

Intelligent Bio-Inspired Detection of Food Borne Pathogen by DNA Barcodes: The Case of Invasive Fish Species *Lagocephalus Sceleratus*

Konstantinos Demertzis^(✉) and Lazaros Iliadis^(✉)

Department of Forestry & Management of the Environment & Natural Resources,
Democritus University of Thrace, 193 Pandazidou st. 68200, Orestiada, Greece
{kdemertz, liliadis}@fmenr.duth.gr

Abstract. Climate change combined with the increase of extreme weather phenomena, has significantly influenced marine ecosystems, resulting in water overheating, increase of sea level and rising of the acidity of surface waters. The potential impacts in the biodiversity of sensitive ecosystems (such as Mediterranean sea) are obvious. Many organisms are under extinction, whereas other dangerous invasive species are multiplied and thus they are destroying the ecological equilibrium. This research paper presents the development of a sophisticated, fast and accurate Food Pathogen Detection (FPD) system, which uses the biologically inspired Artificial Intelligence algorithm of Extreme Learning Machines. The aim is the automated identification and control of the extremely dangerous for human health invasive fish species “*Lagocephalus Sceleratus*”. The matching is achieved through extensive comparisons of protein and DNA sequences, known also as DNA barcodes following an ensemble learning approach.

Keywords: Extreme learning machines · Ensemble learning · Food pathogen detection · DNA barcoding · *Lagocephalus sceleratus* · Invasive species · Climate change

1 Introduction

1.1 Climate Change and Invasive Species

The effect of climate change in marine ecosystems is obvious in various levels of biological organization and especially in the disturbances of biodiversity and in the extinction of organisms due to the appearance of invasive species (IS). The IS are intruders in new strange for them habitats where they can disturb the natural flora and fauna, harming the environment. The social and financial consequences are considered very crucial. For example they can affect human health, agriculture, fishing and food production. Especially, regarding marine ecosystems, various species like fish are traveling searching for colder water, either due to the fact that their natural environment does not satisfy the range of temperatures required for their survival, or because they follow various plant-organism species moving to colder waters [1].

The *Lagocephalus Sceleratus* is a characteristic case of invasive species whose presence in the Mediterranean Sea, causes serious problems [2]. It is probably the

© Springer International Publishing Switzerland 2015

L. Iliadis and C. Jayne (Eds.): EANN 2015, CCIS 517, pp. 89–99, 2015.

DOI: 10.1007/978-3-319-23983-5_9

most dangerous from the 29 “lessepsian” immigrants that passed in the Mediterranean Sea through the Suez Canal. Its uncontrolled invasion and its reproduction, threatens the marine environment with an irreparable imbalance. Its presence causes an intense competition with the native fish regarding the available food. Moreover its consumption by humans is extremely dangerous as it contains the toxic substance “tetradoxine” which causes stomach pain, then diarrhea, vomiting, breathlessness, paralysis and even death. Pattern recognition of the *Lagocephalus Sceleratus* exclusively with phenotypic markers is an extremely difficult and dangerous process. This is due to the fact that this species is unknown to the public and also neither the big morphological differences nor the significant similarities reflect the affinity of the organisms. The need for a comprehensive and absolutely valid identification of the specific species is vital in cases of food production or food canning as it is obviously a serious threat for public health.

1.2 Literature Review

Akova et al. [3] proposed a novel method using an advanced statistical model that boosts the ability of computers to detect the presence of bacterial contamination in tested samples. These formulae drive machine-learning, making possible the identification of known and unknown classes of food pathogens. In [4] an optical sensor system is proposed, for rapid and noninvasive classification of FBP. This system was used for image acquisition and focused in comparing its potential with multivariate calibrations in classifying three categories of popular bacteria. Both linear (LDA, KNN and PLSDA) and nonlinear (BPANN, SVM and OSELM) pattern recognition methods were employed comparatively for modeling. Also [5] describes an application of the BARDOT (Bacteria Rapid Detection using Optical scattering Technology), which is a technique for pathogen belongs to the broad class of optical sensors and relies on forward-scatter phenotyping to classify bacteria belonging to the *Salmonella* class in a non-exhaustive framework. They use a Bayesian approach in learning with a non-exhaustive training dataset to allow for the automated detection of unknown bacterial classes. Similarly [6] demonstrates a pattern recognition technique to classify particles on the basis of their discrete scatter patterns collected at just five different angles, and accompanied by the measurement of axial light loss. The proposed approach can be potentially used with existing instruments because it requires only the addition of a compact enhanced scatter detector. It has been shown that information provided just by five angles of scatter and axial light loss can be sufficient to recognize various bacteria with 68-99% success rate.

In [7] Pan Yi discusses the use of machine learning methods with various advanced encoding schemes and classifiers in order to improve the accuracy of protein structure prediction. Also [8] proposes a machine learning method for classifying DNA-binding proteins from non-binding proteins based on sequence information. Finally paper [9] introduces three ensemble machine learning methods for analysis of biological DNA binding by transcription factors (TFs). The goal is to identify both TF target genes and their binding motifs. Subspace-valued weak learners (formed from an ensemble of different motif finding algorithms) combine candidate motifs as probability weight matrices (PWM), which are then translated into subspaces of a DNA k-mer (string)

feature space. Assessing and then integrating highly informative subspaces by machine methods gives more reliable target classification and motif prediction.

2 Conventional Methods for Food Pathogen Detection

Conventional methods for testing food-borne pathogens are based on the cultivation of pathogens, a process that is complicated and time consuming. Three methods for Food Pathogen Detection can be used:

- ✓ Polymerase Chain Reaction (PCR) sequencing,
- ✓ Restriction Fragment Length Polymorphisms (RFLP) and
- ✓ DNA barcoding.

The most common method, PCR, involves extracting DNA from the food followed by amplification of specific pieces of DNA through an enzymatic process. The amplified DNA fragments are separated by size using a technique called agarose gel electrophoresis and are compared with DNA fragments of known size to enable their identification.

In RFLP analysis, the DNA sample is broken into pieces by restriction enzymes i.e. enzymes that can recognize specific base sequences in DNA and cut the DNA at that site (the restriction site). The resulting restriction fragments are separated according to their size using gel electrophoresis. RFLP analysis was the first DNA profiling technique inexpensive enough for widespread application.

DNA barcoding is a molecular based system, which is based on the analysis of a short genetic marker called the “DNA barcode” in an organism's DNA. The most commonly used barcode region, for animals, at least, is a segment of approximately 600 base pairs of the mitochondrial gene cytochrome oxidase I (COI). By comparing the DNA barcode to a compiled database of barcodes it can be identified as belonging to a particular species. It differs from molecular phylogeny in that the main goal is not to determine patterns of relationship but to identify an unknown sample in terms of a preexisting classification. So there is demand for alternative methods to test for food-borne pathogens that are simpler, quick and applicable to a wide range of potential applications [10].

3 Novelty of the Proposed Intelligent System

The main contribution of this research is the development of the sophisticated ISDNAEF approach (Intelligent System for DNA Extraction in Food testing) embedded in Droplet Digital System [11], for the automated recognition of the *Lagocephalus Sceleratus* fish. This process includes extended comparisons of proteins and DNA sequences (barcodes). Initially a mapping of the existing DNA barcodes to a set of spatial points is done, in order to determine the complex relations that characterize these datasets. Pattern recognition in this research is performed by Employing Ensemble Learning (ENL) an approach that incorporates the biologically inspired Extreme Learning Machine (ELM) Artificial Intelligence (AI) algorithm. ELMs are a modern, fast and reliable Machine Learning (ML) approach used successfully for

classification problems. A comparative analysis of the convergence of the ISDNAEF bio-inspired algorithm with the performance of other relative ML methodologies has shown its reliability for cases of high complexity datasets.

The actual innovation of this research is the AI incorporation in the digital machines that perform the accurate identification, offering safety in food production and protection of consumers' health. Not only the DNA extraction is done extremely accurately but also the required time is minimized significantly, due to the exploitation of the ELMs.

Another strong advantage of this research is the gathering and selection of data which is done not only by employing heuristic and bibliographic approaches but also with the use of the FASTA algorithm [12] that creates scenarios of high complexity and rational data sets that can generalize to new ones.

Feature selection was performed by using Particle Swarm Optimization (PSO) [13] in order to have the optimal convergence and the minimum requirement of computational resources. To the best of our knowledge the ELMs are used for the first time in the case of a Food borne Pathogen Detection System, performing DNA barcode analysis.

4 Machine Learning Methodologies

4.1 Ensemble Learning

Ensemble methods [14] use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms. Usually they refer only to a concrete finite set of alternative models, but typically they allow for much more flexible structures to exist between those alternatives. Also, they are primarily used to improve the performance of a model, or to reduce the likelihood of an unfortunate selection of a poor one. Other applications of ensemble learning include assigning a confidence to the decision made by the model, selecting optimal (or near optimal) features, data fusion, incremental learning, non-stationary learning and error-correcting. The novel concept of combining learning algorithms is proposed as a new direction of ensemble methods for the improvement of the performance of individual algorithms. These algorithms could be based on a variety of learning methodologies and could achieve different ratios of individual results. The goal of the ensembles of algorithms is to generate more certain, precise and accurate system results. Numerous methods have been suggested for the creation of ensembles of learning algorithms:

- ✓ Using different subsets of training data with a single learning method.
- ✓ Using different training parameters with a single training method (e.g. using different initial weights or learning methods for each neural network in an ensemble).
- ✓ Using different learning methods.

Herein the 1st approach was applied in order to develop the ELM ensembles.

4.2 Extreme Learning Machines (ELM)

The ELMs are applying the Single hidden Layer Feed Forward Networks' approach (SLFNs) and the conventional (both single hidden layer and multi hidden layer) feed forward ANN one, with the particularity that the hidden layer (known as feature mapping) of the SLFNs does not necessarily work in concert. In other words the hidden nodes/neurons of the generalized feed forward networks might have been developed randomly. Also all the hidden node parameters are independent from the target functions or from the training datasets. The output weights of the ELMs might be determined in various ways (e.g. with or without iterations, with or without incremental implementations). Finally the ELMs might create the hidden node/neuron parameters randomly before seeing the training data vectors and they can handle non-differentiable activation functions without any problems, like stopping criterion, learning rate and learning epochs [15].

5 Description of the Proposed Algorithm

Before applying ensembles of algorithms using different subsets of training data with the ELM learning method, resampling of the datasets was performed by employing the Bootstrap with replacement methodology. Bootstrapping is a statistical method for estimating the sampling distribution of an estimator by sampling with replacement from the original sample, with the purpose of deriving robust estimates of standard errors and confidence intervals of a population parameter, like mean, median, proportion, odds ratio and correlation coefficient. In the methodology proposed herein, three successive data resampling's were performed in order to obtain the average performance of the algorithm and determine its accuracy. The overall algorithmic approach that was proposed herein is described clearly and in details in the figure 1.

Subsequently, the ELMs were used in order to classify the developed datasets. This was done by employing the Gaussian Radial Basis Function kernel according to equation 1 [16].

$$K(u,v)=\exp(-\gamma\|u-v\|^2) \quad (1)$$

The hidden neurons $k=20$, w_i are the assigned random input weights and b_i the biases, where $i=1,\dots,N$ and H is the hidden layer output matrix.

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & h_L(x_1) \\ \vdots & & \vdots \\ h_1(x_N) & \cdots & h_L(x_N) \end{bmatrix} \quad (2)$$

$h(x) = [h_1(x), \dots, h_L(x)]$ is the output (row) vector of the hidden layer with respect to the input x .

Function $h(x)$ actually maps the data from the d -dimensional input space to the L -dimensional hidden-layer feature space (ELM feature space) H and thus, $h(x)$ is

indeed a feature mapping. ELM aims to minimize the training error as well as the norm of the output weights as shown in equation 3 [16]:

$$\text{Minimize : } \|H\beta - T\|^2 \text{ and } \|\beta\| \tag{3}$$

To minimize the norm of the output weights $\|\beta\|$ is actually to maximize the distance of the separating margins of the two different classes in the ELM feature space $2/\|\beta\|$. The calculation of the output weights β is done according to equation 4 [16]:

$$\beta \left(\frac{1}{c} + H^T H \right)^{-1} H^T T = \tag{4}$$

where c is a positive constant and T is obtained from the *Function Approximation of SLFNs* with additive neurons

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} \tag{5}$$

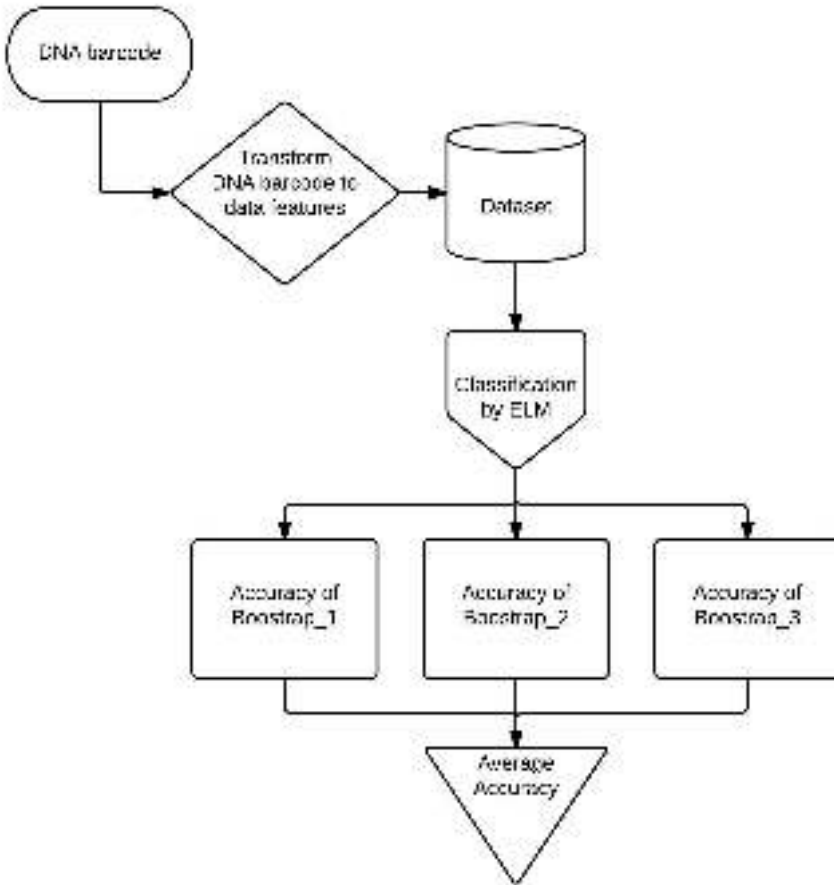
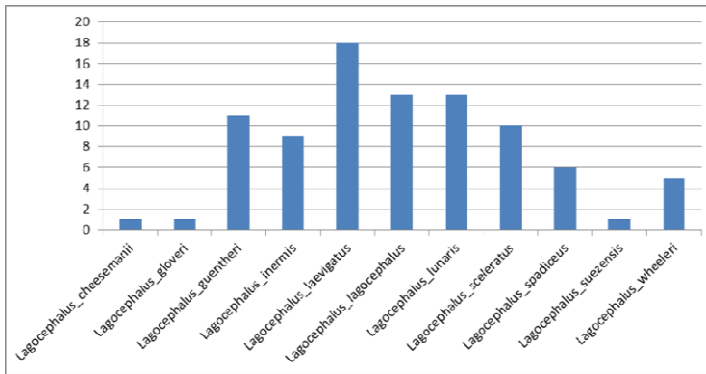


Fig. 1. The structure of the proposed algorithm

6 Data and Testing

Three datasets were developed to be used in the training process of the algorithm. These datasets emerged from the numerical conversion of the genetic information encoded as A, T, C and G (abbreviations of adenine, thymine, cytosine, and guanine) included in the DNA barcodes of the species under examination [17] [18].

The first *lago_family_dataset* (Lagocephalus Family Dataset) comprised of 582 independent parameters and 11 classes, coming from the DNA barcodes 88 fish samples of the Lagocephalus family as shown in the following graph 1. It should be mentioned that three out of the eleven classes include only one case.

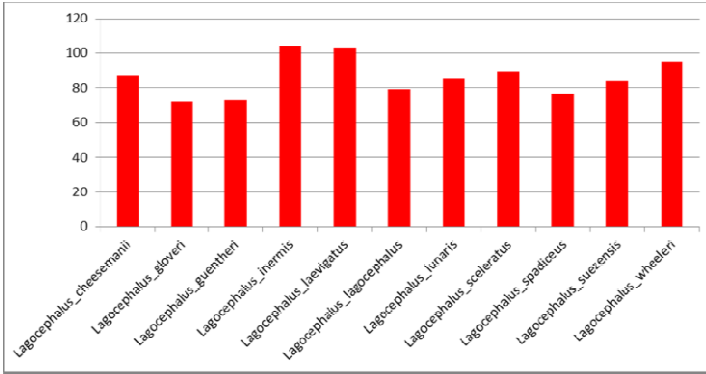


Graph 1. Original Lagocephalus Family Dataset

Actually an imbalanced dataset was developed. Imbalanced datasets are a special case of classification where the distribution is not uniform among the classes. The problem in such cases is that standard learners are often biased towards the majority class. That is because these classifiers attempt to reduce global quantities such as the error rate, without taking the data distribution into consideration. As a result examples from the overwhelming class are well-classified whereas examples from the minority class tend to be misclassified.

The Synthetic Minority Oversampling technique (SMOTE) was employed to resolve the above problem by performing re sampling of the minority class [19]. Re sampling the minority class provides a simple way of biasing the generalization process. It can do so by generating synthetic samples accordingly biased and controlling the amount and placement of the new samples. SMOTE is a technique which combines informed oversampling of the minority class with random undersampling of the majority class. Also it produces the best results as far as re-sampling and modifying the probabilistic estimate techniques. For each minority sample, SMOTE works as follows:

- ✓ Find its k -nearest minority neighbors.
- ✓ Randomly select j of these neighbors.
- ✓ Randomly generate synthetic samples along the lines joining the minority sample and its j selected neighbors (j depends on the amount of oversampling desired).



Graph 2. Lagocephalus Family Dataset after application of SMOTE

The *lago_family_dataset*, was re-created by following the SMOTE technique and its new version comprised of 948 cases, distributed in classes as shown in the following graph 2.

Additionally the *usual_fish_dataset* was created, including 1127 fish samples using 519 independent parameters and 14 classes representing widely used Mediterranean fish that can be confused as *Lagocephalus Sceleratus*.

Finally extended protein and DNA sequence comparisons related to *Lagocephalus Sceleratus*, were performed with fish species that have high genetic similarities, in order to create a new dataset of high complexity by following the Protein Similarity Search (PSS) method. PSS method provides sequence similarity searching against protein databases using the FASTA algorithm. The FASTA algorithm searches protein and deoxyribonucleic acid (DNA) databases for sequences with statistically significant similarity and compares proteins, DNA, short peptides and oligonucleotides [20].

The *lago_fasta_dataset* (*Lagocephalus* FASTA Dataset) that have emerged includes the DNA barcode of 772 fish, related to 558 independent parameters and 16 classes, corresponding to fish of high genetic similarity with *Lagocephalus Sceleratus*.

6.1 Results

The basic hypothesis that the use of datasets with high genetic similarity produces lower efficiency machine learning models was confirmed. We have found higher correlation between the independent parameters and the classes. More specifically for the *usual_fish_dataset*, whose data are related to usual Mediterranean fish with high level of phenotype similarity and low similarity of genetic indices to the *Lagocephalus Sceleratus* Machine Learning algorithms managed to classify the data with an accuracy as high as 99%. Respectively in the case of the *lago_family_dataset* which includes only data of the *Lagocephalus* species the classification performance was quite lower which is something due to the high genetic correlation between the fish species under examination. This led to high complexity between the independent parameters and the dataset classes. This problem was re confirmed more intensively in the *lago_fasta_dataset* where the data had the highest similarity index in terms of protein and DNA sequences. These similarities were determined after using the

Protein Similarity Search method and the FASTA algorithm. However the fact that the classification efficiency results still remain very high (in the worst case 94%) is very encouraging. This fact enhances and confirms that our proposal to embed AI algorithms in automatic devices for DNA extraction for food testing is correct.

6.1.1 Comparative Analysis

An analysis was performed for the datasets under consideration, between the ELMs method that we have used, versus Multilayer Perceptron (MLP), Support Vector Machines (SVM) and Random Forest (RF). The results are presented in the following table 1. In this case the hold out approach was used (70% Training, 15% Validation and 15% Testing).

Table 1. Accuracy (ACC) and Performance Matrices (PM) comparison between ELM, MLP, SVM and RF algorithms in *usual_fish_dataset*

Classifier	<i>usual_fish_dataset</i>			<i>lago_family_dataset</i>			<i>lago_fasta_dataset</i>		
	ACC and PM			ACC and PM			ACC and PM		
	ACC	RMSE	ROC	ACC	RMSE	ROC	ACC	RMSE	ROC
ELM	98.6%	0.0416	0.998	96.7%	0.0879	0.987	95.9%	0.0810	0.983
MLP	98.2%	0.0473	0.994	95.6%	0.1091	0.986	94.4%	0.0899	0.979
SVM	97.1%	0.0635	0.985	94.8%	0.1139	0.973	94.1%	0.0817	0.980
RF	97.9%	0.0609	0.994	95.5%	0.1055	0.974	94.8%	0.0826	0.979

Taking into consideration the superiority of the ELMs for every dataset we have used the ensembles of ELMs approach with different subsets of training data. The Bootstrap with replacement was used for the re-sampling three times for each dataset and the final average classification accuracy was considered. The results are displayed in table 2.

Table 2. Accuracy (ACC) and Performance Matrices (PM) of ELM after resampling by Bootstrap with replacement method

	<i>usual_fish_dataset</i>			<i>lago_family_dataset</i>			<i>lago_fasta_dataset</i>		
	ACC and PM			ACC and PM			ACC and PM		
	ACC	RMSE	ROC	ACC	RMSE	ROC	ACC	RMSE	ROC
Bootstrap_1	99.1%	0.0314	0.999	96.9%	0.0647	0.990	95.8%	0.0768	0.987
Bootstrap_2	98.6%	0.0418	0.998	96.8%	0.0662	0.994	96.1%	0.0715	0.994
Bootstrap_3	99.0%	0.0394	0.998	97.0%	0.0658	0.995	96.2%	0.0709	0.995
Average	98.9%	0.0375	0.998	96.9%	0.0656	0.993	96.0%	0.0731	0.992

Due to the fact that the number of the used features was too high for all three datasets (*lago_family_dataset* 582 independent parameters, *usual_fish_dataset* 519 and *lago_fasta_dataset* 558) several feature selection attempts were done for the reduction of the training time and for the enhancement of the generalization in order to avoid overfitting. The Particle Swarm Optimization was used to search for the optimal feature subset. The assessment of each subset was done by considering the value of each subset which is based on the contribution and the degree of redundancy of each cha-

racteristic. The parameters considered for the final decision are related to the classification accuracy and to the correlation of the classification errors in comparison to the accuracy of the initial parameters set. After the feature selection the *lago_family_dataset* has 231 independent parameters (reduction by 60%), the *usual_fish_dataset* 197 ones (reduction by 62%) and the *lago_fasta_dataset* 235 (reduction by 57%). The final results are presented in the following table 3.

Table 3. Accuracy (ACC) and Performance Matrices (PM) of ELM after feature selection by PSO search method

	<i>usual_fish_dataset</i>			<i>lago_family_dataset</i>			<i>lago_fasta_dataset</i>		
	ACC and PM			ACC and PM			ACC and PM		
	ACC	RMSE	ROC	ACC	RMSE	ROC	ACC	RMSE	ROC
Bootstrap_1	99.3%	0.0309	0.999	97.0%	0.0642	0.991	96.2%	0.0688	0.995
Bootstrap_2	98.9%	0.0406	0.999	97.1%	0.0639	0.996	96.3%	0.0685	0.994
Bootstrap_3	99.2%	0.0382	0.999	97.3%	0.0631	0.996	96.4%	0.0679	0.996
Average	99.1%	0.0366	0.999	97.1%	0.0637	0.994	96.3%	0.0684	0.995

The final conclusion is that the proposed method has proven to be reliable and efficient and has outperformed at least for these datasets the other approaches.

7 Discussion–Conclusions

This research paper presents an innovative ensemble classifier method using ELMs which was embedded successfully in an advanced food pathogen detection system, aiming in the pattern recognition of the dangerous invasive species *Lagocephalus Sceleratus* by digital systems. Machine Learning was employed successfully that performs the pattern recognition based on extended comparisons of protein and DNA sequences known as bar codes. It is really important to identify this species in food production and canning. The proposed algorithm uses a sophisticated technique of combined learning, improving the potential efficiency of the classification and the generalization ability. The whole process is automated by running fast bio-inspired machine learning approaches instead of complicated and costly biochemical processes. The most significant innovation of this methodology is that it uses embedded AI algorithms in digital control machines of food quality control that ensure public health. The system was tested thoroughly under various scenarios showing very high accuracy and reliable performance. Future research could involve its extension under a hybrid scheme, which will combine semi supervised methods and online learning for the trace and exploitation of hidden knowledge between the inhomogeneous data that might emerge. Also the use of the unsupervised learning for the re sampling of the datasets will create subsets of higher homogeneity, that will potentially train a model with higher accuracy. Finally, it would be very interesting to perform optimization of the proposed algorithm, by using heuristic approaches like genetic or particle swarm optimization based selective ensemble.

References

1. Frank, J.R., Olden, J.D.: Assessing the Effects of Climate Change on Aquatic Invasive Species. *Conservation Biology* **22**(3), 521–533 (2008). doi:10.1111/j.1523-1739.2008.00950.x. Society for Conservation Biology
2. Kheifets, J., Rozhavsky, B., Solomonovich, Z.G., Rodman, M., Soroksky, A.: Severe Tetrodotoxin Poisoning after Consumption of *Lagocephalus sceleratus* (Pufferfish, Fugu) Fished in Mediterranean Sea, Treated with Cholinesterase Inhibitor. *Case Reports in Critical Care* **2012**, Article ID 782507, 3 p. (2012). doi:10.1155/2012/782507
3. Akova, F., Dundar, M., Davisson, V.J., Hirtleman, D.E., Bhunia, A.K., Robinson, J.P., Rajwa, B.: A Machine-Learning Approach to Detecting Unknown Bacterial Serovars. *Statistical Analysis and Data Mining* (2011). doi:10.1002/sam.10085
4. Pan, W., Zhao, J., Chen, Q.: Classification of foodborne pathogens using near infrared laser scatter imaging system with multivariate calibration (2015). doi:10.1038/srep09524
5. Rajwa, B., Dundar, M.M., Akova, F., Bettasso, A., Patsekina, V., Hirtleman, E.D., Bhunia, A.K., Robinson, J.P.: Discovering the Unknown: Detection of Emerging Pathogens Using a Label-Free Light-Scattering System. *Cytometry Part A* **77A**, 1103–1112 (2010)
6. Rajwa, B., Venkatapathi, M., Ragheb, K., Banada, P.P., Hirtleman, E.D., Lary, T., Robinson, J.P.: Automated classification and recognition of bacterial particles in flow by multi-angle scatter measurement and a support-vector machine classifier. *Cytometry A* **73**(4), 369–379 (2008). doi:10.1002/cyto.a.20515
7. Pan, Y.: Protein structure prediction and understanding using machine learning methods. In: 2005 IEEE Granular Computing, vol. 1 (2005). doi:10.1109/GRC.2005.1547225
8. Ma, X., Hu, L.: Extracting sequence features to predict DNA-binding proteins using support vector machine. In: 2013 Fifth International Conference on Computational and Information Sciences (ICICIS) (2013). doi:10.1109/ICICIS.2013.48
9. Yu, D.-J., Hu, J., Li, Q.M., Tang, Z.M., Yang, J.Y., Shen, H.B.: Constructing Query-Driven Dynamic Machine Learning Model With Application to Protein-Ligand Binding Sites Prediction. *NanoBioscience, IEEE*, **14**(1) (2015)
10. Leigh, D., Thredgold, E.A.V., Lenehan, C.E.: Direct detection of histamine in fish flesh using microchip electrophoresis with capacitively coupled contactless conductivity detection. *Anal. Methods*, 1802–1808 (2015). doi:10.1039/C4AY02866J
11. <http://www.bio-rad.com/>
12. Lipman, D.J., Pearson, W.R.: Rapid and sensitive protein similarity searches. *Science* **227**(4693), 1435–1441 (1985). doi:10.1126/science.2983426. PMID 2983426
13. Moraglio, A., Di Chio, C., Poli, R.: Geometric Particle Swarm Optimization **2008**, Article ID 143624, 14 p. (2008). doi:10.1155/2008/143624
14. Rokach, Lior: Ensemble-based classifiers. *Artificial Intelligence Review* **33**(1–2), 1–39 (2010). doi:10.1007/s10462-009-9124-7
15. Cambria, E., Huang, G.-B.: Extreme Learning Machines. *IEEE Intelligent Systems* (2013)
16. Huang, G.-B.: An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels (2014). doi:10.1007/s12559-014-9255-2, Springer
17. <http://www.cabi.org/isc/>
18. <http://www.boldsystems.org/>
19. Nitesh, V., Chawla, B.K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* **16** (2002)
20. <http://www.ebi.ac.uk/Tools/sss/fastal/>