

Extreme deep learning in biosecurity: the case of machine hearing for marine species identification

Konstantinos Demertzis^a, Lazaros S. Iliadis^b and Vardis-Dimitris Anezakis ^c

^aSchool of Engineering, Department of Civil Engineering, Lab of Mathematics and Informatics, Democritus University of Thrace, Xanthi, Greece; ^bSchool of Engineering, Department of Civil Engineering, Lab of Mathematics and Informatics, Democritus University of Thrace, Xanthi, Greece; ^cDepartment of Forestry and Management of the Environment and Natural Resources, Democritus University of Thrace, Orestiada, Greece

Biosafety is defined as a set of preventive measures aimed at reducing the risk of infectious diseases' spread to crops and animals, by providing quarantine pesticides. Prolonged and sustained overheating of the sea, creates significant habitat losses, resulting in the proliferation and spread of invasive species, which invade foreign areas typically seeking colder climate. This is one of the most important modern threats to marine biosafety. The research effort presented herein, proposes an innovative approach for Marine Species Identification, by employing an advanced intelligent Machine Hearing Framework (MHF). The final target is the identification of invasive alien species (IAS) based on the sounds they produce. This classification attempt, can provide significant aid towards the protection of biodiversity, and can achieve overall regional biosecurity. Hearing recognition is performed by using the Online Sequential Multilayer Graph Regularized Extreme Learning Machine Autoencoder (MIGRATE_ELM). The MIGRATE_ELM uses an innovative Deep Learning algorithm (DELE) that is applied for the first time for the above purpose. The assignment of the corresponding class 'native' or 'invasive' in its locality, is carried out by an equally innovative approach entitled 'Geo Location Country Based Service' that has been proposed by our research team.

ARTICLE HISTORY

Received 30 November 2017
Accepted 14 July 2018



KEYWORDS

Biosecurity; invasive species; machine hearing; geolocation-based services; extreme learning machines; big data

1. Introduction

1.1. Bio-safety and bio-pollution

Bio-safety (BISA) is a strategic and integrated approach. It includes policy and regulatory frameworks, plus risk analysis and management rules. It is related to food safety, animal and plant life and health, in certain environments. It covers the introduction of plant pests, animal pests, diseases and zoonoses, and the release of genetically modified organisms. BISA is a hot topic that has emerged recently is the introduction-management of Invasive Alien Species and their genotypes. It is essentially a holistic approach that is

CONTACT Vardis-Dimitris Anezakis  danezaki@fmenr.duth.gr  Department of Forestry and Management of the Environment and Natural Resources, Democritus University of Thrace, Pantazidou 193, PC 68200 Orestiada, Greece

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

directly related to sustainability in areas such as agriculture and food safety and it refers to the protection of the environment, mainly focusing in biodiversity. The threats to biosecurity, are related to small-scale risks that emerge rapidly, making the application of an effective policy, a very challenging task. This is due to time and resources' limitations, which make the analysis and assessment of threats likelihood, a tedious task.

IAS are a result of generalized climate change and they constitute a serious and rapidly worsening threat to natural biodiversity in Europe. They enter new foreign habitats and they can stifle natural flora or fauna, causing serious harm to the environment (Demertzis & Iliadis, 2017a). The impacts are socio-economic and there are negative consequences in public health, in fisheries, in agriculture and more generally in food production. The invasion of these species is the second biggest threat to local biodiversity worldwide and is called 'bio-pollution'. The impact of IAS and their rapid expansion in new seas have direct economic consequences in various areas (e.g. financial) such as the risk of indigenous species' extinction, entails costs to restore natural balance. The expansion of organisms harmful to the human health might lead (in mid-term) in the reduction of tourist development. Ecological impacts such as food grid disruption are very important, and we cannot ignore the risk of introducing new diseases that can destroy sensitive indigenous species. The changes in biodiversity entail a change in relative abundance of species. Moreover, the risk to public health should not be overlooked as these species may be toxic, such as 'Lagocephalus' fish, which contains 'Tetrodotoxin' a very dangerous substance, capable of causing serious health problems, even death (Demertzis & Iliadis, 2015, 2017b). European Union spends at least 12 billion Euros per year, trying to control IAS and to overcome their consequences.

1.2. Our research approach

Modern IT technologies and innovations are introduced by Computational Intelligence (COIN). These automated high-tech solutions, create the preconditions for designing proper biosecurity and biodiversity protection methods that can evolve and reform the existing framework. They literally reinforce and simplify IAS detection processes with Machine Learning (ML) identification systems. They allow fast and easy data collection and thus logging of indigenous and non-resident populations in an area. This is achieved by applying an automated process with low financial requirements. Moreover, they create proper conditions for studying the behaviours of different species and their seasonal variation. In other words, they are helping to accurately map the overall invasions. In this way, they are contributing substantially to slowing the uncontrolled expansion of IAS. The significance of these intelligent models towards the identification and classification of IAS has been supported by some recent researches (Demertzis & Iliadis, 2015; Demertzis & Iliadis, 2017b, 2017c; Demertzis, Iliadis, & Anezakis, 2017).

This study proposes a new identification method for IAS which is presented for the first time in the international literature. It is an advanced intelligent Deep Extreme Learning Machine framework (DELM) towards Marine Specification Identification aided by Machine Hearing (MAHE). The proposed framework uses the Online Sequential Multilayer algorithm (OSML) plus a Graph Regularized Extreme Learning Machine Autoencoder (GRELMA). It is a highly efficient and original DELE architecture that uses a Multilayer Extreme Learning Machine (MELM) with online learning classification capabilities.

Checking whether the recognized item is native or not in its locality, is carried out by an innovative Geo Location Country Based Services (GCBS) (Demertzis et al., 2017).

The implementation of this proposed framework was based on the DELE philosophy. An important aspect of the framework is the usage of ELM which has proven to be capable, of solving a multidimensional and complex IT problem. The Deep ELM simulates the functioning of biological brain cells in a most realistic mode. This creates the potential for a fully automated configuration of the model with high accuracy classification. An innovative aspect of this work is the development of this model using Online Sequential Multi-layer algorithm plus a Graph Regularized Extreme Learning Machine Autoencoder that aims to optimize the choice of the input layer weights and bias. This has been done in order to achieve a higher level of generalization. This method combines two highly effective machine learning algorithms, for solving a multidimensional and complex machine hearing problem. Another interesting point is the performance of feature extraction using an intelligent method of audio signal analysis. The proposed method supports the autonomous operation of the system comprising the geolocation capability (using global position systems) of the identified species and the native ones. This system is pouring artificial intelligence to digital sensors that can easily (quickly and at low cost) identify invasive or rare species on basis of their phenotypes. This would result in strengthening biosecurity programmes.

1.3. Literature review

Through a series of new learning architectures and algorithms, have been transformed; DELE methods are now the state-of-the-art in object, speech and audio recognition. Deng and Yu (2013) had proposed methods and applications of DELE. A large deep convolutional neural network trained (Krizhevsky, Sutskever, & Hinton, 2012) to classify 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into 1000 different classes. In recent years, Convolutional Neural Networks (CNNs) have become very popular and have achieved great success in many computer vision tasks particularly in object recognition. Cellular Simultaneous Recurrent Networks (CSRNs) applied (Alom, Alam, Taha, & Iftekharruddin, 2017) to generate initial filters of CNNs for features extraction and Regularized Extreme Learning Machines (RELM) for classification. Experiments were conducted on three popular data sets for object recognition (such as face, pedestrian, and car) to evaluate the performance of the proposed system.

An object recognition algorithm constructed efficient features automatically without relying on human experts in order to design features for fish species classification by Zhang, Lee, Zhang, Tippetts, and Lillywhite (2016). Results from experiments showed that the proposed method obtained an average of 98.9% classification accuracy, with a standard deviation of 0.96%. The data set comprised of 8 fish species and a total of 1049 images. Moreover, DELE has been the driving force behind large leaps in accuracy and model robustness in audio related domains, like speech recognition.

Hinton et al. (2012) presented four successful DELE research approaches for acoustic modelling in sound recognition, namely: Bing-voice-search speech recognition, Switch-board speech recognition, Google voice input speech recognition and YouTube speech recognition.

Moreover (Hinton et al., 2012) proposed the employment of (Deep Neural Networks (DNNs) to extract high-level features from raw data and to show that they are effective for speech and emotion recognition. They first produced an emotion state probability distribution for each speech segment using DNNs. Then they constructed utterance-level features from segment-level probability distributions. Then these features were fed into an Extreme Learning Machine (ELM), a special simple and efficient single-hidden-layer neural network, to identify utterance-level emotions. The automatic Sound Event Classification (SEC) has attracted a growing attention in recent years. Feature extraction is a critical factor in SEC system, and DNNs. Actually, they have achieved the state-of-the-art performance for SEC. The Extreme Learning Machine Auto Encoder (ELMAE) is a new DELE algorithm, which has both an excellent representation performance and it follows a very fast training procedure. A bilinear multi-column ELMAE algorithm has been proposed by Zhang, Yin, Zhang, Shi, and Li (2017) in order to improve the robustness, stability, and feature representation of the original approach. This method was applied towards feature representation of sound signals. Moreover, a similar ELMAE model combined with a two-stage ensemble learning and classification framework was developed to perform the robust and effective automatic SEC.

Additionally, many studies on automatic audio classification and segmentation are using several features and techniques. Zhao et al. (2017) proposed a new method for automated field recording analysis with an improved segmentation and a robust bird species classification. They used a Gaussian Mixture Model (GMM) with an event-energy-based sifting procedure that selected representative acoustic events. Moreover, they used a classification Support Vector Machine (SVM).

1.4. Paper outline

The remainder of this paper is as follows: A general description of the theoretical background is presented in Section 2. The theoretical concepts of the proposed system are discussed in Section 3. The data sets used and the feature extraction method is included in Section 4. Finally, the marine species identification framework and the experimental setup that was used in this research are discussed in detail in Section 5. Moreover, Section 6 is about the obtained results which are compared to existing approaches. The conclusions are summarized and further discussed in Section 7, whereas future research is shown in Section 8.

2. Theoretical background

2.1. Machine hearing

Machine Hearing (Lyon, 2017) is a scientific field of artificial intelligence that attempts to reproduce the sense of hearing algorithmically. Audio Signal analysis is the general procedure which refers to the extraction of knowledge based on the correlation among the content and the nature of audio signals.

In general, the process involves the extraction of certain features capable to differentiate their values, according to the content and the structure of the corresponding

signals. Based on the application, the final algorithmic step is classification, segmentation, automatic retrieval or synthesis.

2.2. Intelligent methods of audio signal analysis

A MAHE data collection can be so large or so heterogeneous and complex in its structure, that traditional data processing software is impossible to manage. Solving a MAHE problem requires high availability of resources. Critical factors are time complexity of the employed algorithms, memory availability as a function of input data, as well as individual analysis related to other resources, such as the number of parallel processors required to solve the problem.

Therefore, MAHE cases require resolution with modern computational intelligence methods. The analysis of data extracted from audio fragments requires not only large storage space (Ramírez-Gallego, Fernández, García, Chen, & Herrera, 2018) but coordination of complex analysis processes solved in polynomial time as well.

3. Theoretical background of the proposed system

3.1. A brief reference to the extreme learning machines

Before introducing our proposed algorithm, it would be essential to discuss the existing basic theoretical framework. The question of whether a small Neural Network architecture can learn a lot, even from huge training data sets, was answered in the affirmative by ELM. An ELM (Cambria & Guang-Bin, 2013) is a Single-Hidden Layer Feed Forward Neural Network (SLFFNN) with N hidden neurons, randomly selected input weights and random values of bias in the hidden layer, while the weights at its output are calculated with a single multiplication of vector matrices. SLFFNNs are used in ELMs because of their ability to approach any continuous function and to classify any discontinuous areas. An ELM can accurately learn N samples, and its learning speed can be even thousands of times greater than the speed of conventional Back Propagation Feed Forward Neural Networks (BP_FFNN).

ELMs use the SLFFNN's general methodology, with the specificity that the Hidden Layer (feature mapping) is not required to work in a coordinated fashion. All hidden-layer parameters are independent from the activation functions and from the training data.

ELMs can randomly create hidden nodes or hidden level parameters, before seeing the training data, while it is remarkable that they can handle non-differential activation equations and they do not address known NN problems such as stopping criterion, learning rate and learning epochs (Cambria & Guang-Bin, 2013; Huang, 2014, 2015).

A mathematical basis has been provided for the understanding of ELM (Equations (1)–(9)) (Cambria & Guang-Bin, 2013; Huang, 2014, 2015).

For an ELM using SLFFNN and random representation of hidden neurons, input data is mapped to a random L -dimensional space with a discrete training set N , where (x_i, t_i) , $i \in 1, N$ with $x_i \in R^d$ and $t_i \in R^c$. The specification output of the network is the following:

$$f_L(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x)\beta \quad i \in 1, N. \quad (1)$$

Vector matrix $\beta = [\beta_1, \dots, \beta_L]^T$ is the output of the weight vector matrix connecting hidden and output nodes. On the other hand, $h(x) = [g_1(x), \dots, g_L(x)]$ is the output of the hidden nodes for input x , and $g_1(x)$ is the output of the i th neuron. Based on a training set $\{(x_i, t_i)\}_{i=1}^N$, an ELM can solve the Learning Problem $H\beta = T$, where the target labels (target outputs) are $T = [t_1, \dots, t_N]^T$ and the output vector matrix of the Hidden Layer H is the following:

$$H(\omega_j, b_j, x_i) = \begin{bmatrix} g(\omega_1 x_1 + b_1) & \cdots & g(\omega_l x_1 + b_l) \\ \vdots & \ddots & \vdots \\ g(\omega_1 x_N + b_1) & \cdots & g(\omega_l x_N + b_l) \end{bmatrix}_{N \times l}. \quad (2)$$

Input weight vector matrix of hidden layer ω (before training) and bias vectors b are created randomly in the interval $[-1, 1]$, with $\omega_j = [\omega_{j1}, \omega_{j2}, \dots, \omega_{jm}]^T$ and $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}]^T$. Output weight vector matrix of hidden layer H is calculated by the use of the Activation function in the training data set, based on the following function:

$$H = g(\omega x + b). \quad (3)$$

The output weights β can be estimated by using function 4:

$$\beta = \left(\frac{1}{C} + H^T H \right)^{-1} H^T X, \quad (4)$$

where $H = [h_1, \dots, h_N]$ is the output vector matrix of the hidden layer and $X = [x_1, \dots, x_N]$ is the input vector matrix of the hidden layer. Indeed, β can be calculated by the following general relation 5:

$$\beta = H^+ T, \quad (5)$$

where H^+ is the generalized inverse vector matrix Moore–Penrose for matrix H .

3.1.1. Description of the ELM model employed in our approach

This approach is employing ELM with Gaussian Radial Basis Function (RBF) kernel $K(u, v) = \exp(-\gamma \|u - v\|^2)$. The hidden neurons are $k = 20$. Subsequently, the assigned random input weights are w_i and the biases are denoted as b_i , $i = 1, \dots, N$.

To calculate the hidden layer output matrix H we have used the following function (6):

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & h_L(x_1) \\ \vdots & \ddots & \vdots \\ h_1(x_N) & \cdots & h_L(x_N) \end{bmatrix} \quad (6)$$

where $h(x) = [h_1(x), \dots, h_L(x)]$ is the output (row) vector of the hidden layer with respect to the input x . Moreover, $h(x)$ maps the data from the D -dimensional input space to the L -dimensional hidden-layer feature space (ELM feature space) H . Thus, $h(x)$ is indeed a feature mapping. ELM is to minimize the training error as well as the norm of the output weights:

$$\text{Minimize: } \|H\beta - T\|^2 \text{ and } \|\beta\| \quad (7)$$

where H is the hidden-layer output matrix of function (6).

Minimization of the norm of the output weights $\|\beta\|$ is actually achieved by maximizing the distance of separating the margins of the two different classes in the ELM feature space $2/\|\beta\|$.

We used the following function (8) to calculate the output weights β :

$$\beta = \left(\frac{1}{C} + H^T H \right)^{-1} H^T T, \tag{8}$$

where the value of C (a positive constant) and the value of T are obtained from the *Function Approximation of SLFFNs with additive neurons*:

$$t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m \text{ and } T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}. \tag{9}$$

3.2. Online sequential extreme learning machines

The Online Sequential ELM (OSELM) (Huang, Liang, Rong, Saratchandran, & Sundararajan, 2005; Liang, Huang, Saratchandran, & Sundararajan, 2006) is an alternative technique for large scale computing and ML which is employed when data becomes available in a sequential manner to determine mapping to data set corresponding labels. The main difference between the Online Learning (ONL) and the Batch Learning (BL) techniques is that in ONL mapping is updated after the arrival of every new data point in a scale fashion, whereas BL techniques are used when one has access to the entire training data set at once. It is a versatile sequential **L**earning **A**lgorithm (LA) because training observations are introduced sequentially one-by-one, or chunk-by-chunk, with varying or fixed chunk length. At any moment, only the newly arrived single record or chunk are used and learned. A single record or a chunk of training observations is discarded as soon as the learning procedure for that particular (single or chunk) observation(s) is completed. The LA has no prior knowledge as to how many training observations will be presented. Unlike other sequential learning algorithms which have many control parameters to be tuned, OSELM only requires the number of hidden nodes to be specified (Huang et al., 2005; Liang et al., 2006).

The OSELM consists of two main phases namely: Boosting Phase (BPh) and Sequential Learning Phase (SLPh). BPh trains the SLFFNs using the primitive ELM method with some batch of training data in the initialization stage. This set of training data is discarded as soon as the boosting phase is completed. The required batch of training data is very small, which can be equal to the number of hidden neurons (Huang et al., 2005; Liang et al., 2006).

3.3. The innovative OSML-GRELMA proposed algorithm

Considering and combining the features of ELM (that was presented above) we introduce and propose a new Deep architecture by creating an Online Learning Multilayer Graph Regularized Extreme Learning Machine Auto-Encoder (OSML-GRELMA).

This is a multi-layered neural network model that receives successive OL data streams and uses the unsupervised GRELMA algorithm as a basic building block in which the outputs of each level are used as inputs to the next one (Sun, Zhang, Zhang, & Hu, 2017).

An autoencoder is an ANN used for unsupervised learning of efficient coding. The aim of an autoencoder is to learn a representation (encoding) for a set of data, but with the output layer having the same number of nodes as the input layer, and with the purpose of reconstructing its own inputs (instead of predicting target value Y given inputs X). The algorithm is described below (Sun et al., 2017):

Algorithm 1. GRELMA Algorithm for Clustering (Sun et al., 2017)

Input: Data $\{X\} = \{x_i\}_{i=1}^N$ the number of hidden neurons n_h , the penalty coefficient κ and λ
Output: The cluster results.
Step 1: Initialize an ELM of n_h hidden neurons with random input weights and biases.
Step 2: If $n_h \leq N$ Compute the output weights β by employing the following equation

$$\beta^* = (I_{n_h} + H^T C H + \lambda H^T L H)^{-1} H^T C X$$
Else Compute the output weights β with the following equation $\beta^* = H^T (I_N + C H H^T + \lambda L H H^T)^{-1} C X$
Step 3: $X_{new} = X \beta^T$
Step 4: Treat each row of X_{new} as a point and cluster the N points into K clusters using the k-means algorithm.
The overall function of the OSML-GRELMA is presented in the following algorithm 2.

Algorithm 2. OSML-GRELMA Algorithm for Classification

Input: A small initial training set $N = \{(x_i, t_i) | x_i \in R^p, t_i \in R^m, i=1, \dots, \tilde{N}\}$
The model depth: m ;
The number of hidden nodes in each GRELMA: $n_{h1}, n_{h2}, \dots, n_{hm}$
The new activation function: h_{new} .
Output: The classification results of the M data.
Phase 1 (BPh) (Huang et al., 2005; Liang et al., 2006)
Initialize $X_i = X_{train}$
For $i = 1: m$
Assign arbitrary input weight w_i and bias b_i , $i=1$, of the i_{th} layer GRELMA by using some random numbers;
Calculate the initial hidden layer output matrix $H_0 = [h_1, \dots, h_{\tilde{N}}]^T$, where $h_i = [h_{new}(w_1 \cdot x_i + b_1), \dots, h_{new}(w_{\tilde{N}} \cdot x_i + b_{\tilde{N}})]^T$, $i = 1, \dots, \tilde{N}$, where h_{new} is the activation function.
Train the output weights β_i of the i_{th} layer GRELMA;
Estimate the initial output weight $\beta^{(0)} = M_0 H_0^T T_0$, where $M_0 = (H_0^T H_0)^{-1}$ and $t_1, \dots, t_{\tilde{N}}$.
Set $k = 0$.
Compute the outputs $X_{i+1} = h_{new}(X_i \beta_i^T)$.
Phase 2 (SLPh) (Huang et al., 2005; Liang et al., 2006)
The essential step of this phase for each further coming observation (x_i, t_i) , where $x_i \in R^p$, $t_i \in R^m$ and $i = \tilde{N} + 1, \tilde{N} + 2, \tilde{N} + 3$, is described as follows:
Calculate the hidden layer output vector $h_{(k+1)} = [h_{new}(w_1 \cdot x_i + b_1), \dots, h_{new}(w_{\tilde{N}} \cdot x_i + b_{\tilde{N}})]^T$
Calculate the latest output weight $\beta^{(k+1)}$ by using the algorithm $\hat{\beta} = (H^T H)^{-1} H^T T$ which is known as the Recursive Least-Squares (RLS) algorithm.
Set $k = k + 1$
End For
Map X_{m+1} , the output of the m_{th} layer, to the output layer.
Compute the classification results by using the above trained OSML-GRELMA model.

The main objective and training success of the proposed OSML-GRELMA approach is based on the evolutionary identification of the underlying structure of the input data flows to produce the final model. It basically uses the knowledge of labelled data to investigate the distribution of the input data, aiming at enhancing the outcome of the learning process using an adaptive scheme. In this sense, it includes procedures that approach unsupervised learning, where inputs come from the same marginal distribution or they follow a common cluster structure.

3.4. The innovative proposed geolocation country-based services algorithm

The proposed GCBS algorithm performs the precise determination of the country to which coordinates (obtained by a GPS device) are referenced each time. It has been introduced by (Demertzis et al., 2017). According to this method, the world borders were originally taken from the states, as they appear in the shapefile available at the following URL http://thematicmapping.org/downloads/world_borders.php.

The following example is a Python script included in this shapefile and checks the coordinates 39.35230, 24.41232 belonging to Greece¹:

```
import countries
cc = countries.CountryChecker(TM_WORLD_BORDERS-0.3.shp')
print cc.getCountry(countries.Point(39.35230, 24.41232)).iso
print Greece
```

Experiments were performed on a PC with an i7 at 3.6 GHz CPU and 16 GB RAM. The feature extraction and classification processes run under a Linux Ubuntu 16.04 LTS Operating System (OS), with PyLab (NumPy, SciPy, Matplotlib and IPython).

4. Data set used in this research

Bioacoustics is an interdisciplinary branch that combines biology with acoustics to study the production of sound, its propagation through elastic means, and its intake from animals. It deals with objective electrophysiological measurements performed on animals for the study of the hearing instrument, with emphasis on the bioelectric potentials. The *Sea Audio Dataset* used in this study was created for the extension of a previous research effort of our team (Demertzis et al., 2017). More details are presented in the next paragraphs.

4.1. Underwater sounds

Four main categories of sounds have been determined in order to create highly complex scenarios. They can potentially include the most likely cases to be detected underwater. They are suitable for the training phase of the proposed framework. [Graphs 1–3](#) are used to depict the three of them below.

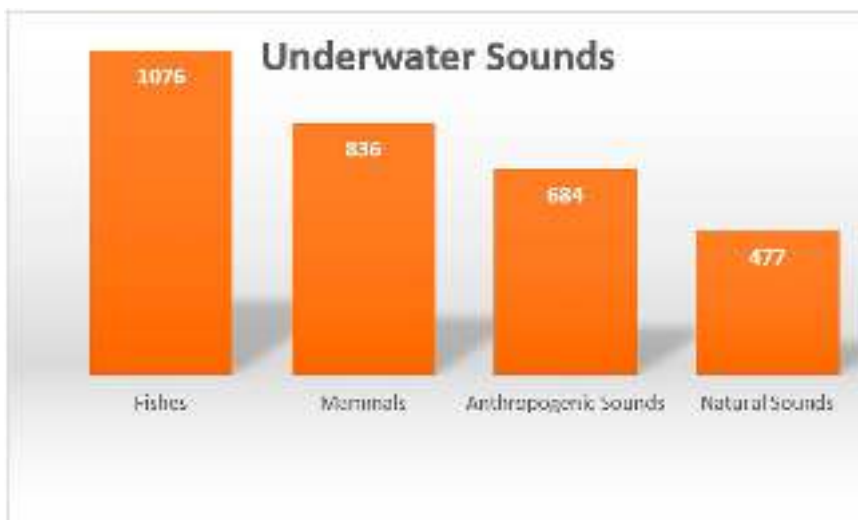
The first graph presents the four categories of underwater sounds.

The categories of the underwater sounds are presented below:

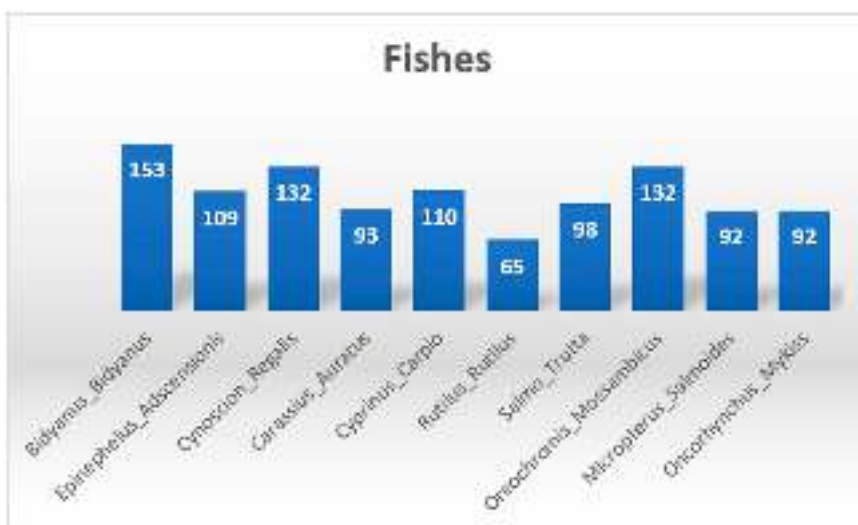
- Fish: Several species of fish produce sounds with various mechanisms such as teeth, pharynx, fins, and shuttle bladder. 1076 sounds belonging to 10 fish species have been included in this category (e.g. *Bidyanus Bidyanus*, *Epinephelus Adscensionis*, *Cynoscion Regalis*, *Carassius Auratus*, *Cyprinus Carpio*, *Rutilus Rutilus*, *Salmo Trutta*, *Oreochromis Mossambicus*, *Micropterus Salmoides*, *Oncorhynchus Mykiss*).

The second graph shows the categorization of 1076 sounds belonging to 10 fish species.

- Mammals: Marine mammals produce and use sounds to orientate and communicate with each other. Totally 836 sounds belonging to 8 species of mammals are included



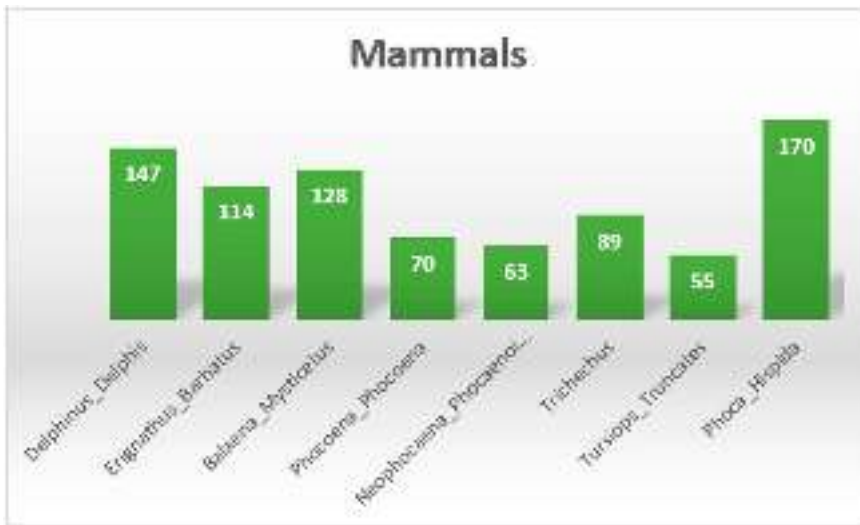
Graph 1. The four categories of underwater sounds.



Graph 2. The categorization of 1076 sounds belonging to 10 fish species.

in this category (e.g. *Delphinus Delphis*, *Erignathus Barbatus*, *Balaena Mysticetus*, *Phocoena Phocoena*, *Neophocaena Phocaenoides*, *Trichechus*, *Tursiops Truncates*, *Phoca Hispida*).

The third graph indicates the categorization of 836 sounds belonging to 8 species of mammals.



Graph 3. The categorization of 836 sounds belonging to 8 species of mammals.

- Anthropogenic Sounds: It comprises of 684 sounds belonging to nine classes (Ship, Sonar, Zodiac, Torpedo, Wind Turbine, Scuba Noise, Bubble Curtain, Personal Water Craft, Airgun).
- Natural Sounds: It includes totally 477, which are classified in six clusters (*Earthquake, Hydrothermal Vents, Ice Cracking, Rainfall, Lightning, Waves*).

4.2. Feature extraction

The Feature Extraction process (Giannakopoulos, 2015) enables capturing the characteristics that precisely determine the uniqueness of each sound and helps to distinguish between acoustic categories. The categories distinction is based on 34 characteristics related to statistical measurements obtained from the signal frequency information, while in general two stages are followed in the characteristics extraction methodology (Giannakopoulos, 2015):

- Short-term feature extraction: This method separates the input signal into short-term windows (frames) and calculates a series of attributes for each frame, thus leading and discovering the sequence of short-term vector characteristics for the signal.
- Mid-term feature extraction: In many cases, the signal is represented by the relative statistics of extracted features of the Short-term feature extraction operation described above. For this reason, the Mid-term feature extraction function extracts a series of statistics (e.g. mean and standard deviation) over each short-term feature sequence.

In this research effort, we have extracted the short-term feature sequences for an audio signal, using a frame size of 50 msec and a frame step of 25 msec (50% overlapping). All sounds have a sampling rate of 44.1 kHz, 16-bit stereo resolution while their average duration is 10.3 sec.

5. The marine species identification framework

The first step of the proposed algorithmic approach includes the process of *audio feature extraction*, obtaining the proper features related to each sound of the *Sea Audio Dataset*. In the second step, these attributes are introduced to the proposed DELMF model and the classification process is performed to determine if the detected sound comes from an IAS.

If this sound is described as noise that comes from a usually human sea-related process, then it is rejected and there is no further development in the process.

If this sound comes from a species of fish, or mammal, and once this species is identified, the coordinates are taken from a GPS device and they are assigned to the country they belong to, through the GCBS. Then a check is made as to whether the species identified is native to that country, otherwise it is recorded as IAS.

Lists with indigenous and invasive species were extracted from the Invasive Species Compendium² the most valid and comprehensive database world-wide. Algorithm 3 has been developed in an earlier research of our team in order to classify the IAS as native or invasive species (Demertzis et al., 2017). It is presented below:

Algorithm 3. Algorithm for classifying IAS as native or invasive species

Input:
 Recognized_Species;
 Country;
 Country_Native_Species;
 1: Read Recognized_Species, Country, Country_Native_Species;
 2: **for** i=1 to Country_Native_Species [max] **do**
 3: **if** Country_Native_Species [i]=Recognized_Species **then**
 4: Recognized Species=Native_Species
 5: **else**
 6: Recognized Species=Invasive_Species
 7: **end if**
 8: **end**
Output:
 Species Identity;

The overall algorithmic procedure is presented in [Figure 1](#).

6. Results and comparative analysis

The evaluation of the classification is mainly based on the Confusion Matrix (CM), where the main diagonal values (top left corner to bottom right) correspond to the correct classifications and the rest of the numbers correspond to very few cases that were misclassified. The misclassifications frequencies are related to the False Positive (FP) and False Negative (FN) indices appearing in the confusion Matrix. An FP is the number of cases where you wrongfully receive a positive result and the FN is exactly the opposite. On the other hand, the True Positive (TP) is the number of records where you correctly receive a Positive result. The True Negative (TN) is defined respectively. The True Positive rate (TPR) known as Sensitivity, the True Negative rate also known as Specificity (TNR) and the Total Accuracy (TA) are defined by using equations 10, 11, and 12 respectively

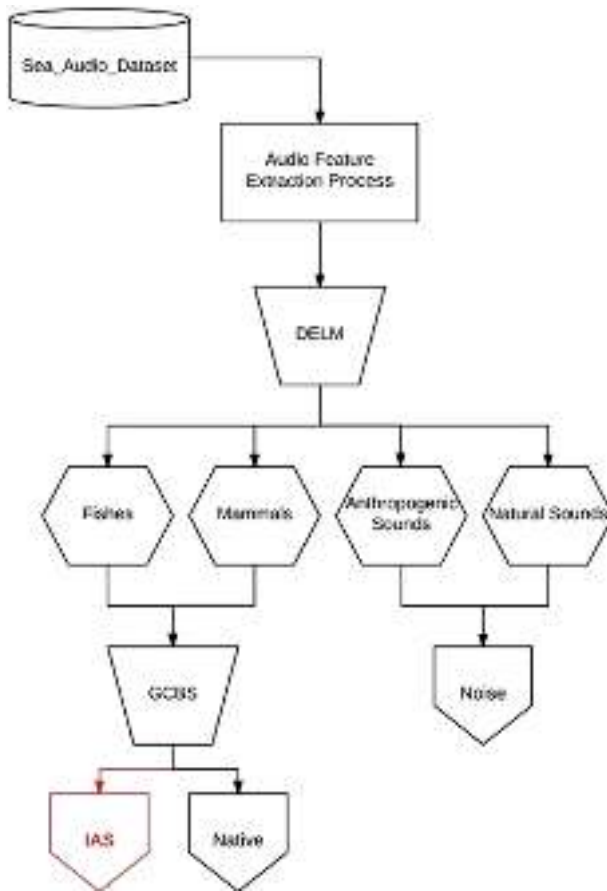


Figure 1. The flowchart of the proposed algorithmic procedure.

(Fawcett, 2006; Mao, Jain, & Duin, 2000):

$$TPR = \frac{TP}{TP + FN} \tag{10}$$

$$TNR = \frac{TN}{TN + FP} \tag{11}$$

$$TA = \frac{TP + TN}{P + N} \tag{12}$$

P represents the number of true and false positives while N represents the number of true and false negatives respectively (total testing class), in total accuracy.

The Precision (PRE) the Recall (REC) and the F-Score indices are defined as in equations 13, 14 and 15, respectively (Fawcett, 2006; Mao et al., 2000):

$$PRE = \frac{TP}{TP + FP} \tag{13}$$

$$\text{REC} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (14)$$

$$\text{F - Score} = 2 * \frac{\text{PRE} * \text{REC}}{\text{PRE} + \text{REC}} \quad (15)$$

The 10-fold cross validation (10_FCV) is employed in this stage in order to obtain performance indices. Analytical values of the predictive capacity of the DELMF algorithm are presented in the following Tables 1–3 where a comparison is made with the Spiking Convolutional Neural Network (SCNN) algorithm, which was the subject of an earlier research effort by our team (Demertzis et al., 2017). The Receiver Operating Characteristic (ROC) curve, that is a graphical plot that illustrates the diagnostic ability of the classifier system are presented in graphs 4–6. Authors have designed 3 ROC curves. The first ROC curve (Graph 4) represents the underwater sounds, the second curve (Graph 5) indicates the fishes and finally the third curve (Graph 6) depicts the mammals. In all simulations below, the testing hardware and software characteristics are listed as follows: *Laptop Intel-i7 2.4G CPU, 16G DDR3 RAM, Windows 10, MATLAB R2015a*.

In general, the proposed system has achieved equal and in many cases higher accuracy than the comparable alternatives. The most important fact to be highlighted in this respect, which justifies this research, is the significant reduction in DELMF's implementation time, which was reduced up to 23% in any case, even when compared with the extremely fast SCNN algorithm. This achievement is very important for the scientific research in Machine Hearing, which, as has been said, is an extremely complex field with specific requirements.

The performance results in Tables 1–3 are related to the testing process. The validity of the proposed approach has been proven by using all known indices namely: Accuracy,

Table 1. Performance metrics of categories_dataset.

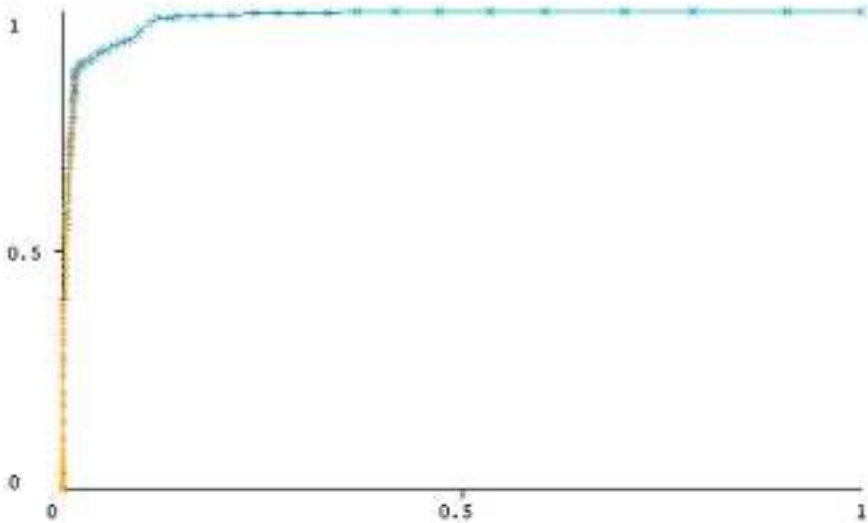
Classifier	Classification Accuracy (ACC) & Performance Metrics in the Test_Dataset						Time
	ACC	RMSE	Precision	Recall	F-Score	ROC	
DELMF	96.08%	0.1376	0.960	0.960	0.960	0.970	189 sec
SCNN	96.25%	0.1368	0.963	0.963	0.963	0.975	248 sec

Table 2. Performance metrics of mammals_dataset.

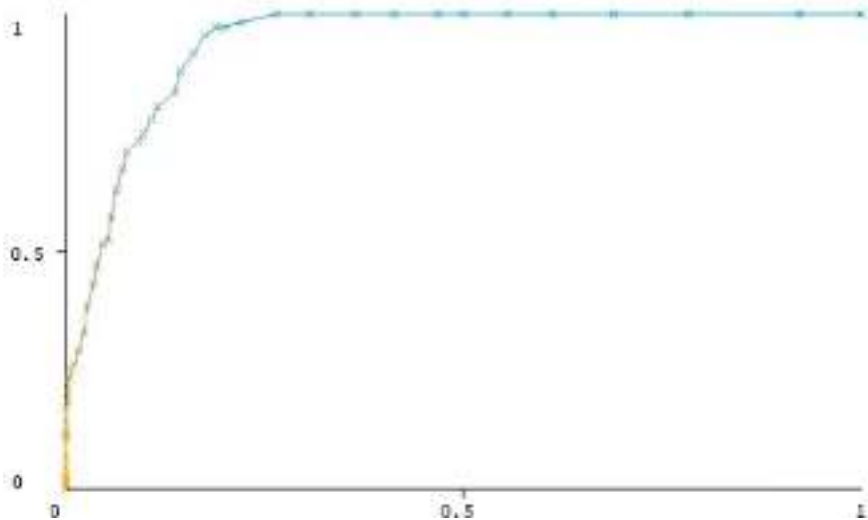
Classifier	Classification Accuracy (ACC) & Performance Metrics in the Test_Dataset						Time
	ACC	RMSE	Precision	Recall	F-Score	ROC	
DELMF	92.18%	0.1571	0.922	0.922	0.922	0.955	186 sec
SCNN	92.10%	0.1577	0.921	0.921	0.921	0.952	259 sec

Table 3. Performance metrics of fishes_dataset.

Classifier	Classification Accuracy (ACC) & Performance Metrics in the Test_Dataset						Time
	ACC	RMSE	Precision	Recall	F-Score	ROC	
DELMF	87.91%	0.1512	0.879	0.879	0.879	0.920	214 sec
SCNN	89.03%	0.1481	0.892	0.890	0.890	0.939	277 sec



Graph 4. The ROC curve of the underwater sounds.



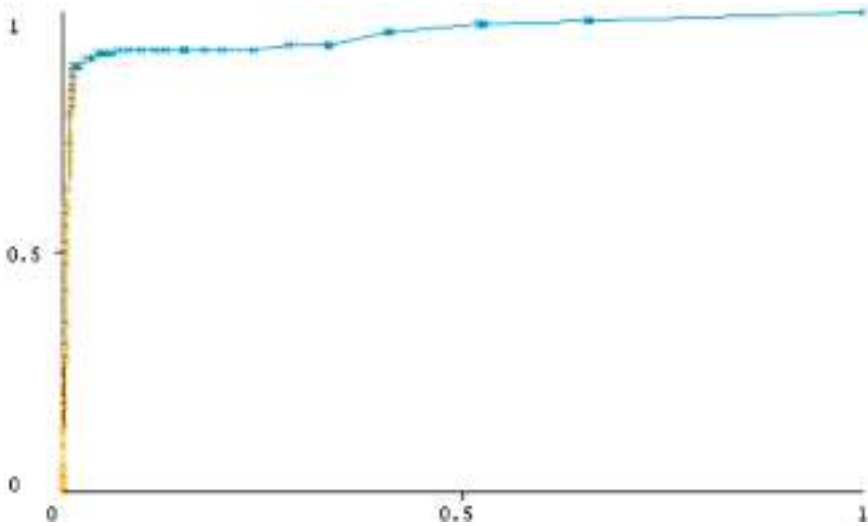
Graph 5. The ROC curve of the fishes.

RMSE, PRECISION, RECALL, F-Score and ROC. So the results are very well supporting the reliable performance.

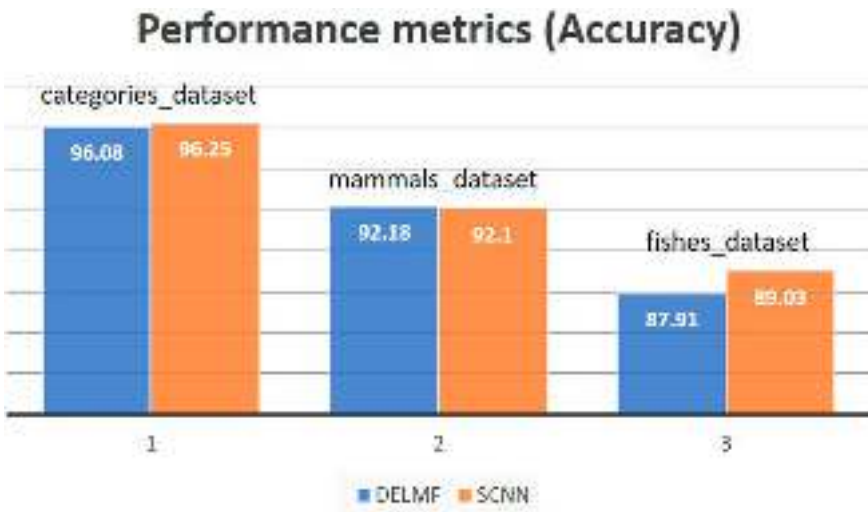
Moreover, a graph represents the accuracy for the data sets that have been considered in the testing process ([Graph 7](#)).

7. Discussion and conclusions

An innovative, reliable, low-demand and effective computer-based hearing system, based on sophisticated computational intelligence, was presented in this work. The described



Graph 6. The ROC curve of the mammals.



Graph 7. The accuracy for the datasets that have been considered in the testing process.

application significantly reduces the running time. It is coupled with the highly optimistic obtained results. It is a credible innovative proposal in the standardization and design of biosecurity and biodiversity protection methods.

As it has been proven, ELMs are an important approach for handling and analysing Big Data as they require the minimum training time relative to the corresponding engineering learning algorithms. Moreover ELMs do not require fine manipulations to determine their operating parameters and finally they can determine the appropriate output weights towards the most effective resolution of a problem. What is most important, they have the potential to generalize, in contrast to corresponding methods which adjust their

performance based solely on their training data set. It is obvious that the emerging use of ELM in Big Data analysis as well as DELE creates serious prerequisites for complex systems' development by low-cost machines.

8. Future research

Proposals for development and future improvements of this system should focus on further optimizing the parameters of the DELMF algorithm used to achieve an even more efficient, accurate, and faster categorization process. It would be very important to select a heuristic optimization method to search for the optimal system parameters and possibly to automatically calculate the contribution of the independent variables and assign them as entry variables to the system (Demertzis & Iliadis, 2017c). This would be a *Meta-learning* approach and it would lead to a totally automated and self-adaptive system.

Notes

1. <https://github.com/che0/countries>
2. <http://www.cabi.org/isc/>

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Konstantinos Demertzis is an officer of the Hellenic Army, serving in Research and Informatics Corps. He holds a PhD in Environmental Informatics and Computational Intelligence from the Democritus University of Thrace and an MSc in the Communication & Computer Networking Technologies, from the University of the Aegean. He is a Part-time Lecturer at the Computer and Informatics Engineering department of the Eastern Macedonia & Thrace Institute of Technology, Research Associate at the National and Kapodistrian University of Athens and Research Assistant at the University of Peloponnese. His research interests include Computational Intelligence (Big Data Analytics, Machine Learning), Environmental Informatics (Invasive Species, Climate Change) and Cyber Security (Critical Infrastructure Protection, Malware Analysis). He has published 10 research papers in international scientific journals and 14 in Proceedings of international conferences.

Lazaros S. Iliadis (BSc in Mathematics AUTh, MSc Computer Science, Wales UK, PhD Expert systems, AUTh) is Professor of Applied Informatics in the School of Engineering Department of Civil Engineering, Lab of Mathematics and Informatics of the Democritus University of Thrace. He has authored, co-authored 60 publications in international scientific journals, more than 100 publications in Proceedings of international conferences corresponding to more than 1000 citations and he has been the organizer/chair of more than 18 scientific conferences. He is author and co-author of 2 scientific books and a member of ACM and IEEE. He has supervised 3 PhD dissertations and 24 MSc ones. He has lectured as a visiting Professor in many Greek and foreign Universities (UOM, UTh, UOL, UEL, CRAN, COV).

Vardis-Dimitris Anezakis is a PhD candidate (Forest Informatics) in the Department of Forestry and Management of the Environment and Natural Resources of the Democritus University of Thrace, Greece. His research interests include Hybrid Computational Intelligence modelling of environmental risks and threats. More specifically his PhD research is related to intelligent modelling of

the impacts of atmospheric conditions-pollution on public health, considering the climate change contribution. He has published 7 research papers in international scientific journals and 7 in Proceedings of international conferences.

ORCID

Vardis-Dimitris Anezakis  <http://orcid.org/0000-0002-4012-0651>

References

- Alom, M. Z., Alam, M., Taha, T. M., & Iftekharruddin, K. M. (2017, May). Object recognition using cellular simultaneous recurrent networks and convolutional neural network. *Proceedings of the International Joint Conference on Neural Networks (IJCNN2017)* (pp. 2873–2880). Anchorage. doi:10.1109/IJCNN.2017.7966211
- Cambria, E., & Guang-Bin, H. (2013, November–December). Extreme learning machines. *IEEE Intelligent Systems*, 28(6), 30–59. doi:10.1109/MIS.2013.140
- Demertzis, K., & Iliadis, L. (2015, September). Intelligent Bio-inspired detection of food borne pathogen by DNA barcodes: The case of invasive fish species ligocephalus scleratus. In L. Iliadis & C. Jayne (Eds.), *Engineering applications of neural networks EANN2015. Communications in computer and information science, Vol. 517* (pp. 89–99). Rhodes: Springer Verlag, Cham. doi:10.1007/978-3-319-23983-5_9
- Demertzis, K., & Iliadis, L. (2017a). The impact of climate change on biodiversity: The ecological consequences of introduced species in Greece. *Handbook of Climate Change Communication, Vol. 1, Theory of Climate Change Communication, Chapter 2*, In: W. Leal Filho, E. Manolas, A.M. Azul, U.M. Azeiteiro, & H. McGhie (Eds.), Springer. doi:10.1007/978-3-319-69838-0_2
- Demertzis, K., & Iliadis, L. (2017b, June). Detecting invasive species with a bio-inspired semi-supervised neurocomputing approach: The case of ligocephalus scleratus. *Neural Computing and Applications*, 28(6), 1225–1234. London: Springer. doi:10.1007/s00521-016-2591-2
- Demertzis, K., & Iliadis, L. (2017c, October). Adaptive elitist differential evolution extreme learning machines on Big Data: Intelligent recognition of invasive species. In P. Angelov, Y. Manolopoulos Y, L. Iliadis, A. Roy, & M. Vellasco (Eds.), *INNS 2016. Advances in intelligent systems and computing, Vol.529* (pp. 333–345). Thessaloniki: Springer Verlag, Cham. doi:10.1007/978-3-319-47898-2_34
- Demertzis, K., Iliadis, L., & Anezakis, V. D. (2017, July). A deep spiking machine-hearing system for the case of invasive fish species. *Proceedings of 2017 IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA)* (pp. 23–28). Gdynia, Poland. IEEE. doi:10.1109/INISTA.2017.8001126
- Deng, L., & Yu, D. (2013). Deep learning: Methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4), 197–387. doi:10.1561/20000000039
- Fawcett, T. (2006, June). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. Elsevier Science Inc. doi:10.1016/j.patrec.2005.10.010
- Giannakopoulos, T. (2015). Pyaudioanalysis: An open-source python library for audio signal analysis. *PLoS ONE*, 10(12), e0144610. doi:10.1371/journal.pone.0144610
- Hinton, G., Deng, L., Yu, D., Dahl, G., Rahman Mohamed, A., Jaitly, N., ... Kingsbury, B. (2012, November). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6), 82–97. doi:10.1109/MSP.2012.2205597
- Huang, G.-B. (2014). An insight into extreme learning machines: Random neurons, random features and kernels. *Cognitive Computation*, 6(3), 376–390. Springer New York LLC. doi:10.1007/s12559-014-9255-2
- Huang, G.-B. (2015). What are extreme learning machines? Filling the Gap between Frank Rosenblatt's dream and John von Neumann's Puzzle. *Cognitive Computation*, 7(3), 263–278. Springer New York LLC. doi:10.1007/s12559-015-9333-0

- Huang, G.-B., Liang, N.-Y., Rong, H.-J., Saratchandran, P., & Sundararajan, N. (2005, July). On-line sequential extreme learning machine. In M. H. Hamza (Eds.), *CI 2005. Proceedings of the IASTED International Conference on Computational Intelligence* (pp. 232–237). Calgary, Alberta, Canada.
- Krizhevsky, A, Sutskever, I., & Hinton, G. E. (2012, December). ImageNet Classification with Deep Convolutional Neural Networks. *NIPS 2012. Advances in Neural Information Processing Systems Vol.2* (pp. 1097–1105). Lake Tahoe, NV.
- Liang, N.-Y., Huang, G.-B., Saratchandran, P., & Sundararajan, N. (2006). A fast and accurate On-line sequential learning algorithm for feedforward networks. *IEEE Transactions on Neural Networks*, 17(6), 1411–1423. doi:10.1109/TNN.2006.880583
- Lyon, R. (2017, April). *Human and machine hearing: Extracting meaning from sound*. Cambridge: Cambridge University Press. doi:10.1017/9781139051699
- Mao, J., Jain, A. K., & Duin, P. W. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4–37. doi:10.1109/34.824819
- Ramírez-Gallego, S., Fernández, A., García, S., Chen, M., & Herrera, F. (2018). Big data: Tutorial and guidelines on information and process fusion for analytics algorithms with MapReduce. *Information Fusion*, 42, 51–61. doi:10.1016/j.inffus.2017.10.001
- Sun, K., Zhang, J., Zhang, C., & Hu, J. (2017). Generalized extreme learning machine autoencoder and a new deep neural network. *Neurocomputing*, 230, 374–381. ISSN 0925-2312, doi:10.1016/j.neucom.2016.12.027.
- Zhang, D., Lee, D.-J., Zhang, M., Tippetts, B. J., & Lillywhite, K. D. (2016). Object recognition algorithm for the automatic identification and removal of invasive fish. *Biosystems Engineering*, 145, 65–75. doi:10.1016/j.biosystemseng
- Zhang, J., Yin, J., Zhang, Q., Shi, J., & Li, Y. (2017). Robust sound event classification with bilinear multi-column ELM-AE and two-stage ensemble learning. *Eurasip Journal on Audio, Speech, and Music Processing*, (1). art. No. 11. doi:10.1186/s13636-017-0109-1
- Zhao, Z., Zhang, S.-H., Xu, Z.-Y., Bellisario, K., Dai, N.-H., Omrani, H., & Pijanowski, B. C. (2017). Automated bird acoustic event detection and robust species classification. *Ecological Informatics*, 39, 99–108. doi:10.1016/j.ecoinf.2017.04.003