

Self-Organising and Self-Learning Model for Soybean Yield Prediction

Mona Alghamdi
Computer Science
Imam Abdulrahman Bin Faisal
University
Jubail, SA
msmalghamdi@iau.edu.sa

Plamen Angelov
Computer and Communication
Lancaster University
Lancaster, UK
p.angelov@lancaster.ac.uk

Raul Gimenez
Lancaster Environment Centre
Lancaster University
Lancaster, UK
gimenezgea@gmail.com

Mariana Rufino
Lancaster Environment Centre
Lancaster University
Lancaster, UK
m.rufino1@lancaster.ac.uk

Eduardo Soares
Computer and Communication
Lancaster University
Lancaster, UK
e.almeidasoares@lancaster.ac.uk

Abstract— Machine learning has arisen with advanced data analytics. Many factors influence crop yield, such as soil, amount of water, climate, and genotype. Determining factors that significantly influence yield prediction and identify the most appropriate predictive methods are important in yield management. It is critical to consider and study the combination of different crop factors and their impact on the yield. The objectives of this paper are: (1) to use advanced data analytic techniques to precisely predict the soybean crop yields, (2) to identify the most influential features that impact soybean predictions, (3) to illustrate the ability of Fuzzy Rule-Based (FRB) sub-systems, which are self-organizing, self-learning, and data-driven, by using the recently developed Autonomous Learning Multiple-Model First-order (ALMMo-1) system, and (4) to compare the performance with other well-known methods. The ALMMo-1 system is a transparent model, which stakeholders can easily read and interpret. The model is a data-driven and composed of prototypes selected from the actual data. Many factors affect the yield, and data clouds can be formed in the feature/data space based on the data density. The data cloud is the key to the IF part of FRB sub-systems, while the THEN part (the consequences of the IF condition) illustrates the yield prediction in the form of a linear regression model, which consists of the yield features or factors. In addition, the model can determine the most influential features of the yield prediction online. The model shows an excellent prediction accuracy with a Root Mean Square Error (RMSE) of 0.0883, and Non-Dimensional Error Index (NDEI) of 0.0611, which is competitive with state-of-the-art methods.

Keywords—Autonomous Learning, Fuzzy Rule-Based, Multi-Modal, Linear regression, Soybean yield prediction

I. INTRODUCTION

Estimation of agricultural crop yields and production is of paramount importance for global food security. Crop yields result from the interaction of multiple factors such as soil and weather conditions, water availability, crop species, and management practices (sowing date, planting rate, etc). However, the relative relevance of the different factors varies strongly from place to place and from season to season [21]. Understanding the effect of these factors on crop yields is crucial for guiding farmer's decisions to optimize farming practices and outputs in each region. Although high quality data on these factors is becoming increasingly available at the plot level [22], its analysis remains challenging [23]. Machine

learning can be used as advanced data analytics to predict crop yields and determine the influence of the various factors on prediction accuracy in an intelligible and understandable manner, thereby providing support for farm decisions. Thus, machine learning can be used to promote successful agriculture and tackle the challenges in agricultural production. Extracting knowledge from data facilitates intelligent decision-making by humans. Over the last two decades, many machine learning algorithms were introduced that provide good prediction of linearly separable classes [1]. However, processing large and complex data streams requires non-linearly separable classes. Numerous non-linear methods are used to solve complex problems related to data streams [2, 3]. The Fuzzy Rule-Based (FRB) system is considered a universal approximator [4], and the neuro-fuzzy system is an instance of FRB, integrating an Artificial Neural Network (ANN) and an FRB system [5], by making use of ANN's layered structures to represent a set of fuzzy rules. It could be interpreted as a multi-layer feed-forward neural network, or as a degree of likelihood in a probability distribution [2]. In 2018, the ALMMo-1 system, which is an extended development of a neuro-fuzzy system was proposed. The model uses an empirical approach for data partitioning with the robust design of an FRB system. It has a data-driven mechanism that learns autonomously from the underlying patterns of data streams [5]. ALMMo-1 can be a competitive proposed solution for the task in use. It is an online evolving mechanism that does not need any prior assumptions or parameters and can be easily adapted to the environment, due to the associated self-learning and self-organizing mechanisms. The objective of this paper is the following: to use ALMMo-1's method to precisely predict soybean yields, from a database that combines environmental information, agricultural management and remote sensing information from commercial plots of Central Argentina. Furthermore, we determine the coefficient between factors from underlying patterns of data, thereby identifying the most important factor affecting soybean yield prediction, and compare the performance of the model with other well-known methods.

In simple terms, the ALMMo-1 model transforms the inputs, which are factors that affect the soybean yield, to the corresponding output, which is the prediction of soybean yield as shown in fig.1.

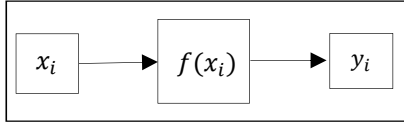


Fig. 1. Transferring the input x_i to the output y_i after the self-learning and self-organising of data

The simplest form of the model can be represented using the following equation:

$$y_i = f(x_i) \quad (1)$$

Where x_i represents the factors that affect the yield, $f(x_i)$ is the model during the learning process using the antecedent of FRB, and y_i is the consequent parameters. The remainder of this paper is structured as follows. In Section (II) we discuss the agricultural/ economic relevance of machine learning methods for crop prediction. In Sections (III), (IV) and (V), the proposed computational approach, experimental results and evaluation, and alternative methods are presented, respectively. Finally, in Section (VI) we summarize the conclusions of this work.

II. AGRICULTURAL/ ECONOMIC RELEVANCE

Soybean is one of the most valuable crops in the world, not only as an oil seed crop and feed for livestock, but also as a good source of vegetable protein for the human diet and as a biofuel feedstock [24]. Argentina is an important soybean supplier being the third largest soybean exporter in the world, and the largest exporter of soybean derivatives (cake, oil and biodiesel) [25]. In addition, soybean is the main crop grown in the country, with ~ 19 Mha harvested per year across very different environmental conditions [26]. The central role of this crop to the economy of the country demands a better understanding of the multiple factors involved in yield determination, to optimize farming practices and production outputs. In this paper, we analyse a database containing soybean yield and management practices recorded for commercial plots in Central Argentina over 13 crop seasons (2005/06 to 2017/18), by the private farmer's association AACREA [27]. In addition to the provided data, we included rainfall and remote sensing data from the Moderate Resolution Imaging Spectroradiometer (MODIS) [28] as important predicting factors, to account for the weather and vegetation growth conditions for each plot and season, respectively. We identified five important factors that determine soybean yields in Central Argentina: three management factors which include "sowing date" (expressed as the number of days from Jan. 1st), "plant density" (the number of plants per square meter), "maturity group" (the length of the crop lifecycle); "rainfall amount" (total rainfall between 1-Nov and 31-Mar, which accounts for the weather conditions) and "the maximum vegetation index" (the maximum value of the Enhanced Vegetation Index (EVI) registered during the crop cycle as a proxy of the crop condition).

III. THE PROPOSED COMPUTATIONAL APPROACH

The proposed data-driven approach uses an FRB framework with the recently introduced ALMMo model [5]. The model has many robust characteristics with which it makes decisions: learning from the stream data, self-adapting to environmental changes, self-monitoring from the data stream [2], [6], [7]. The FRB system is a multi-model system with the ability to learn autonomously and requiring neither human experts nor prior assumptions [6]. In this paper, the model predicts the yield of soybeans based on the densities of underlying data represented by the yield's factors. In general, there are two types of ALMMo models: the zero-order ALMMo-0 and the first order ALMMo-1. The former, is used for classification, while the latter (considered in this paper) is used for regression. The structure of the model consists of data clouds, which represent the building blocks of the multi-model system [5]. Meta-parameters are generated from data clouds, which are formed completely from observed data without either user-specified or problem-specified parameters, and they are updated recursively. The model has sophisticated characteristics, which distinguish it from other methods because it is built based on Empirical Data Analysis (EDA), which employs local and global typicality, cumulative proximity, square centrality, and density [8]. Due to the recursive update of data, the model can run with a few samples of data, which promotes the computational efficiency and reduces memory consumption [5]. In the following subsections, more details about the approach will be presented.

A. Normalisation

Normalization is used to adjust the values of each feature input into a specific range. In statistics, there are many types of normalization. In this paper, minimum and maximum feature scaling is used according to the following expression [1]:

$$x_i = \frac{x' - \min(x'_i)}{\max(x'_i) - \min(x'_i)} \quad (2)$$

The key notation used in this paper is described in table I. After applying the above equation, the values of each feature lie in the range $[0,1]$.

B. The Architecture of ALMMo-1 model

The architecture of the ALMMo-1 model is essentially composed of multiple models, which are used in adaptive, control, prediction, and classification [6]. Multi-model systems follow the principle of "divide and rule" [2, 6, 9] by decomposing complex problems into a set of simpler ones and combining these afterwards. Examples of this model include, but are not limited to, FRB systems with a set of AnYa type local sub-systems initially proposed as an alternate to Mamdani and Takagi-Sugeno types. These FRB systems follow the actual data distribution and do not involve any aggregation or scalar membership functions because the data cloud density replaces the scalar membership function. These data clouds represent the data density distribution of observed data and derive the fuzzy membership of a particular IF... THEN of FRB. The construction of the IF part

consists of focal points that represent the data clouds. While, the consequent THEN part defines the local linear model [10],[5]. Each local linear model is obtained from an AnYa type fuzzy rule with prototype p_z and parameter vector a_z in the consequence part as follows [5]:

$$\text{IF } (x \sim p_z) \text{ THEN } (y_z = a_z x^T) \quad (3)$$

$$y = \sum_{z=1}^N \lambda_z a_z x^T \quad (4)$$

The key novelty of the ALMMo-1 model is the use of the EDA framework. The EDA depends entirely on the actual data; therefore, it is free from parameters, and assumptions [5],[11]. The main characteristics that have been used in the proposed model are unimodal discrete density D [8] and multimodal continuous typicality τ^{MM} [12]. R^M is the set of the actual samples in the data stream and is denoted $\{x_S\} = \{x_1, x_2, \dots, x_S\}$, where the subscript S indicates the time instance. Unimodal discrete density D is generated from the ensemble properties and distributions of actual data samples and identifies the distance of a particular data sample in relation to all other data samples. The Euclidean distance is the most common type of distance used. The definition of the unimodal discrete density of the data sample x ($x(x) \in \{x\}_S$) formed from Cauchy function is [8]:

$$D_S(x) = \frac{1}{1 + \frac{\|x - \mu_S\|^2}{\sigma_S^2}} \quad (5)$$

The average scalar product X_S and the mean μ_S updates recursively as follows [2]:

$$\mu_S = \frac{S-1}{S} \mu_{S-1} + \frac{1}{S} x_S; \mu_1 = x_1 \quad (6)$$

$$X_S = \frac{S-1}{S} X_{S-1} + \frac{1}{S} \|x_S\|^2; X_1 = \|x_1\|^2 \quad (7)$$

1) The identification of the antecedent (IF) part

When each data cloud arrives, known as x_{S+1} , the model updates the global mean μ_S and the average scalar product X_S using (6) and (7) to obtain μ_{S+1} and X_{S+1} , respectively. Then, the unimodal discrete density D and the focal points of the current data cloud $\mu_{S,z}$ for the fuzzy rule z is updated using equation (5), denoted by $D_{S+1}(x_{S+1})$ and $D_{S+1}(\mu_{S,z})$. The first condition checks if a new rule will be generated as Cond.1:

$$\text{IF } (D_{S+1}(x_{S+1}) > \max_{z=1,2,\dots,N} (D_{S+1}(\mu_{S,z})) \text{ OR } (D_{S+1}(x_{S+1}) < \min_{z=1,2,\dots,N} (D_{S+1}(\mu_{S,z})) \text{ THEN } (x_{S+1} \text{ is a new focal point}) \quad (8)$$

When the first condition is met, a new focal point for the new data cloud is identified. Then, the second condition examines whether the new data cloud overlaps as Cond.2:

$$\text{IF } (D_{S+1,z}(x_{S+1}) \geq \frac{1}{1+n^2} \text{ THEN } \left(\begin{array}{l} \text{the } N\text{th focal point and the respective data} \\ \text{cloud needs} \\ \text{to be replaced by a new one} \end{array} \right) \quad (9)$$

In the case of the first and second conditions are not fulfilled, the model creates a new rule with a new associated focal point x_{S+1} , and the model initialises other parameters. However, when the first and second condition are fulfilled, the intersecting data cloud is removed and replaced by a new rule with a new associated focal point N_{S+1} . When the first condition is not satisfied, x_{S+1} merges to the closest data cloud. The activation level of the z^{th} fuzzy rule at the S^{th} time instance is $\lambda_{S,z}$, which is defined as the normalised unimodal density per rule as follows:

$$\lambda_{S,z} = \frac{D_{S,z}(x_S)}{\sum_{z=1}^{N_S} D_{S,z}(x_S)} \quad (10)$$

TABLE I. DESCRIPTION OF THE KEY NOTATIONS IN THIS PAPER

Notation	Description
x'	The feature input before normalisation, $x' \in R^2$.
x_i	The feature input after normalisation. $x_i = [x_{i,1}, x_{i,2}, \dots, x_{i,M}]^T \in R^M$, $i = 1, 2, \dots, S$, and M is the features number.
$\min(x'_i)$	The minimum value of all the data input in the dataset, $i = [1, 2, \dots, S]$.
$\max(x'_i)$	The maximum value of all the data input in the dataset, $i = [1, 2, \dots, S]$.
p_z	The z^{th} data cloud that represents a fuzzy rule, and $z = 1, 2, \dots, N$; in this paper $N=16$.
y_z	The output of the z^{th} fuzzy rule.
x^T	The parameters of the linear model, $x^T = [1, x^{T-1}] = [1, x_1, \dots, x_M]$.
a_z	The coefficient of each (consequent) parameter of the z^{th} fuzzy rule.
$a_{S,z}$	The coefficient of each (consequent) parameter of the z^{th} fuzzy rule at time S , and $a_{S,z} = [a_{0,z}, a_{1,z}, \dots, a_{M,z}]^T$.
$a_{S,N_{S+1}}$	The new coefficient of each (consequent) parameter associated with new focal point N_{S+1} .
λ_z	The activation level of the z^{th} fuzzy rule.
$\lambda_{S+1,z}$	The activation level of the z^{th} fuzzy rule at instance time $S + 1$.
D_S	The unimodal discrete density of the data set $\{x\}_S$.
μ_S	The mean of the data samples in S dataset.
σ_S^2	The deviation is a measure of difference between the actual value of a variable and other data samples on the data set $\{x\}_S$, $\sigma_S^2 = X_S - \ \mu_S\ ^2$, X_S is the average scalar product and μ_S is the mean.
N_{S+1}	N_S is a number of rules at time S , when a data sample arrives. N_{S+1} is a new rule and focal point x_{S+1} .
X_S	The average scalar product.
$C_{S+1,z}$	The covariance matrix at time $S + 1$ of the fuzzy rule z in the consequence parameter.
$C_{S,N_{S+1}}$	The covariance matrix at time S and the focal point N at time $S+1$ in the consequence parameter.
x_{S+1}	The input feature at time $S + 1$. x_S is the input feature at time S .

2) The identification of consequence (THEN) part

The system updates the consequence parameters. When a new rule is generated, the consequence parameters are established as follows:

$$a_{S,N_{S+1}} \leftarrow \frac{1}{N_S} \sum_{z=1}^{N_S} a_{S,z} \quad (11)$$

$$C_{S,N_{S+1}} \leftarrow \Omega I_{(M+1) \times (M+1)} \quad (12)$$

The new rule inherits the consequence parameters $C_{S,z}$ and $a_{S,z}$ of the old rule, when both the first and second conditions are met. The omega parameter Ω is a constant parameter, and is set to 1000 in this paper. To optimize the consequent parameters of each rule, and the fuzzy weighted recursive least squares is used [13] as follows:

$$C_{S+1,z} = C_{S,z} - \frac{\lambda_{S+1,z} C_{S,z} x_{S+1} x_{S+1}^T C_{S,z}}{1 + \lambda_{S+1,z} x_{S+1} x_{S+1}^T C_{S,z}} \quad (13)$$

$$a_{S+1,z} = a_{S,z} + \lambda_{S+1,z} C_{S+1,z} x_{S+1} (y_{S+1} - x_{S+1}^T a_{S,z}) \quad (14)$$

IV. EXPERIMENTAL RESULTS AND EVALUATION

The dataset used in this experiment relates to agricultural data of soybean crop yields. The most influential factors affecting soybean yields are used as inputs to predict the yield and produce an autonomous learning multi-model regression model based on FRB. The importance of the method is that the model allows the stakeholders to navigate into the inner model representation and interpret the outcomes. The model has linguistic IF... THEN statements, which are interpretable, understandable, and explainable [29]. In addition, the stakeholders can identify and recognise the different factors that affect the yield. Therefore, they can control the yield based on the factors, “sowing data”, “density”, “maturity group”, “amount of rain”, and “the maximum of vegetation index”. For example, the yield can be increased based on the factors used in the FRB model that influence the crop. The main contributions of this paper are the use of the ALMMo-1 model in soybean yield prediction and achieving high prediction accuracy. The performance of a model in the learning and validation phases, which are competitive with state-of-the-art methods, are demonstrated in table II.

TABLE II. THE PERFORMANCE OF THE MODEL IN THE LEARNING AND VALIDATION PHASES

Phases	Correlation	RMSE	NDEI
Learning	0.8408	0.0130	0.0621
Validation	0.9142	0.0883	0.0611

A. The formulation of IF... THEN FRB rule

The model that was used generating an identification of fuzzy rules during the learning process from the input features or factors, by creating data clouds formed from many data densities. A prototype of data density represents a component of one fuzzy rule, which consists of five parameters representing the soybeans factors (“sowing date”, “plant density”, “maturity group”, “rainfall amount”, and “the maximum vegetation index”) represented by $x_{S,1}$, $x_{S,2}$, $x_{S,3}$, $x_{S,4}$, and $x_{S,5}$, respectively. In this paper, 16 rules were generated. Three examples of these rules are shown in table III.

TABLE III. EXAMPLES OF FUZZY RULES FORMED FROM LEARNING PROCESS OF DATA CLOUDS

Rule #	The expression
1	IF $\begin{bmatrix} x_{S,1} \sim 169.82 \\ x_{S,2} \sim 42.39 \\ x_{S,3} \sim 4.34 \\ x_{S,4} \sim 520.89 \\ x_{S,5} \sim 7565.95 \end{bmatrix}$ THEN $y = 0.069 + 4.65 x_{S,1} + 0.7134 x_{S,2} - 0.0269 x_{S,3} + 1.952 x_{S,4} + 0.334 x_{S,5}$
2	IF $\begin{bmatrix} x_{S,1} \sim 165.85 \\ x_{S,2} \sim 41.19 \\ x_{S,3} \sim 4.40 \\ x_{S,4} \sim 449.98 \\ x_{S,5} \sim 4757.14 \end{bmatrix}$ THEN $y = 0.02 + 1.66x_{S,1} + 0.37x_{S,2} + 0.02x_{S,3} + 2.27x_{S,4} + 0.32x_{S,5}$
3	IF $\begin{bmatrix} x_{S,1} \sim 165.72 \\ x_{S,2} \sim 41.55 \\ x_{S,3} \sim 4.40 \\ x_{S,4} \sim 549.1 \\ x_{S,5} \sim 7965.96 \end{bmatrix}$ THEN $y = 0.14 + 2.59x_{S,1} + 0.62x_{S,2} + 0.03x_{S,3} + 1.88x_{S,4} + 0.18x_{S,5}$

B. The prediction of soybean yield

This section presents the performance of the model prediction. Fig. 2 shows samples of data ordered from the minimum to the maximum yield. This demonstrates the convergence between the actual and predicted values of the soybean yield.

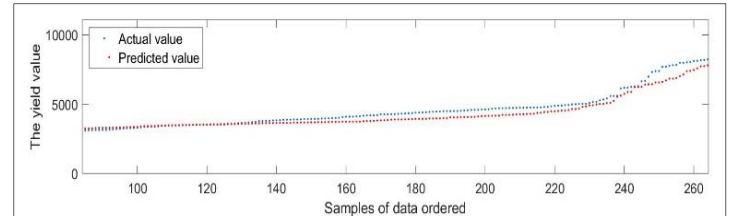


Fig. 3. Comparing actual and predicted yields.

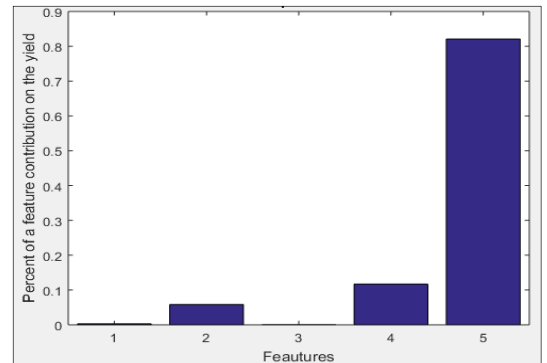


Fig. 2. Percentage of each feature contribution in IF part of FRB.

The percentage of each feature contribution in the antecedent IF part is shown in fig. 3. The features from 1 to 5 indicate “sowing date”, “plant density”, “maturity group”, “rainfall amount”, and “the maximum vegetation index”, respectively.

C. Comparison between ALMMo-1 model and other algorithms

Comparison between the ALMMo-1 model and other popular algorithms is shown in table IV. The model achieves excellent results compared to other well-known algorithms.

TABLE IV. OVERALL PREDICTION PERFORMANCE COMPARISON BETWEEN DIFFERENT ALGORITHMS

ALGORITHM	RMSE
ALMMO-1	0.0883
Ensemble- Bagged Trees	0.0935
Tree- Coarse Tree	0.0994
SVM- Coarse Gaussian SVM	0.1064
Linear regression- Robust Linear	0.1183
Stepwise Linear Regression	0.1120

In addition, the non-dimensional error index (NDEI) is used to evaluate the performance of the proposed method. The NDEI expression is given by:

$$NDEI = \sqrt{\frac{\sum_{i=1}^S (y_{i,actual} - y_{i,estimated})^2}{\sigma_{actual}^2}} \quad (15)$$

Where $y_{i,estimated}$ is a value of the prediction and corresponds to the system output, $y_{i,actual}$ is the actual truth-value, and σ_{actual} is the standard deviation of the actual value. The performance of the proposed algorithm using NDEI is 0.0611.

V. ALTERNATIVE METHODS

Many methods exist to predict crop yield in precision agriculture [14, 15], [16]. The most commonly used methods for predicting crop yields are correlation and multiple regression models. These methods are used to identify essential features affecting the yield. Non-linear statistical methods have been introduced to investigate yield estimation [17], using simple linear correlations of yield with soil properties. However, there are variations between fields and years. Many studies used ANN to predict yields [18], [14]. For example, in a study on setting target corn yields, the root mean square error (RMSE) was 20% [18]. In a comparison of the efficiency of multiple linear regression models and ANN models to predict the soybean yields in Maryland, the ANN model performed better, with an r^2 and RMSE of 0.81 and 214 compared with 0.46 and 312 for the linear regression model [14]. Other methods have also been used. For example, a study of potato yield prediction via satellite image and precision agriculture techniques, employed Vegetation Indices (VIs), which include NDVI and SAV. The method gained an accuracy of between 7.9 and 13.5% for Landsat-8 images and between 3.8 and 10.2% for Sentinel-2 images [19]. A few studies have been proposed with regards to knowledge-based systems using fuzzy logic

techniques in agricultural domain [15], [16], and some authors proposed a framework of a fuzzy rule-based cognitive map [20], [17].

VI. CONCLUSION

In this paper, the ALMMo-1 system is implemented to predict soybean crop yields from factors that affect the yield. The model achieves high accuracy. Distinctive characteristics of the model are the interpretability and readability that allows stakeholders to navigate through the internal model representation, therefore, the factors that impact soybean yields can be adjusted. Also, the model uses nonparametric data clouds, which are assumption-free, constraint-free, totally data-driven, and recursively updated. The model evolves and updates with each data sample entry, improving memory and computation efficiency. In future work, the method could be extended to process the crop images, map yields with satellite data, and evaluate the resulting yield predictions.

REFERENCES

- [1] S. Theodoridis, Pattern recognition, 4th ed. ed. Amsterdam: Elsevier/Academic Press, 2009.
- [2] P. Angelov, Autonomous Learning Systems: From Data Streams to Knowledge in Real-time. 2012.
- [3] I. Skrjanc, J. A. Iglesias, A. Sanchis, D. Leite, E. Lughofer, and F. Gomide, "Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A Survey," Information Sciences, vol. 490, p. 344, 2019, doi: 10.1016/j.ins.2019.03.060.
- [4] L. X. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning," IEEE Transactions on Neural Networks, vol. 3, no. 5, pp. 807-814, 1992, doi: 10.1109/72.159070.
- [5] P. P. Angelov, X. Gu, and J. C. Principe, "Autonomous Learning Multimodel Systems From Data Streams," IEEE Transactions on Fuzzy Systems, vol. 26, no. 4, pp. 2213-2224, 2018, doi: 10.1109/TFUZZ.2017.2769039.
- [6] K. S. Narendra, J. Balakrishnan, and M. K. Ciliz, "Adaptation and learning using multiple models, switching, and tuning," IEEE Control Systems, vol. 15, no. 3, pp. 37-51, 1995, doi: 10.1109/37.387616.
- [7] P. Angelov, "Fuzzily Connected Multimodel Systems Evolving Autonomously From Data Streams," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 41, no. 4, pp. 898-910, 2011, doi: 10.1109/TSMCB.2010.2098866.
- [8] P. Angelov, X. Gu, and D. Kangin, "Empirical Data Analytics," International Journal of Intelligent Systems, vol. 32, no. 12, pp. 1261-1284, 2017, doi: 10.1002/int.21899.
- [9] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, no. 3, pp. 665-685, 1993, doi: 10.1109/21.256541.
- [10] P. Angelov and R. Yager, "A new type of simplified fuzzy rule-based system," International Journal of General Systems, vol. 41, no. 2, pp. 163-185, 2012, doi: 10.1080/03081079.2011.634807.
- [11] P. Angelov and X. Gu, Empirical Approach to Machine Learning. Switzerland: Springer Natural Switzerland, 2018.
- [12] P. P. Angelov, X. Gu, and J. C. Principe, "A Generalized Methodology for Data Analysis," IEEE Transactions on Cybernetics, vol. 48, no. 10, pp. 2981-2993, 2018, doi: 10.1109/TCYB.2017.2753880.
- [13] P. P. Angelov and D. P. Filev, "An approach to online identification of Takagi-Sugeno fuzzy models," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 34, no. 1, pp. 484-498, 2004, doi: 10.1109/TSMCB.2003.817053.
- [14] M. Kaul, R. Hill, and C. Walthall, "Artificial neural networks for corn and soybean yield prediction," Agric. Syst., vol. 85, no. 1, pp. 1-18, 2005, doi: 10.1016/j.agsy.2004.07.009.

- [15] J. R. Ambuel, T. S. Colvin, and D. L. Karlen, "A Fuzzy Logic Yield Simulator For Prescription Farming," *Transactions of the ASAE*, vol. 37, no. 6, pp. 1999-2009, 1994, doi: 10.13031/2013.28293.
- [16] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," *Computers and Electronics in Agriculture*, vol. 143, no. C, pp. 23-37, 2017, doi: 10.1016/j.compag.2017.09.037.
- [17] E. I. Papageorgiou, A. T. Markinos, and T. A. Gemtos, "Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture application," *Applied Soft Computing Journal*, vol. 11, no. 4, pp. 3643-3657, 2011, doi: 10.1016/j.asoc.2011.01.036.
- [18] J. Liu, C. E. Goering, and L. Tian, "A neural network for setting target corn yields," *Trans. ASAE*, vol. 44, no. 3, pp. 705-713, 2001, doi: 10.13031/2013.6097.
- [19] A. A.-G. Khalid et al., "Prediction of Potato Crop Yield Using Precision Agriculture Techniques," *PLoS ONE*, vol. 11, no. 9, p. e0162219, 2016, doi: 10.1371/journal.pone.0162219.
- [20] M. Khan and W. Khor, "A Framework for Fuzzy Rule-based Cognitive Maps," presented at the Pacific Rim International Conference on Artificial Intelligence, Berlin, Heidelberg, 2004, 454-463.
- [21] E. Beza *et al.*, "Review of yield gap explaining factors and opportunities for alternative data collection approaches." *European Journal of Agronomy*, vol. 82, pp: 206-222, 2016.
- [22] B. Gambin *et al.*, "Exploring genotype, management, and environmental variables influencing grain yield of late-sown maize in central Argentina." *Agricultural Systems*, vol. 146, pp. 11-19, 2016.
- [23] S. Mourtzinis *et al.* "Sifting and winnowing: Analysis of farmer field data for soybean in the US North-Central region." *Field crops research*, vol. 221 pp. 130-141, 2018.
- [24] T. Masuda and P.D. Goldsmith, "World Soybean Production: Area Harvested, Yield, and Long-Term Projections." *International Food and Agribusiness Management Review*, vol. 12, pp.143-162, 2009.
- [25] F. A. Merlos *et al.*, "Potential for crop production increase in Argentina through closure of existing yield gaps." *Field Crops Research*, vol. 184, pp.145-154, 2015.
- [26] FAOSTAT. Food and Agriculture Organization of the United Nations. <http://www.fao.org/faostat/en/#data/QC>
- [27] AACREA. Argentine Association of Consortia for Regional Agricultural Experimentation. <http://www.crea.org.ar>
- [28] MODIS. Moderate Resolution Imaging Spectroradiometer. <https://modis.gsfc.nasa.gov/data/dataproduct/mod13.php>
- [29] Adadi, Amina, and Mohammed Berrada. "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)." *IEEE Access*, vol. 6, 2018, pp. 52138–52160.