sin, log
Crossover Probability	0.7
Sub-tree Mutation Probability	0.1
Hoist Mutation Probability	0.05
Point Mutation Probability	0.1
Parsimony Coefficient	0.05 – 0.1

TABLE I: The setup of hyper-parameters for the `gplearn` tool.

## IV. RESULTS

Having discussed the architecture and parameters of the developed ANN and SR models, this section is devoted to the presentation of the derived results. Let us begin with Table II,

<sup>4</sup><https://gplearn.readthedocs.io/en/latest/intro.html>

which, for each infectious disease, presents the Mean Absolute Error (MAE) obtained by the corresponding ANN and SR models in the *validation dataset*. The validation dataset is distinct from the training dataset, and its size is about 40% of the size of the training dataset. It is clear that, for each one of the three considered diseases, the ANN model trained for the respective disease managed to achieve a *lower* MAE than the corresponding SR model, an observation that testifies that the ANN model *outperforms* the SR model in terms of the obtained error-metric.

Disease	ANN Model MAE	SR Model MAE
Pasteurella	60	65.1
Vibrio Harveyi	18	20
Myxobacteria	3	9.13

TABLE II: Mean Absolute Error (MAE) of the ANN and SR models in the validation dataset.

Thereafter, Figures 4, 5 and 6 depict the estimation of the developed ML models in the validation dataset, for each one of the examined diseases. A graphical comparison of the curves of the aforementioned figures confirms the fact that the ANN models *outperform* the SR models. This is, perhaps, more evident in the case of fish-deaths caused by Vibrio Harveyi and Myxobacteria (Figures 5 and 6), where the estimation of the corresponding SR models is unfortunately quite poor.

Let us now turn to the symbolic knowledge of the trained SR models. Accordingly, we present subsequently the mathematical expressions encoded into each SR model developed for the estimation of fish-deaths caused by Pasteurella, Vibrio Harveyi and Myxobacteria:<sup>5</sup>

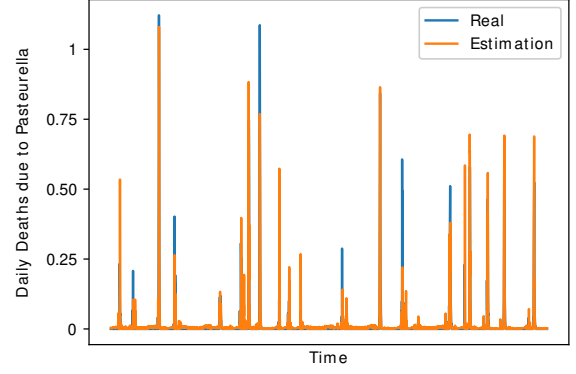
$$Deaths\_Past_{+3} = -2 \cdot Deaths\_Past_0 \cdot \sqrt{Temp_{-3}} + Deaths\_Past_0 - \frac{Deaths\_Past_0}{Food_{-4}}$$

$$Deaths\_Vib_{+3} = -VACCINE\_I_{-1} + ANTIBIOTIC\_I_{-1}$$

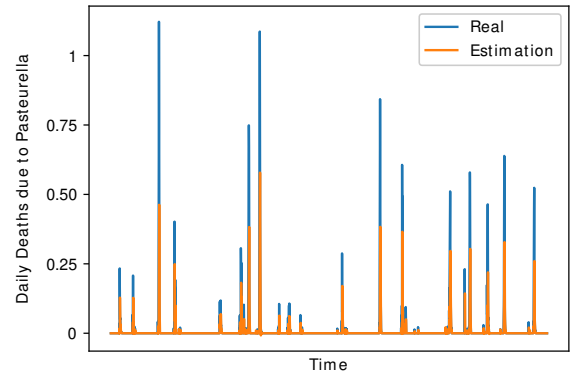
$$Deaths\_Myx_{+3} = Deaths\_Myx_{-9} \cdot MED\_FOOD\_I_{-1}$$

The first thing worth noting is that all three mathematical expressions are quite simple and comprehensible. Some comments on these equations are in order. The first equation asserts that the estimated number of fish-deaths due to Pasteurella on the 3-rd forthcoming day ( $Deaths\_Past_{+3}$ ) is a function of the number of fish-deaths due to Pasteurella recorded today ( $Deaths\_Past_0$ ), of the water temperature recorded three days ago ( $Temp_{-3}$ ), as well as of the amount of food administered four days ago ( $Food_{-4}$ ). The second equation asserts that the estimated number of fish-deaths due to

<sup>5</sup>Due to the randomness employed by the SR models, not all runs converge into the same mathematical expressions. The mathematical expressions shown herein are indicative.



ANN model



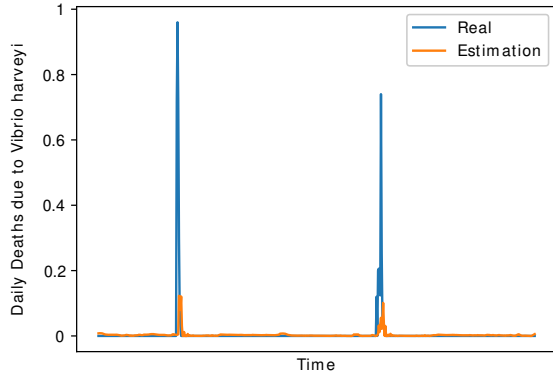
SR model

Fig. 4: Estimation (in the validation set) of the ANN and SR models for the number of fish-deaths on the 3-rd forthcoming day, due to Pasteurella.

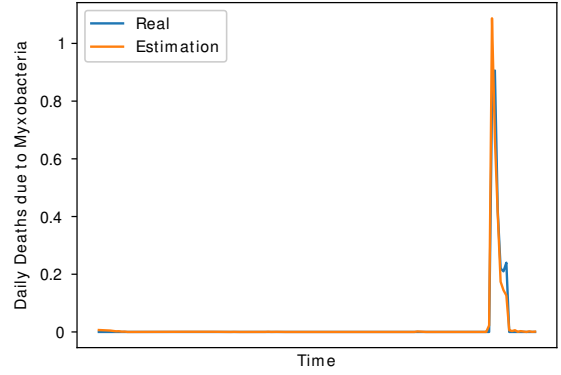
Vibrio Harveyi on the 3-rd forthcoming day ( $Deaths\_Vib_{+3}$ ) is a function of the amount of the vaccine “VACCINE\_I” administered yesterday ( $VACCINE\_I_{-1}$ ), as well as of the amount of the antibiotic “ANTIBIOTIC\_I” administered yesterday ( $ANTIBIOTIC\_I_{-1}$ ). Lastly, the third equation asserts that the estimated number of fish-deaths due to Myxobacteria on the 3-rd forthcoming day ( $Deaths\_Myx_{+3}$ ) is a function of fish-deaths due to Myxobacteria recorded nine days ago ( $Deaths\_Myx_{-9}$ ), as well as of the amount of the medicated food “MED\_FOOD\_I” administered yesterday ( $MED\_FOOD\_I_{-1}$ ).<sup>6</sup> The mathematical equations presented above, essentially, offer fish-farmers an accurate and measurable means to describe the progression of fish-mortalities caused by infectious diseases.

We close this section with the presentation of the results concerning the *total* number of fish-deaths (due to all causes), during the examined time-period. Accordingly, Figure 7 de-

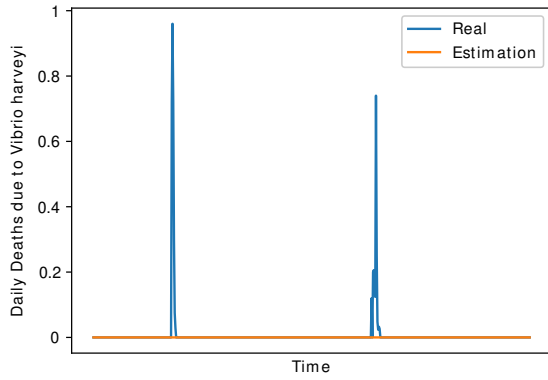
<sup>6</sup>The units of measurement of all variables involved in the equations are identical to those of the available data, discussed in Section II.



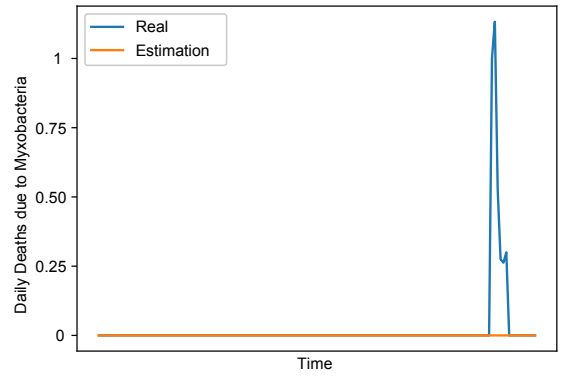
ANN model



ANN model



SR model



SR model

Fig. 5: Estimation (in the validation set) of the ANN and SR models for the number of fish-deaths on the 3-rd forthcoming day, due to Vibrio Harveyi.

Fig. 6: Estimation (in the validation set) of the ANN and SR models for the number of fish-deaths on the 3-rd forthcoming day, due to Myxobacteria.

picts the estimation of the developed ML models in the *validation dataset*, for the total number of fish-deaths on the *next* day. For that case, the MAE of the corresponding ANN model is 1900, whereas, the MAE of the SR model is 2070. It follows then that, in that circumstance also, the ANN model *outperforms* the SR model, a fact which is also confirmed from the graphical comparison of the curves of Figure 7.

Lastly, the mathematical expression encoded into the SR model developed for the total number of fish-deaths is the following:

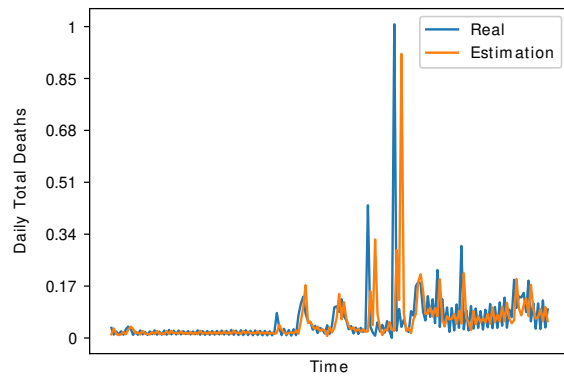
$$D_{+1} = \sqrt{\left(D_{-1} + \sqrt[1/4]{1.8 \cdot \sqrt{D_0 - D_0 - D_{-3}}}\right) \cdot (2 \cdot D_0 - D_{-2} + 0.6)}$$

The derived mathematical expression asserts that the total number of fish-deaths on the next day ( $D_{+1}$ ) is a (rather complex) function of the total number of fish-deaths recorded on each one of the last four days ( $D_0$ ,  $D_{-1}$ ,  $D_{-2}$  and  $D_{-3}$ ).

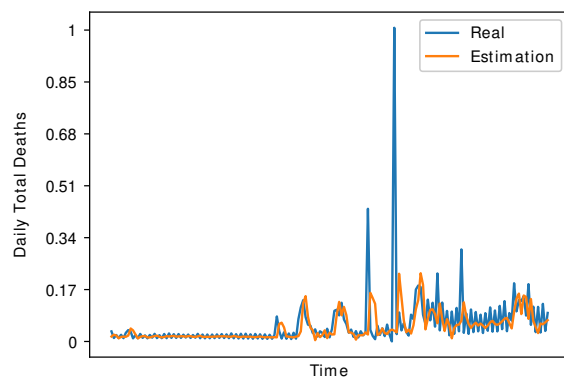
## V. CONCLUSION

In this article, we presented a novel comparison of a collection of Artificial Neural Networks (ANNs) —particularly, Multilayer Perceptrons— and a collection of Symbolic Regression (SR) models, all developed for predicting fish-deaths resulting from infectious diseases in aquaculture. The overall study demonstrated that ANNs outperform SR models for each disease under consideration, achieving a lower Mean Absolute Error (MAE) compared to the latter. However, the increased performance of ANNs in terms of MAE is being paid by their black-box nature, making it challenging to interpret their internal workings and decision-making processes. In contrast, the SR models, despite their lower performance, were able to derive transparent and comprehensible mathematical expressions for each disease. The overall study not only provides a system that generates early warning-alerts of pathogenetic circumstances, contributing to the development of more effective disease-management strategies in the fish-farming industry, but also offers valuable insights into the





ANN model



SR model

Fig. 7: Estimation (in the validation set) of the ANN and SR models for the total number of fish-deaths on the next day.

strengths and limitations of both ANNs and SR models.

We note that the comparison between ANN and SR models conducted in this study is not intended to identify these ML models as competitors. Instead, our intention is to demonstrate how the two models can complement each other to obtain a more accurate and robust estimation of the number of fish-deaths.<sup>7</sup> Indeed, a fish-farming unit can utilize both ML models in conjunction, depending on the specific needs of a task and the desired level of transparency.

Future work is to be devoted to the evaluation of more advanced ANN-based models and more sophisticated SR models (such as the “AI Feynman” [11]) on the same domain.

#### ACKNOWLEDGMENTS

This work is supported by the action “Improving Competitiveness of the Greek Fish Farming Through Development of Intelligent Systems for Disease Diagnosis and Treatment

<sup>7</sup>For example, if both models estimate that a large number of fish-deaths are expected on a particular day, due to an infectious disease, then it would be reasonable to assume that this will be the case.

#### REFERENCES

- [1] Ramón Rueda Delgado, Luis G. Baca Ruíz, Manuel Pegalajar Cuéllar, Miguel Delgado Calvo-Flores, and María del Carmen Pegalajar Jiménez. A comparison between NARX neural networks and symbolic regression: An application for energy consumption forecasting. In Jesús Medina, Manuel Ojeda-Aciego, José Luis Verdegay, Irina Perfilieva, Bernadette Bouchon-Meunier, and Ronald R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications*, pages 16–27. Springer, 2018.
- [2] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.
- [3] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR) 2015*, 2015.
- [4] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, 1992.
- [5] Kevin D. Lafferty, Catherine D. Harvell, Jon M. Conrad, Carolyn S. Friedman, Michael L. Kent, Armand M. Kuris, Eric N. Powell, Daniel Rondeau, and Sonja M. Saksida. Infectious diseases affect marine fisheries and aquaculture economics. *Annual Review of Marine Science*, 7:471–496, 2015.
- [6] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, 2015.
- [7] Daoliang Li, Xin Li, Qi Wang, and Yinfeng Hao. Advanced techniques for the intelligent diagnosis of fish diseases: A review. *Animals*, 12, 2022.
- [8] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324:81–85, 2009.
- [9] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.
- [10] Albert G. J. Tacon and Marc Metian. Fish matters: Importance of aquatic foods in human nutrition and global food supply. *Reviews in Fisheries Science*, 21:22–38, 2013.
- [11] Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6, 2020.
- [12] Gregory J. Zdaniuk, D. Keith Walters, Rogelio Luck, and Louay M. Chamra. A comparison of artificial neural networks and symbolic-regression-based correlations for optimization of helically finned tubes in heat exchangers. *Journal of Enhanced Heat Transfer*, 18:115–125, 2011.
- [13] Zhiming Zhu, Chao Duan, Chuanfu Dong, Shaoping Weng, and Jianguo He. Epidemiological situation and phylogenetic relationship of *Vibrio harveyi* in marine-cultured fishes in China and Southeast Asia. *Aquaculture*, 529, 2020.