2.1 Artificial intelligence and data mining basics:

Artificial Intelligence (AI) is the field of engineering that attempts to reverse engineer functions of the human brain reproduced in silico. Abe (2005) defined two complementary branches, they are “Knowledge based” systems, and the creation of an autonomous agent that can exhibit an adaptive behavior. One of approach in the category of Knowledge-Based Systems is a simulation of human thinking through intelligent programming.

The term Artificial Intelligence has received many definitions during the past decades.

The term ‘intelligence’ is followed by typical definitions including:

1. Ability to comprehend

2. The information acquisition

3. Innate ability for interpreting the environment

4. The ability to comprehend events, facts or situation awareness

5. The capacity for problem solving by planning and reasoning, deriving abstractions, capture ideas, language understanding, and adaptive behavior.

Stair and Reynolds enlarged this definition by including the use and the synthesis of symbolic information capturing experience in the form of heuristic rules.

The primary focus of AI aims to develop intelligent machines by employing principles from phycology and philosophy.

The objective with phycology based modeling concerned the simulation of ‘intelligent’ human behavior through constructing software reproducing human behavior (Pfeifer & Bongard, 2007).

The goal of the philosophy rendered approach was to develop a simulation of intelligence as a computational entity (Russell & Norvig, 2001).

With evolution of sensing technologies, there are huge amounts of the gathered data in real time from the monitoring of various industrial and natural processes. The process steps that concern the analysis of large datasets for supporting decision-making require a framework that is able to extract fit for purpose knowledge for decision algorithms.

Depending on the type of framework used for extracting the necessary knowledge for developing the decision framework the knowledge acquisition procedure appears with terms such as data mining process, knowledge extraction, pattern recognition and machine learning.

More specifically, data mining involves the exploratory analytics procedure of large datasets to reveal recurring patterns to obtain correlations and to construct association rules and inference mechanisms. The process requires collaborative and synergistic activities of application specialists and information technology analysts who use relevant algorithms for capturing knowledge for decision rule formation. Interventions to the development of decision rules by the aforementioned experts during different phases are necessary for prohibiting automation of the current process. The acquired knowledge must be to repeatable and predictable results of measurable accuracy so as it to later verified by the future data and subsequently leads to trustworthy decisions.

Data Mining architecture is defined as the global structure of information processing pipe lying which includes information acquisition from different sources, extraction of meaningful patterns and statistical indicators for further analysis, derivation of abstraction models with predictive characteristics and actionable insights leading to adapted optimized practices and solutions.

The term mathematical learning theory refers to the majority of mathematical models that constitute the data mining process that lead to knowledge extraction.

The data mining procedure based on inferencing algorithms, aim to derive an inferential framework based on data regularities, expressed as patterns that lead to repeatable conclusions and on which we can base future predictions and decisions.

The generalization of these models is the ultimate outcome of the data mining process since a limited set of examples that have formed these models can generalize in a very large number of future examples and draw real-time conclusions in an operational mode by feeding future novel data to these predictive models.

The predictive models have been derived from data can take different shapes, such as linear operations, heuristic rules in inferential form, charts, decision trees, nonlinear models such as Neural Networks (NNs) and Support Vector Machines (SVMs).

The data collection process is considered mostly autonomous and oblivious of the data mining purposes, thus it does not appear to have any similarities with data collection practices that follow any classic statistical scheme. From this aspect, data mining is characterized as a secondary way of data analysis. Taking into account the main aim of data analysis, the Data mining enterprises can be split to two very important research factors: interpretation and prediction.

• **Interpretation** aims to find ordinal motifs in the data, describe them with the help of novel rules and principles to provide specialists with the adequate information and knowledge about the system investigated.

The interpretation is capable of finding associations between data sets that lead to the formulation of a rule that connects these two datasets embodying the knowledge that associates the physical phenomena that can have produced these data and apart from the association which is a statistical feature, it can illuminate the causality between the physical phenomena that have been associated and reveal hidden mechanisms behind the phenomena that are explained by physical principles.

• **Prediction.** The aim of prediction is to estimate the future value of a process at a certain time base on previous observations. There are miscellaneous calibration methods used to develop data mining models. A plethora of techniques had been proposed by computer science research such as classification and decision trees or heuristic rules, and more recent machine learning methods and neural networks.

The Data mining can be summarized in the following steps:

1st Step. The ANNs models are trained with time-series input data for achieving a trustworthy prediction.

2nd Step. Once input data collection is completed, the time-series data are filtered properly so as to be de-noised and alter their time series form to multidimensional before being fed in to the ANN model.

3rd Step. The employed ANN models are fine-tuned with historical data—The accuracy of the ANN models is evaluated with unseen data that are not included in the training set so that the ANN model accuracy is tested with an independent error criterion.

4th Step. The calibrated neural network outputs are used as predictors for predicting future values of a process and therefore can be used for preventive measures and to facilitate actions based on these predictions.

2.3 Artificial neural networks applications in Biosystems engineering:

The leading scientific journal published by the European Society of Agricultural Engineers (EurAgEng), Biosystems Engineering Journal, defines Biosystems Engineering term as the effective combination of the education and research in Physical Sciences and Engineering aiming to improve from a technical point of view, the biological systems function in terms of sustainability, food safety and quality, crop and land use management and the environmental protection.

According to the American Society of Biosystems Engineers (ASBE), Biosystems Engineering: involves the development of efficient and environmentally safe methods for food production as well as the use of renewable energy sources to meet the needs of an ever-increasing world population.

In conclusion, Biosystems Engineering incorporates engineering principles to sciences that are closely related to biology, agriculture and environmental sciences in such a way so as to guarantee the viable food and crop production that meet high national and qualitative standards taking into account the efficient utilization of natural resources.

More specifically, Biosystems Engineering has a broader scope than that of Agricultural Engineering field because it is regarded as the application of a plethora of engineering sciences not only to agricultural systems but also to living organisms.

Biosystems Engineering cover the following scientific fields:

- **Power and Machinery :**

Advances in artificial intelligence and agriculture machinery engines, have enabled the utilization of biofuels by flexible adaptation of the engine operation to optimize the efficiency of the machine by fine-tuning the engine cycles and the emissions profile leading to cost reduction and minimal environmental footprint.

Pantazi, Moshou, Kateris, Gravalos, and Xyradakis (2013) studied the vibrational profile of a four-stroke tractor engine fed by pure gasoline and gasoline mixtures with ethanol and methanol in different proportions (10%,20% and 30%). For the determination of optimal fuel blend, Machine Learning techniques including Neural Networks, Active Learning and original Novelty Detection algorithms have been utilized. An accuracy of 90% was achieved in detecting both type and mixture percentage of bioethanol.

- **Post-harvest Technology:**

Photonic technologies like a reflectance sensing and chlorophyll fluorescence can produce signals that can automate the decision process needed in advanced quality sorting of apples in a nondestructive manner and in real time.

Moshou et al. (2005) has proven the link between the chlorophyll fluorescence and the mealiness presence in apples, supposing mealiness and chlorophyll degradation are correlated. Their study has proven that through the employment of advanced data mining architectures like Self Organizing Maps is more performant compared to other more conventional techniques like discriminant analysis and multilayer perceptrons demonstrating a high accuracy in the automated fruit sorting with the assistance of non-destructive methods.

• Livestock farming:

To ensure the quality of animal and avian production, the volatiles and other emissions are necessary to be reduced to minimize health risk both to animals and to humans. Pan and Yang (2007) proposed an automated volatile sensing system consisted by an olfactory sensing developed specifically for registering and analyzing the odors from animal emissions in farms.

The proposed tool called "Odor Expert" can be used both under lab and farm conditions by employing a typical backpropagation algorithm which achieved a performance of r = 0.932.

Results indicated that the "Odor Expert" demonstrates a superior performance compared to a human panel. Therefore, is regarded as a useful tool for odor management practices in livestock and poultry farms. The timely detection of health degradation in animals can result in better animal health management thus minimizing the impact of antibiotics and avoiding further complications.

Moshou et al. (2001) proposed an ANN based technique concerning swine cough recognition. A neural network was trained in order to discriminate the swine cough from other ambient sound in the piggery area. The best results were obtained by a Self-Organizing map, which was able to recognize the coughs from other sounds by a percentage >95%.

• Food Security through early warning intelligent systems:

Olive oil adulteration is a major concern and most of the times cause of economical profits loss and general threat for human health. Most of the times the olive oil adulteration with hazelnut oil is too difficult to be detected since their similar chemical composition.

Garcı´a-Gonza´lez, Mannina, D'Imperio, Segre, and Aparicio (2004) investigated this problem by proposing an artificial neural network based on hydrogen and carbon based nuclear magnetic resonance signatures (1 HNMR and 13C NMR respectively). In terms of this study, A multilayer perceptron was employed to detect the type of blend of the two oils including pure olive oils, pure hazelnut oils, and mixtures of them.

• Crop Monitoring for Disease Detection and Stress Phenotyping:

The growing conditions of crops should be monitored continuously to achieve incipient diseases, weeds and biotic and abiotic stress factors. Reduction of pesticides can be achieved by targeting the infected areas of the field with site-specific interventions (chemical or mechanical).

This type of crop management results in much higher production with simultaneous reduction of inputs adjusted to the needs of the crop and to lower costs for maintaining the production at high level, while at the same time the environmental impact is minimized due to precise fertilization, phytosanitary interventions like spraying and targeted irrigation.

DeChant et al. (2017) presented a phenotyping approach by recognizing stress symptoms in tomato leaves, with the help of a viral, fungal and bacterial and pest infection images by applying deep ANN architectures from the public domain including AlexNet and GoogLeNet.

All the sub disciplines of Biosystems Engineering rely on mathematical modeling and data acquisition to calibrate and use these models for prediction. Therefore, Data Mining is a indispensable tool for Biosystems Engineering applications since the knowledge discovery and model building enables the different applications that are needed for achieving practical engineering systems for the interpretation and control of biological commodities and environmental monitoring.

A special domain integrating Biosystems applications and data mining in the form of sensor based interpretation and control of agricultural production systems is Precision Agriculture

2.4 Contribution of artificial intelligence in precision agriculture (economical, practical etc.):

A variety of machine learning methods have been employed in the field of precision agriculture including Supervised and Unsupervised Learning for the assessment of plant health condition and invasive plant species recognition by interpreting spectral signatures and optical features.

Advancements in sensor engineering, machine learning and geo-spatial analysis systems have enabled potential for precision agriculture enabling biotic and abiotic stress detection in crops at an early stage .

The current techniques are capable of offering rapid, accurate and trustworthy decisions, in comparison to the conventional ones.

The precise invasive plant species recognition is considered a precondition to determine the type of suitable treatment.

A number of machine learning methods have been applied to weed recognition , the majority of them rely on morphology features, weed leaf shape geometrical features for classification.

Detecting and forecasting crop health status are the principles for sustainable and efficient crop protection including approaches for the discrimination of plant biotic stress through machine learning reaching a classification range from 95% to 99%.

The identification of insects through optical sensing and AI techniques, the derivation of Vegetation Indices (Vis) from reflectance spectral data and the relationship to damage severity is more as more frequent approach.

Prabhakar (2011) investigated the detection of cotton infection infestation by leafhopper by utilizing specific wavebands emanating from linear intensity graphs associating reflectance and the level of infestation.

To sum up, a promising development of a variety of AI techniques in precision agriculture arises, contributing to the security and the sustainability in crop production. According to the current state of the art in sensing technologies and data analytics based on machine learning, the following projections and their respective contribution to precision agriculture are envisaged:

1. Customization of mainstream AI techniques to crop production. Attention should be paid to the fusion of information from the spatial, spectral and time series of crop parameters to detect trends related to crop status. The combination of advanced data analytics and the visualization mapping of the crop problems can lead to timely interventions that lead to reduction of treatment chemicals and crop yield loss avoidance. These projected outcomes guarantee a lower environmental impact and higher profit margins.

2. Technologies combining improvements in hardware of sensors will make them more compact and embeddable in field-deployed devices. These compact sensors will be equipped with board decision algorithms allowing the infield real time detection of crop anomalies and identification of the situation leading to event based interventions and management. The trend towards field automation by integrated sensors and software modules in filed deployable devices will make available big data for real time analysis of crop status and will allow event based decision algorithms that are available in autonomous machinery and the smart mobile devices to manage the crop status in real time. This will result in optimal crop production and quality leading to minimization of crop losses and higher competitiveness for crop producer.

3. The data fusion from different sources will enable the construction accurate and reliable models capable of assessing the potential of future states within a field—and fuse it with climate projections and soil data derived from soil maps to predict the yield potential and assess crop suitability concerning the examined fields. The effective combination of machine learning methods and sensor technologies can function as efficient tools for agricultural advisory boards and experts by assisting them into reaching conclusions, minimizing the risk of possible yield loss, adaptation of Best Management Practices (BMPs) to reduce quality variability and reducing inputs for sustainable crop production by ensuring higher economic benefits.

2.5 Support vector machines (SVMs):

SVMs (Vapnik, 1998) are machine learning technique, mostly employed for pattern recognition, nonlinear function approximation.

SVMs demonstrate important properties, when compared to other machine learning architectures including neural network learning and particularly algorithms for training architectures.

SVM and RBN became highly used image processing methods for data mining and recognition.

In the same fashion as RBN, SVM functions based on calculating a distance metric among data vectors to allocate data samples to clusters (Courant & Hilbert, 1954). With the help of the idea of structural risk minimization, this characteristic of a distance-calculated metric enhances the robustness of SVM by clearing noise in the training procedure, which induces a better performance in classification compared to other classifiers like decision trees, maximum likelihood, and neural network-based approaches (Huang, Davis, & Townshend, 2002; Bazi & Melgani, 2006).

Contrary to the above referred classifiers, an SVM classifier is binary by definition. Provided that more than two classes are usually considered in remote sensing image processing, the way of building a multiple-class SVM classifier is a problem that needs solution.

The main asset of SVMs is their capability of dealing with large volumes of data without encountering calibration bottlenecks that are usual in multilayer architectures that exhibit complex error surfaces with trapping regions.

On the other hand, a further characteristic of SVM is the efficient learning with small amount of data, due to their simpler, more effective architecture and learning procedure. A classification mapping can be defined as follows: $f = X \rightarrow \{+ / - 1\}$

A crucial step is to characterize and predict f due to the sensitivity and accuracy of separating overlapping classes, which is defined as capacity of the SVM.

Low value capacities cannot lead to trustworthy and accurate estimation of multi-parametric functions, while high value capacities will attempt to describe the data with superfluous paramaters leading to estimation error a phenomenon known as “overfitting”. Overfitting is a major problem in training of ANNs. To overcome this problem two main techniques are employed.

The first one is known as “early stopping” and the second method is targeting on reducing the roughness of the error function, which corresponds to spurious noise contribution that might be due to local deviations of the data producing mechanism by penalizing such deviations through regularization of the error function. Similarly, in SVMs the over fitting is avoided by utilizing regularization functions. An advantage of SVMs is their ability of being calibrated with small datasets so they avoid the problem of over fitting (Vapnik, 1998). The less complex functions that are used for reaching decisions are the linear ones

2.6 One class classifiers:

In application where safety is a high priority, novelty detection is crucial forestimating the threshold of deviation from normality. Baseline system state

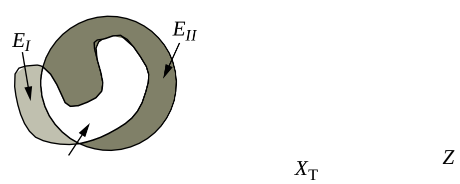

Fig. Error areas defined in connection with the one-class classifier and the targetdomain (Tax & Duin, 2004).

may deviate due to wear, structural alterations, environmental changes are held responsible for altering the operational state of a system. The estimation of robust thresholds is important for credible novelty detection. For novelty detectors which utilize one class classifiers it is crucial to acquire all possible working stages that are attributed to normal operating conditions so that the one class classifier covers all possible normal operation states so that the robust thresholds signify faithful limits of these operations conditions leading to trustworthy outlier detection decisions.

One-class classification scheme can be described:

- Only baseline data are used as target without inclusions of outliers; The boundary is set between two classes is defined only from target classdata;
- The robustness of the boundary condition is associated with the crite- rion of maximizing the correct classification of as many target examples as possible while at the same time keeping the accepted outliers.

figure, demonstrates a target dataset area X_T. E_I represents the error of the correclty rejected target examples while E_{II} symbolizes the incorrectly classified oulier objects as belonging to the baseline data-set. The area in the circle represents the area of definition of the one class classifier.

By assuming that the outlier data are distributed evenly, the consequences of shrinking the domain E_{II} to a minimal data description concerns the mini- misation of false acceptance. This can be achieved by simultaneously mini- misation of the volume of one class classifier and falsely accepted sector E_{I}.

2.7.1 One class classification

classifiers One-class classifiers demonstrate the following features:

• The only available information concerns the target class;

• The determination of the threshold that defines the two classes is constrained on the estimation of data originating completely from the target class;

• A critical problem concerns the threshold definition around the target class having the objective to enclose the maximum of the target objects and simultaneously to exclude outliers. A number of One Class classifiers have employed in the presented research:

• Support Vector Machines
• Autoencoder Neural Network
• Mixtures of Gaussians (MOG) and
• Self-Organizing Maps (SOM).

2.7.2 SVM based one-class classifier

One Class Support Vector Machines construct a model based on the target data. In the operational phase they allocate a new data by using the deviation from target data as a membership metric (Scholkopf, Platt, Shawe- € Taylor, Smola, & Williamson, 2001). The value of the spread parameter determines the sensitivity of the classifier according to the kernel K(x,z)¼exp{kxzk2 /σ 2 } which is more sensitive with a small spread in order to classify a data with nonlinear decision borders while a large spread works better for linearly separable data due to lower sensitivity.

2.7.3 Auto-encoder based one-class classifier

Auto-encoders (Japkowicz, Myers, & Gluck, 1995) are ANNs that are trained so they can reproduce input data as a mirror image in their output layer. In this classifier, only one hidden layer is used with less hidden neurons with sigmoidal transfer functions. The main assumption in auto-encoder networks concerns the full reconstruction of objects from a target dataset resulting at a smaller error compared to outliers. In the special case of only one hidden layer, the network will find a similar solution to Principal Component Analysis (PCA) algorithm (Bourlard & Kamp, 1988). The network tries to achieve small reconstruction error and therefore manages to achieve that by developing a compact mapping from the input data to the subspace of the hidden neurons. The dimension of this subspace depends on the number of number of hidden neurons. Auto-encoder networks offer many degrees of freedom but they demonstrate similar behavior to multilayer perceptron due to their similarity (Tax & Duin, 2001b). This behavior is mainly associated to improper selection of training parameters and reduced flexibility due to predefinition of the Auto-encoder construction.

2.7.4 MOG based one-class classifier

The simplest way for generating a one-class classifier is to estimate the topological variation of the training data samples (Tarassenko, Hayton, Cerneaz, & Brady, 1995) and accordingly to define a border delimiting this area while the same class data are situated. The optimal border defines a minimum volume connected to a model of a probability density of samples that belong to the target distribution. In order to obtain a model with more adaptive and sensitive modeling behavior, the basic spherical distribution can be extended to a mixture of Gaussians (MOG) which provides enhanced flexibility for modeling more complex data distributions (Duda, Hart, & Stork, 2001).

2.7.5 Augmentation of one class classifiers

A One Class Classifier is not flexible enough to be used for solving classification problems with multiple classes due to its inherent structure which is directed to One Class Classification. The augmentation of classified data of the target class and outliers allows performing a multi-step procedure where already classified data can be considered as data with a known class and data that do not have any label from a point of view that they do not belong to the already known target or outlier dataset. An iterative scheme where new data augmented with target data, while outlier data are augmented with an initial data in every step can lead to a multiclass classifier, which can classify an arbitrary number of classes. Such a classification scheme corresponds to Active Learning since the acquisition and the classification procedure are performed in an exploratory iteration.

The following iteration outlines the Active Learning procedure:

1. The first training session concerned the calibration of the initial classifier that corresponds to the one class of the target data. After training, the resultant one-class classifier is validated with unknown data. The performance indicator is the successful classification of the new data as outlier with simultaneous classification of the new data belonging to the target set distribution.

2. At this phase, the new baseline set is the product of augmentation of the target set combined with outlier values from the already classified outlier class. The outlier detection process is continued but with the adaptation, concerning the new baseline set which is the augmented baseline of the already known targets and outliers. In the event that an unknown sample is classified as belonging to the augmented class then the procedure of outlier detection is repeated in order to classify internally in the augmented class aiming to find the identity of the new data sample with respect to the initial target dataset and the already known outlier class.

3. Steps 2 and 3 iterated. More precisely, the discovery of the outliers and the step of augmenting those takes place for newly added data that possibly are member of the existent categories or are outliers to the augmented classes. A significant feature of the proposed active learning scheme concerns that the iterations between steps 1 to 3 do not require supervision but are based on outlier discovery and augmentation steps that are executed in an automatic manner