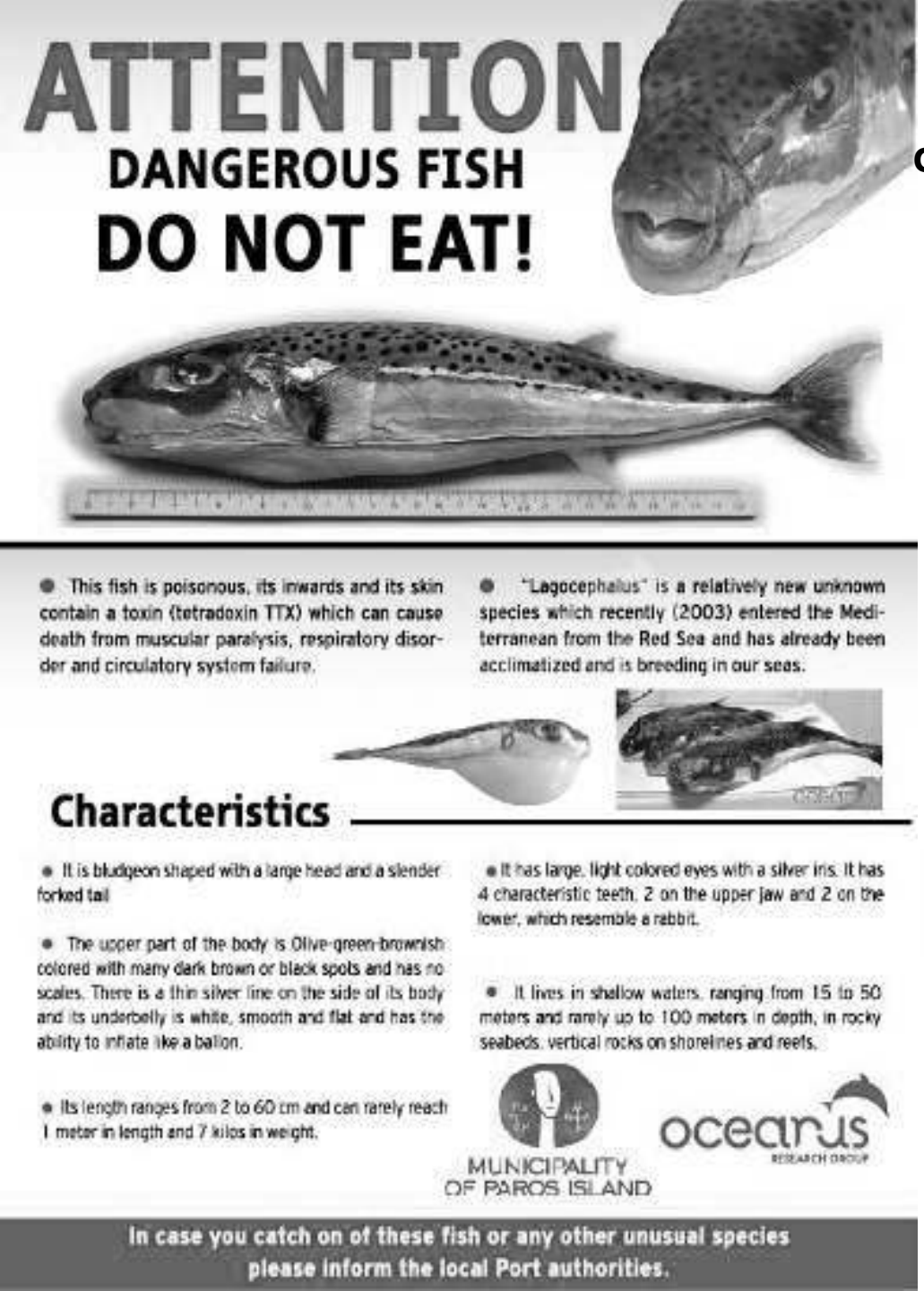



# Adaptive Elitist Differential Evolution Extreme Learning Machines on Big Data: Intelligent Recognition of Invasive Species


**Democritus University of Thrace**
Dep. of Forestry & Management of the Environment & Natural Resources
**Lab of Forest-Environmental Informatics & Computational Intelligence**

**Konstantinos Demertzis – Lazaros Iliadis**

# INNS Conference on Big Data

Thessaloniki, Greece,
23th-25th October 2016
Grand Hotel Palace

INTERNATIONAL NEURAL NETWORK SOCIETY

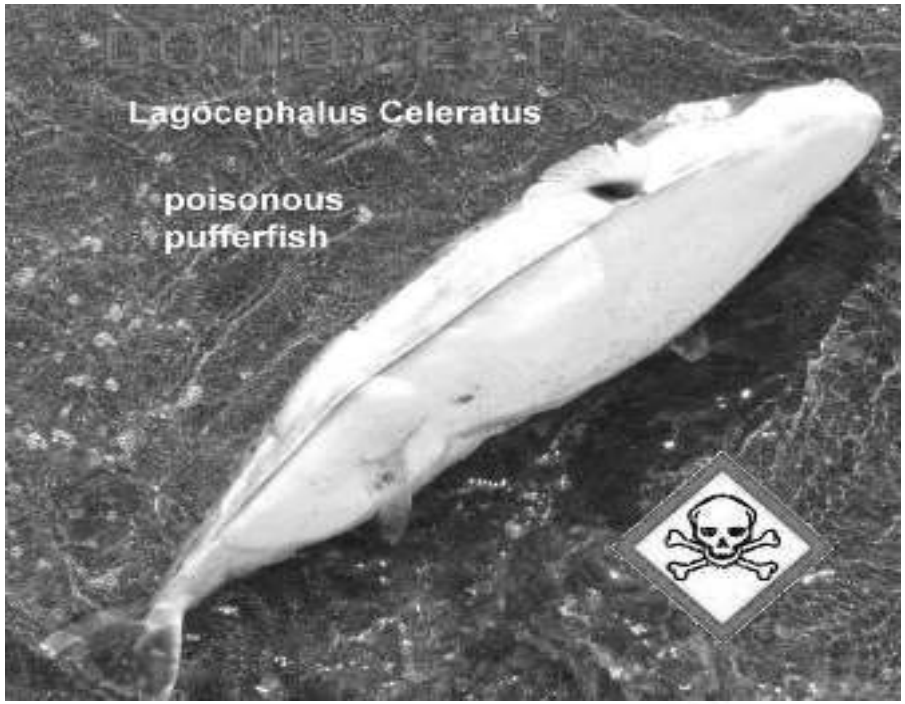
## Adaptive Elitist Differential Evolution Extreme Learning Machines on Big Data: Intelligent Recognition of Invasive Species

1. Introduction
2. Proposed System
3. Areas of Study
4. Data
5. Methodologies
6. Conclusion – Future Work
7. Questions - Discussion

**Adaptive Elitist Differential Evolution Extreme Learning Machines on Big Data: Intelligent Recognition of Invasive Species**

# Invasive Species

- ✓ *Invasive alien species: a growing problem for*
  - ✓ *ecology*
  - ✓ *biodiversity and*
  - ✓ *human health*

# Machine Vision and Big Data

- ✓ *Machine vision is a scientific field of AI which algorithmically attempts to reproduce the sense of sight.*
- ✓ *This is a case of big data processing, as data extracted from images require large storage space and they should be solved by a deterministic polynomial time.*
- ✓ *They also require dynamic assigning of computing resources and coordination of complex data analysis procedures.*


# Proposed Method

- ✓ *This research proposes:*
  - *an ELM model which is optimized by the Adaptive Elitist Differential Evolution (AEDE) algorithm.*
  - *a Geo Location system to detect the invasive species by comparing with the local species of the region under research.*

# Related Literature

- ✓ *Svellingen et al., presented a computer vision-based system that reliably classifies different fish species based on length measurement and weight determination.*
- ✓ *Cabreira et al., proposed an ANN-based platform for fish species identification is presented that uses several statistical methods such as discriminate function analysis and principal component analysis.*
- ✓ *Rova et al., applied SVM algorithm to fish recognition and constructed a texture-based mechanism that distinguishes between the Striped Trumpeter and the Western Butterfish species.*
- ✓ *Lee et al., carried out shape analysis of fish and developed an algorithm for removing edge noise and redundant data point base on nine species with similar shape features.*

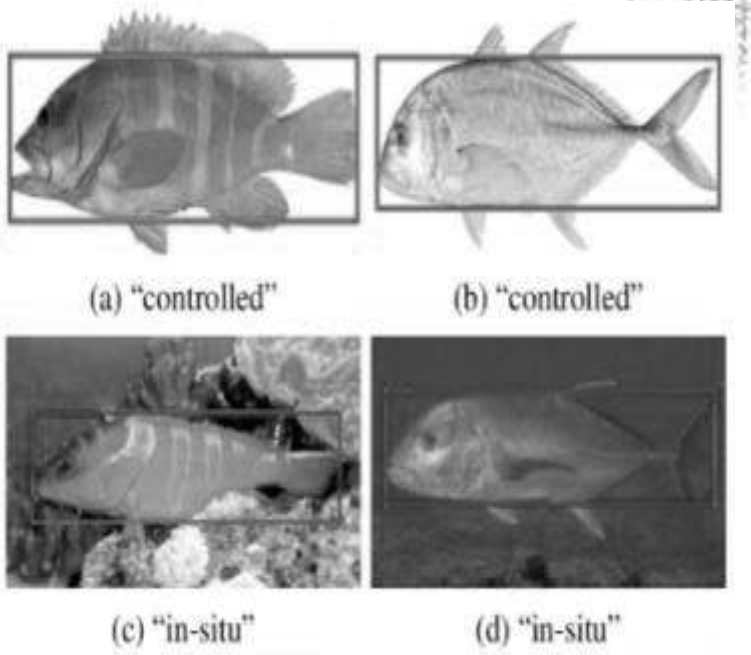
# Related Literature

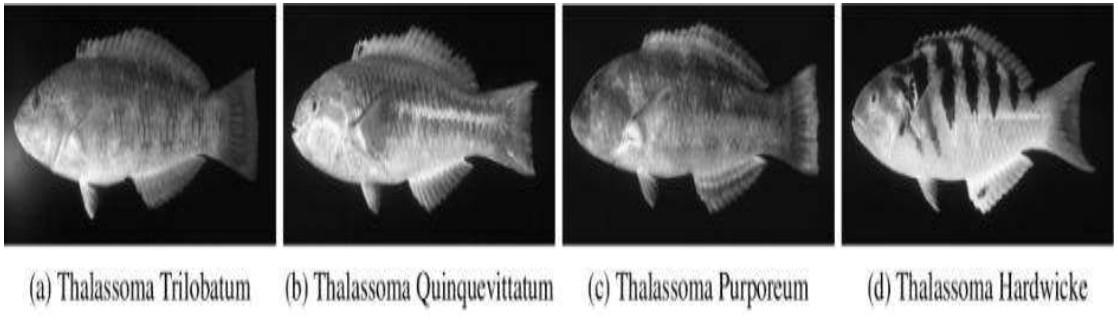
- ✓ *To find the near optimal network parameters of ELM, many scholars have combined with other intelligent optimization algorithms, such as:*
  - *Particle Swarm Optimization (PSO),*
  - *Differential Evolution (DE),*
  - *Cuckoo Search (CS),*
  - *Firefly Algorithm (FA),*
  - *Grey Wolf Optimizer (GWO),*
  - *Harmony Search (HS),*
  - *Biogeography-Based Optimization (BBO),*
  - *Animal Migration Optimization (AMO),*
  - *Gravitational Search Algorithm (GSA) and*
  - *Ant Colony Optimization (ACO).*

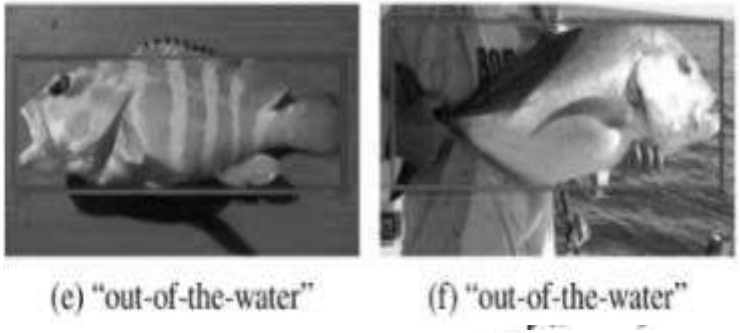
**Adaptive Elitist Differential Evolution Extreme Learning Machines on Big Data: Intelligent Recognition of Invasive Species**

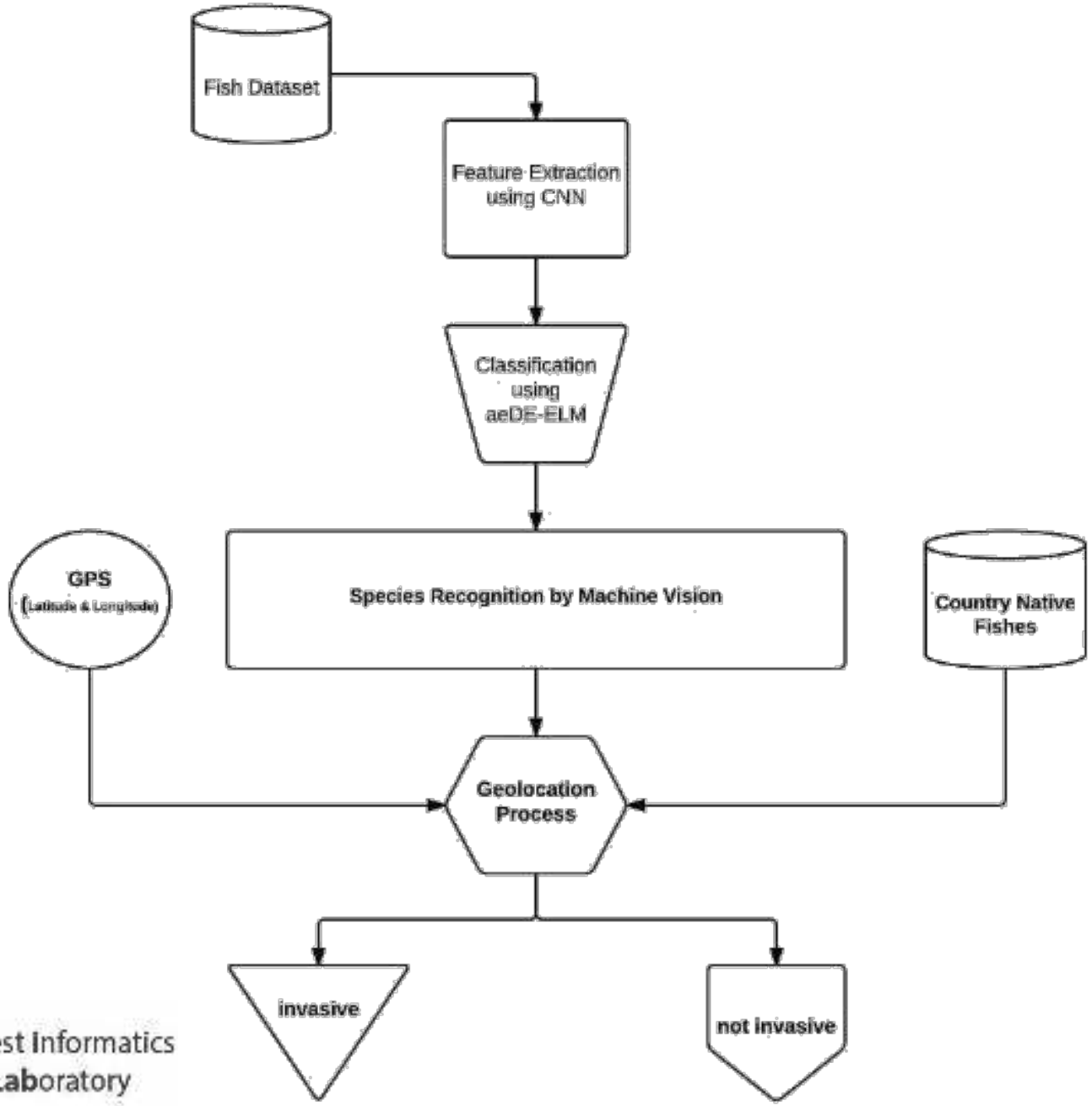
# Data

- ✓ *The fish image data set (3,961 images) is related to 469 species.*
- ✓ *This data consists of real-world images of fish captured in conditions defined as "controlled", "out-of-the-water" and "in-situ".*
- ✓ *Different species are visually similar,*
- ✓ *There is a high degree of variability in the image quality and environmental conditions.*

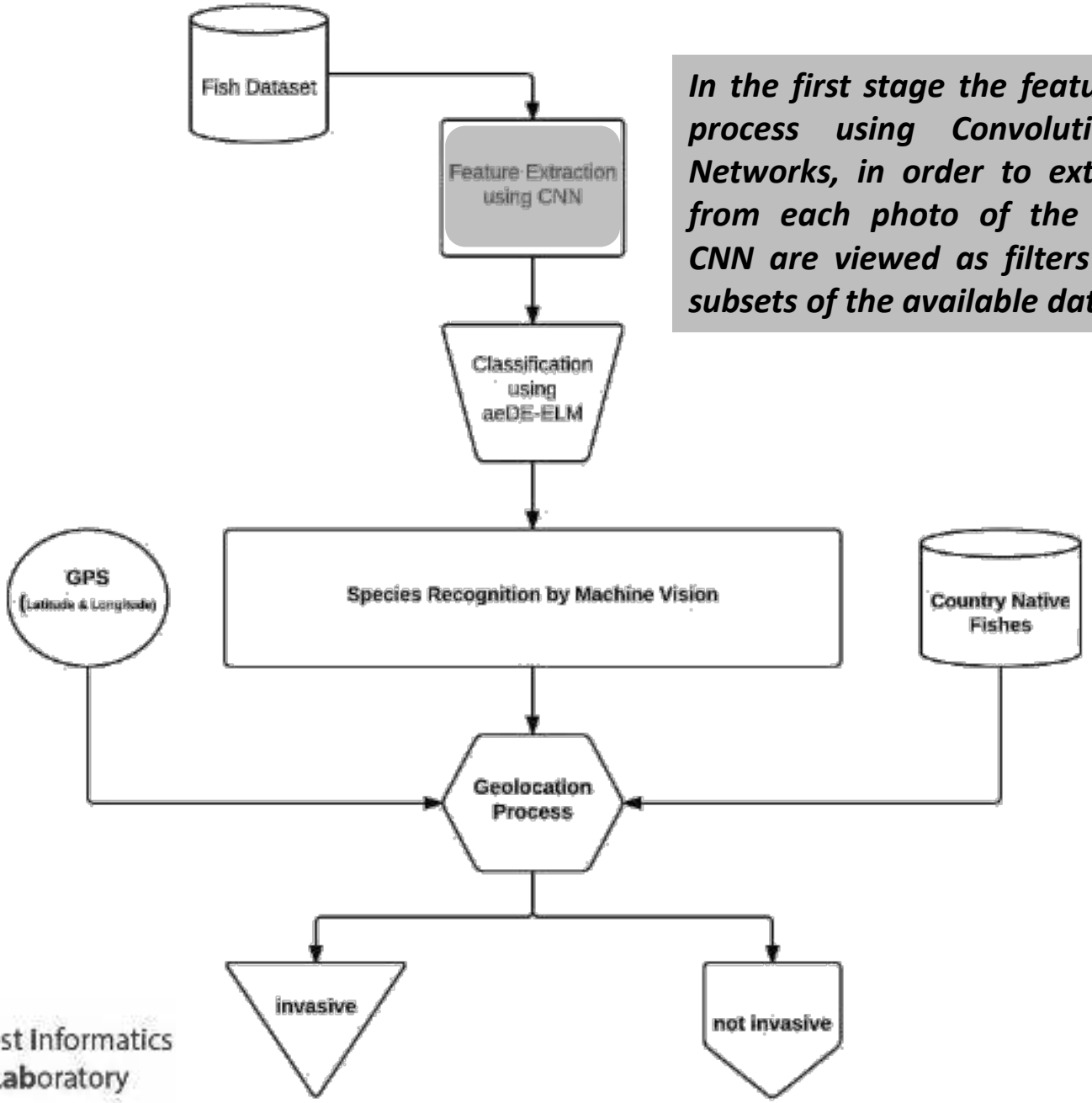





Forest Informatics Laboratory
FILab

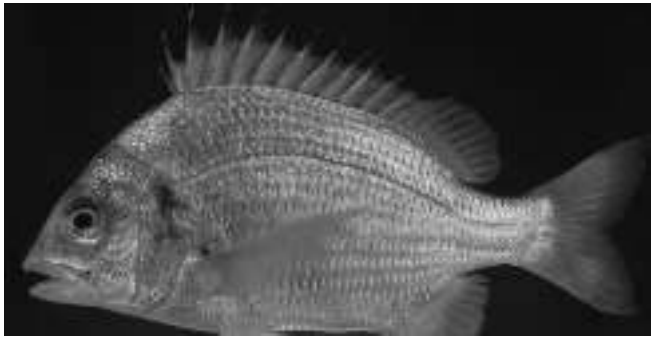



***In the first stage the feature extraction process using Convolutional Neural Networks, in order to extract features from each photo of the fish dataset. CNN are viewed as filters in particular subsets of the available data.***

Forest Informatics Laboratory
FILab

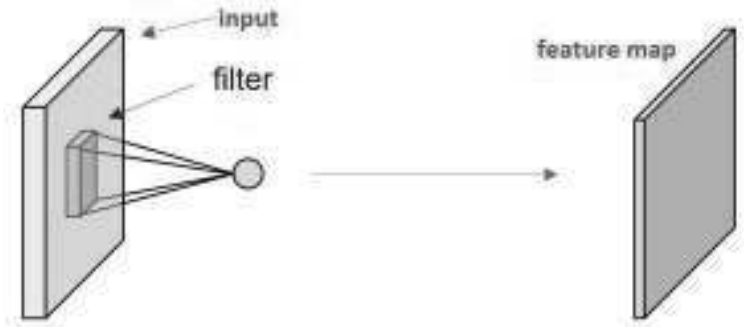
# Description of the proposed method

**81x42 pixel**



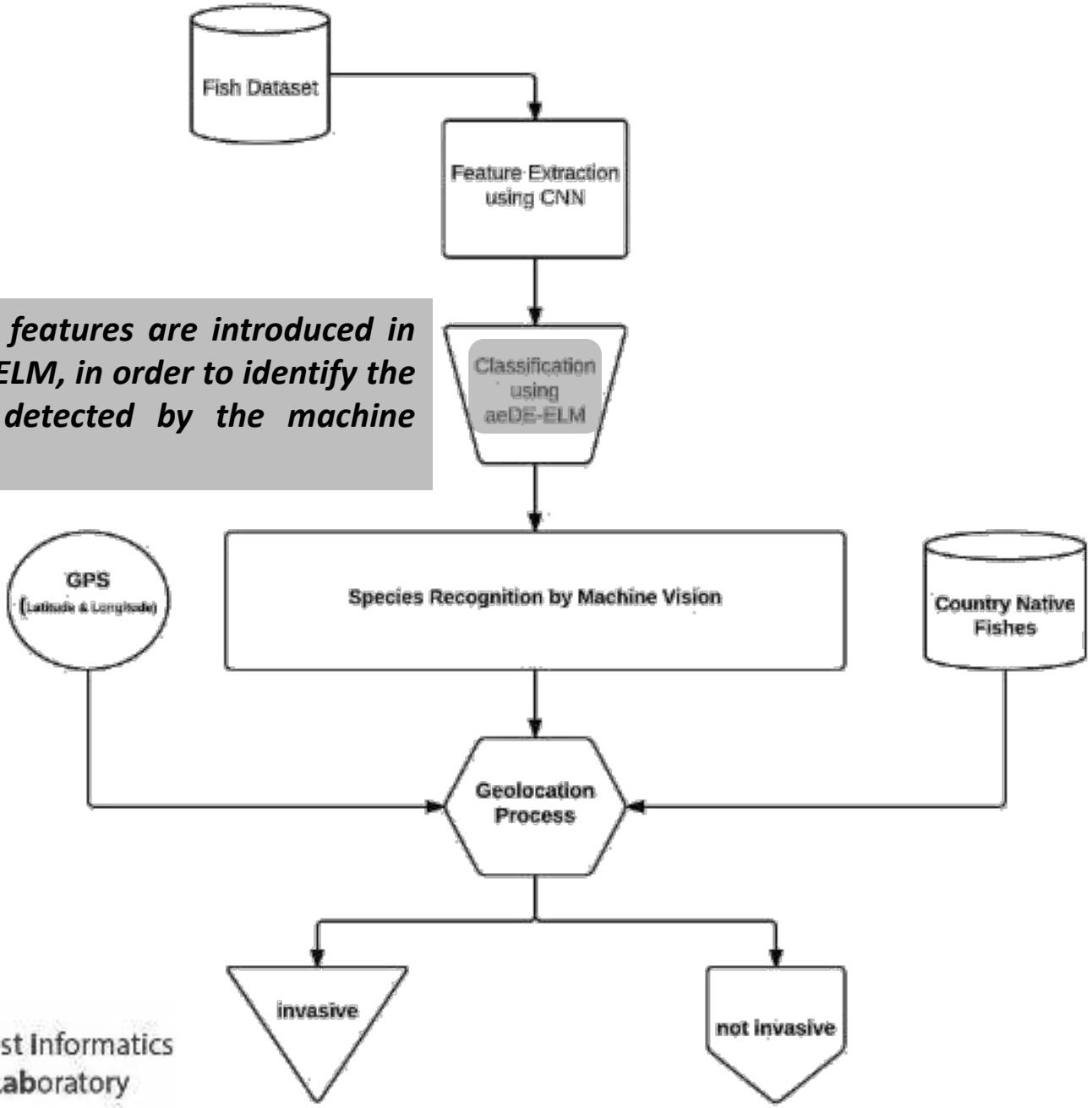
**96 feature map**

*The picture was divided in frames of 9x7 pixels (in order to avoid the frame of 1x1)*




- ✓ **The dimensions of the convolution layer were 81x42=3402 neurons for each one of the 96 filters were applied as feature map.**
- ✓ **We accepted that the neurons belonging to the same filter were assigned the same weights. This means that a filter with 3402 neurons can have only 190 weights (9x7x3+1=190 weights plus a bias *-the number 3 is used due to the RGB color model used that requires 1 channel for each color-*) instead of 3402x190=646,380.**
- ✓ **In this way the weights plus bias were reduced significantly to 96x(9x7x3+1)=18,240.**
- ✓ **This assumption is not arbitrary but it is based on the logical statement that the application of a filter can have useful results regardless the position in which it may be applied.**

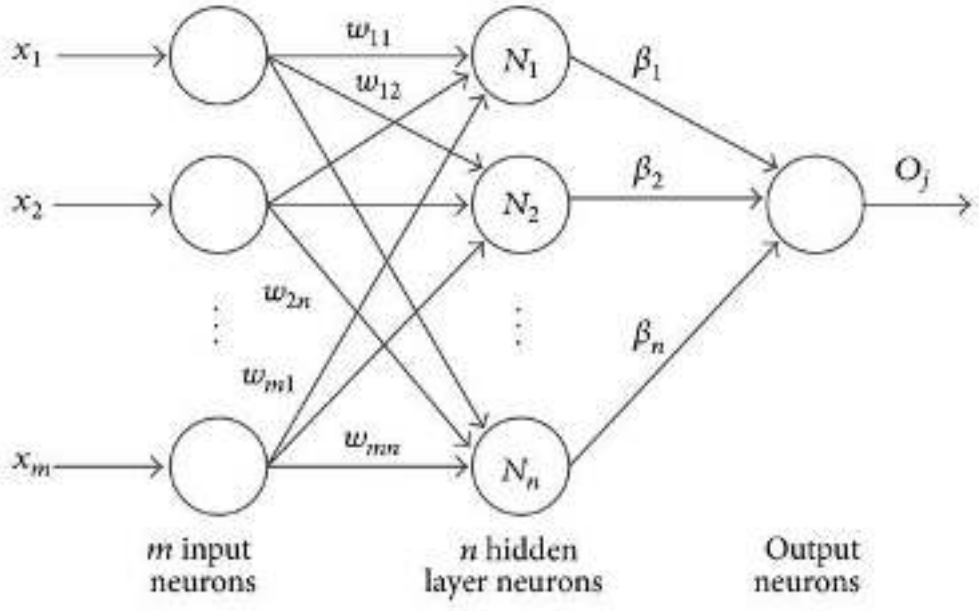
***The extracted features are introduced in the proposed ELM, in order to identify the type of fish detected by the machine vision system.***




Forest Informatics Laboratory
FILab

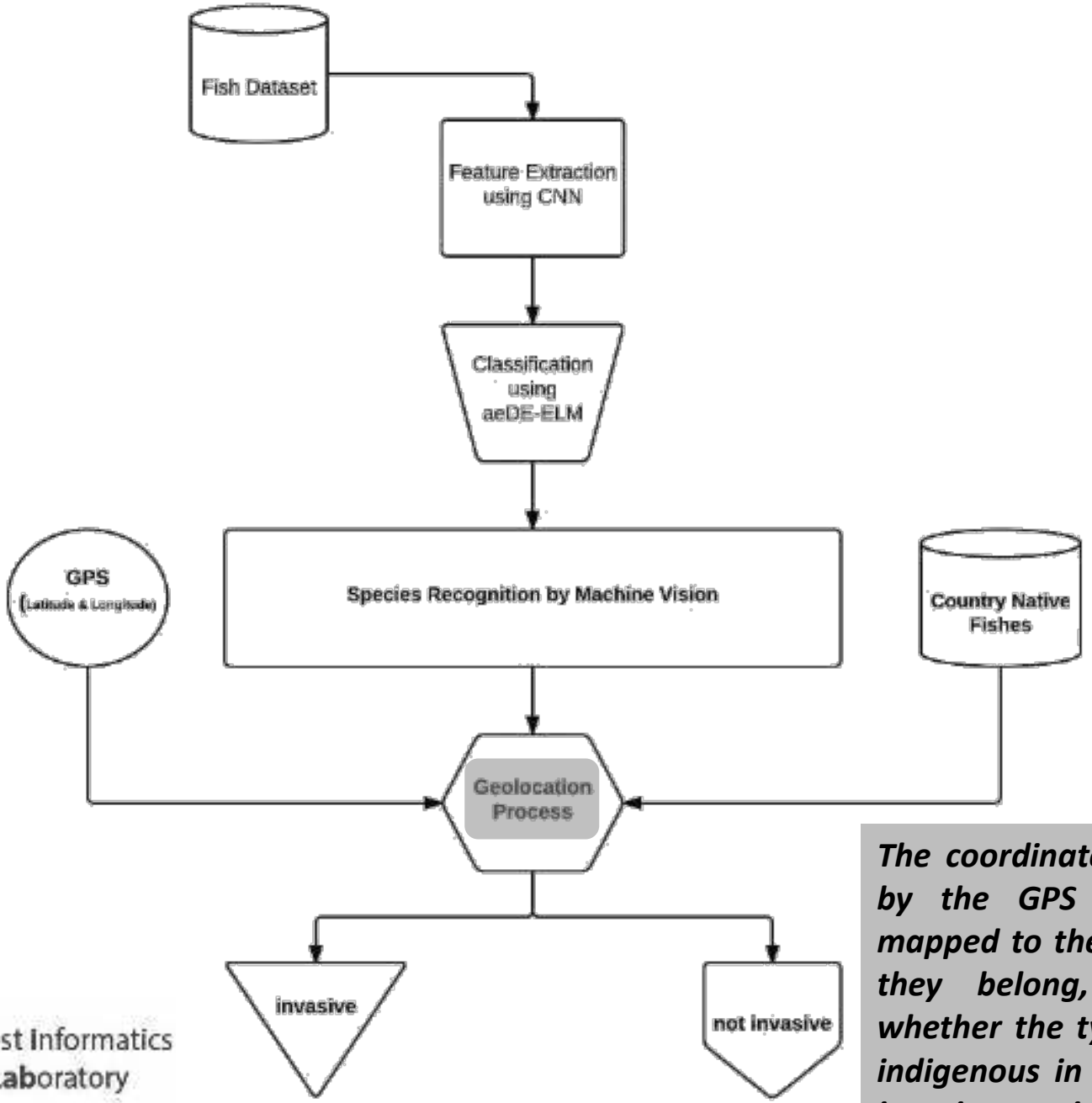
# Description of the proposed method

- ✓ *ELMs*
  - ✓ *are a kind of the Single-Hidden-Layer FFNN,*
  - ✓ *are characterized by the possibility to establish the parameters of hidden nodes at random, before they see the training data vectors,*
  - ✓ *are extremely fast and effective.*



# Description of the proposed method

- ✓ *The AEDE introduces two alternatives in order to enhance the search capability as well as the convergence speed of the DE algorithm:*
  - ✓ *The first one is applied in the **mutation phase**. The new adaptive mutation scheme of the DE uses two mutation operators:*
    - ✓ *The "**rand/1**" which aims to ensure diversity of the population and prohibits the population from getting stuck in a local optimum.*
    - ✓ *The "**current-to-best/1**" which aims to accelerate convergence speed of the population by guiding the population toward the best individual.*
  - ✓ *The second one in the **selection phase**.*
    - ✓ *The children population C consisting of trial vectors is combined with the parent population P of target vectors to create a combined population Q. Then, NP best individuals are chosen from the Q to construct the population for the next generation.*
    - ✓ *In this way, the best individuals of the whole population are always stored for the next generation.*
    - ✓ *This helps the algorithm to obtain a better convergence rate.*




Forest Informatics Laboratory
FILab

***The coordinates are obtained by the GPS and they are mapped to the country where they belong, to find out whether the type identified is indigenous in this country, or invasive species.***

# Description of the proposed method

✓ *Python script example:*

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

✓ *Geolocation Process:*

**Input**: Recognized_Fish; Country; Country_Native_Fishes;
1: Read Recognized_Fish, Country, Country_Native_Fishes;
2: **for** i=1 to Country_Native_Fishes [max]**do**
3: **if** Country_Native_Fishes[i]= Recognized_Fish **then**
4: Recognized Fish=! Recognized Fish;
5: Recognized Fish=Native_Fish
6: **else**
7: Recognized Fish=Invasive_Fish
8: **end if**
9: **end**
**Output:** Fish Identity;

# Results

| Classifier | Classification Accuracy (ACC) & Performance Metrics | | | | | | |
|---|---|---|---|---|---|---|---|
| | ACC | RMSE | Precision | Recall | F-Score | ROC | Evaluation |
| *AEDE-ELM* | *94.81%* | *0.1673* | *0.948%* | *0.948* | *0.949%* | *0.985* | *10-Fold CV* |
| DE-ELM | 93.97% | 0.1711 | 0.940% | 0.940 | 0.940% | 0.985 | 10-Fold CV |
| ELM | 93.13% | 0.1726 | 0.932% | 0.932 | 0.932% | 0.982 | 10-Fold CV |

- ✓ The **Precision** measure shows what percentage of positive predictions where correct.
- ✓ The **Recall** measures what percentage of positive events were correctly predicted.
- ✓ The **F-Score** can be interpreted as a weighted average of the precision and recall.
- ✓ The **ROC curve** is related in a natural way to cost/benefit analysis of diagnostic decision making.

# Results

| Classifier | Classification Accuracy (ACC) & Performance Metrics | | | | | | |
|---|---|---|---|---|---|---|---|
| | ACC | RMSE | Precision | Recall | F-Score | ROC | Evaluation |
| *AEDE-ELM* | *97.52%* | *0.1247* | *0.972%* | *0.972* | *0.972%* | *0.993* | *10-Fold CV* |
| DE-ELM | 97.01% | 0.1289 | 0.970% | 0.970 | 0.971% | 0.990 | 10-Fold CV |
| ELM | 95.86% | 0.1461 | 0.960% | 0.958 | 0.959% | 0.965 | 10-Fold CV |

- ✓ The **Precision** measure shows what percentage of positive predictions where correct.
- ✓ The **Recall** measures what percentage of positive events were correctly predicted.
- ✓ The **F-Score** can be interpreted as a weighted average of the precision and recall.
- ✓ The **ROC curve** is related in a natural way to cost/benefit analysis of diagnostic decision making.

# Conclusions

- ✓ *An innovative biologically inspired hybrid computational intelligence approach suitable for big data was presented in this research paper.*
- ✓ *It is a machine vision system which can recognize various fish species in order to rank them as to whether they are invasive in the region identified or not.*
- ✓ *Specifically, the hybrid and innovative AEDE-ELM model was suggested which uses the innovative and highly effective algorithm AEDE in order to optimize the operating parameters of an ELM.*
- ✓ *The CNN feature extraction method was also used for the training of the model.*
- ✓ *The system can operate in an autonomous mode, because it estimates the geolocation of the species and it decides if it can be considered local or invasive.*

# Future Work

- ✓ *A future extension would be the use and incorporation to the system, of other methods for similar characteristics determination to make the feature map like: Representational Similarity Analysis, Local Similarity Analysis, Isoperimetry and Gaussian Analysis.*
- ✓ *Finally, it would be very important to use and test the performance of the Deep Extreme Learning Machines technology for this specific case.*

Forest Informatics Laboratory
FILab

# References


- ✓ *Carsten S., Ulrich M., Wiedemann C., (2008), Machine Vision Algorithms and Applications. Weinheim: Wiley-VCH. p. 1. ISBN 978-3-527-40734-7.*
- ✓ *Krizhevsky A., Sutskever I., Hinton G. E., (2012), ImageNet Classification with Deep Convolutional Neural Networks, In Advances in Neural Information Processing Systems 25.*
- ✓ *Svellingen C., Totland B., White D. Ovredal T. (2006), Automatic Species Recognition, length measurement and weight determination using the CatchMeter CV System.*
- ✓ *Cabreira A. G., Tripode M. and Madirolas A. (2009). "Artificial neural networks for fish-species identification". – ICES Journal of Marine Science, 66:1119–1129*
- ✓ *Rova A., Mori G., Dill L. M., (2007). One fish, two fish, butterfish, trumpeter: Recognizing fish in underwater video, Conf on Machine Vision Applications, p. 404–407.*
- ✓ *D.J. Lee, R. Schoenberger, D. Shiozawa, X. Xu and P. Zhan (2008). Contour matching for a fish recognition and migration monitoring system. Stud. Comput. Intell., 122: 183-207*
- ✓ *S.O. Ogunlana, O. Olabode , S.A. A. Oluwadare & G. B. Iwasokun, Fish Classification Using SVM, IEEE, African Journal of Computing & ICT, Vol 8. No. 2 June, 2015*


Forest Informatics Laboratory
FiLab



**Adaptive Elitist Differential Evolution Extreme Learning Machines on Big Data: Intelligent Recognition of Invasive Species**

# Thanks

kdemertz@fmenr.duth.gr
http://utopia.duth.gr/~kdemertz/
liliadis@fmenr.duth.gr
http://filab.fmenr.duth.gr/iliadis.html

Forest Informatics Laboratory
FILab