**Indonesian Journal of Electrical Engineering and Computer Science**
Vol. 37, No. 1, January 2025, pp. 648~659
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i1.pp648-659 ❒ 648

# MODIS-NDVI and wheat yield patterns and predictions in Taounate, Morocco

**Sara Bouskour[1], Mohamed Hicham Zaggaf[1], Lhoussain Bahatti[1], Soumaya Zayrit[2]**
[1]Department of Electrical Engineering, Ecole Normale Supérieure de l'Enseignement Technique (ENSET), Hassan II University, Mohammedia, Morocco
[2]Faculty of Science Ain Chock, University Hassan II-Casablanca, Casablanca, Morocco

**Article Info**

***Article history:***

Received Mar 11, 2024
Revised Sep 12, 2024
Accepted Sep 29, 2024

***Keywords:***

Estimate wheat yield
K-means
MODIS
NDVI
SVR
Time series

**ABSTRACT**

This study is devoted to the use of varied analytical methods to elucidate the complex relationship between normalized difference vegetation index (NDVI) and wheat production in Taounate, Morocco based on MODIS Satellite data. Linear regression (LR), with a coefficient of determination ($R^2$) of 0.93, provided a solid basis, while the decision tree (DT) showed significant performance with an $R^2$ of 0.81. Support vector regression (SVR) performed well with an $R^2$ of 0.96, highlighting its ability to capture the non-linear nuances of the data. Given the complexity inherent in the observed relationships, characterized by non-linear variations, we opted for a combined approach. K-means, closely linked to SVR, was integrated for its ability to identify homogeneous subgroups in the data ($R^2$ up to 0.98). This combination made it possible to circumvent the limits of strictly linear methods, thus reinforcing the robustness of our analysis. These results underline the capacity of the chosen methodology to decode the interactions between NDVI and wheat production in the complex context of Taounate. By providing clear and nuanced perspectives, this study helps inform agricultural decisions and build resilience to climate challenges in the region.



***Corresponding Author:***

Sara Bouskour
Department of Electrical Engineering, Ecole Normale Supérieure de l'Enseignement Technique (ENSET)
Hassan II University
Mohammedia, Morocco
Email: sara.bouskour@gmail.com

## 1. INTRODUCTION

Smart agriculture is emerging as an essential response to food security and climate change issues. At the heart of this innovative approach is the fusion of cutting-edge technologies and informed agricultural practices [1], aimed at optimizing productivity while mitigating environmental impacts. Our study follows the momentum of smart agriculture specifically in Morocco, where wheat cultivation is of particular importance. Faced with constantly evolving climate challenges, the need for adaptive strategies emerges as a priority, and it is in this perspective that our work is deployed [2].

Traditional methods of forecasting crop yields during the growing season can estimate yields before harvest and help make more informed management decisions. Although based on sound physiological and physical concepts, these methods prove ineffective in cases of spatial soil variability, stress, or different management practices. In contrast, plant cover remote sensing is highlighted as a potentially valuable tool for agricultural monitoring, owing to its synoptic coverage and its ability to observe in many spectral wavelengths. Many studies have recognized that plant development, stress, and yield capabilities are manifested in the spectral reflectance of plant canopies and can be quantified using vegetation spectral indices (NDVI, EVI, SAVI, ARVI, GCL, ReCI).

A study conducted by Shafiee *et al.* [3] examines the application of two machine learning-based regression methods to predict wheat grain yield from vegetation indices extracted from UAV images. The objective of this research was to evaluate the effectiveness of support vector regression (SVR) in combination with sequential forward selection (SFS) for grain yield prediction. Another study, by Ashourloo *et al.* [4], explored various remote sensing methods to improve crop yield forecasting. The main objective of this study is to determine the best remote sensing characteristics and the most efficient regression model for wheat yield prediction in the Hamedan region, Iran. To this end, we used several vegetation indices (VI) and reflectance values obtained from Sentinel-2 images, analyzed in different time windows, as input data for the regression models. Among the models tested, the gaussian regression model (GPR) gave the best results, with a root mean square error (RMSE) of 0.228 t/ha and a coefficient of determination $R^2$ of 0.73. A recent study by Ashfaq *et al.* [5] demonstrated that combining multiple datasets using three machine learning algorithms significantly improved yield prediction performance, with a coefficient of determination ranging from 0.74 to 0.88. Incorporating spatial information and other properties into the reference models improved prediction accuracy from 0.08 to 0.12.

Our analysis approach is based on the normalized difference vegetation index (NDVI) which is considered the indicator par excellence. This parameter, derived from satellite data, provides an objective measure of crop health and vigor, based on the normalized difference between reflected light in the red and near-infrared range [6]. Its relevance lies in its ability to quantify the vegetation presence and density, thereby providing detailed insight into crop performance [7].

In an agricultural context, the NDVI allows the crop evolution to be observed in a non-intrusive and continuous manner [8]. Its sensitivity to changes in plant cover makes it a preferred tool for assessing seasonal variations, water stress, and the impacts of climatic conditions on crop health [9], [10]. By choosing NDVI as the pivot of our study, we rely on a proven scientific method to dissect the complex relationship between vegetation and wheat production, providing rich and nuanced insights into the agricultural landscape of Taounate.

In this quest for in-depth understanding, we have mobilized a range of analytical tools. Linear regression (LR) outlined the relationships between NDVI and wheat production, while SVR delved into the intricacies of these links, capturing crucial non-linear nuances. The decision tree (DT) with its ability to unravel complex patterns enriched our analysis by revealing delicate relationships that are often difficult to identify. K-means, closely integrated with the SVR, partitioned the data, revealing homogeneous groups and thus consolidating our perception of different behaviors within the NDVI-wheat production continuum. This approach combination has shaped a holistic perspective, transcending the boundaries of traditional methods to address complexity in an integrated way.

At the heart of this work lies our scientific approach which ensures our commitment effectiveness to unveiling exploitable knowledge for making informed agricultural decisions, such as irrigation and fertilization [11]. By probing the dynamic relationship between NDVI and wheat production, we aim to offer nuanced perspectives, thereby contributing to strengthening the resilience of agricultural practices in the face of current climate challenges and ensuring food security in the Taounate region, Morocco.

This article is composed of different sections aimed at an examination of wheat production in Taounate. The first section is devoted to the precise delimitation of our studied region. The methodology sets out the *modus operandi*, from the collection of NDVI data to the study of correlation with wheat production. The results are presented graphically, followed by a discussion. The conclusion summarizes key findings and identifies potential directions for future research.

## 2. STUDY ZONE

Our study is carried out in the Taounate region, located in the north of Morocco, 80 km from Fez, the capital of the Fez-Meknes region. This province as highlighted by the red hexagon in Figure 1 extends over an area of 561,600 ha = 5,616 $km^2$, between the coordinates 34° 32′ 09″ north and 4° 38′ 24″ west. When it comes to climate, Taounate takes advantage of the Mediterranean weather. The latter is characterized by hot, dry summers and rainy winters, which creates ideal conditions for varying agricultural activities. At the same time, this region presents wide-ranging soil types and favorable conditions for diversifying crops. We find clay soils providing moisture reserves used during dry periods. There is also a type called loamy with an intermediate texture between clay and sand, thus promoting drainage and maintaining certain water retention. Finally, there is also a sandy soil creating rapid drainage conditions ideal for specific crops such as root vegetables, asparagus, and aromatic herbs.

A study was conducted by Belmahi *et al.* [12] in the Fez-Meknes region to study the relationship between cereal yield and NDVI data based on the Pearson coefficient and LR. This study reveals our choice of Taounate as a study area. The graph linked to this study proves that Taounate is the main producer of cereals (30%) in the Fez-Meknes Region.

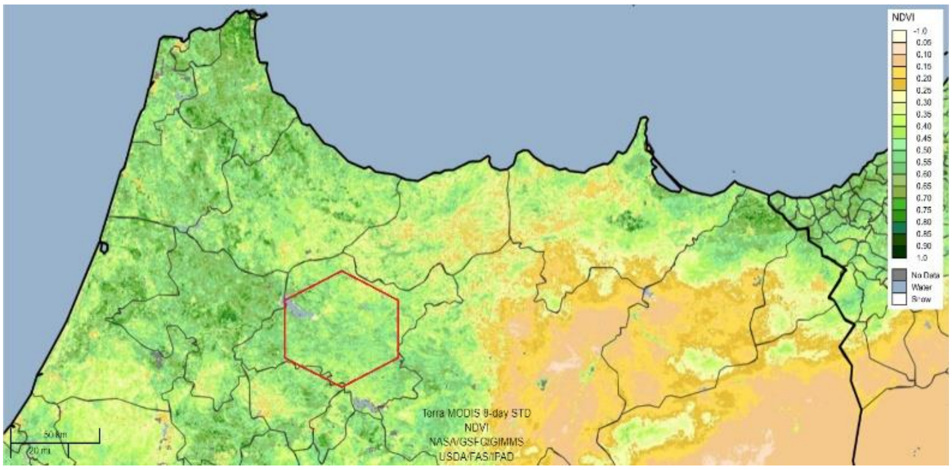


Figure 1. Location of the study area

## 3. METHOD

### 3.1. Remote sensing dataset

Our analysis of wheat crops in Taounate relies on the GIMMS global agricultural monitoring (GLAM) interface, a platform that provides access to NDVI images from MODIS and VIIRS satellites, as well as the retrieval of NDVI data time series [13]. The global inventory modeling and mapping studies/global agricultural monitoring (GIMMS/GLAM) system provides 8-day global datasets in near real-time. This data is derived from satellite images captured by sensors on NASA’s Terra spacecraft, at an average spatial resolution of 500 meters. Before being operational, this data undergoes atmospheric correction and cloud filtering, carried out pixel by pixel, thanks to the efforts of the MODIS scientific team.

The development and provision of this interface are ensured by the National Aeronautics and Space Administration/Goddard Space Flight Center/The Global Inventory Modeling and Mapping Studies (NASA/GSFC/GIMMS) group for the United States Department of Agriculture/Foreign Agricultural Service/International Production Assessment Division (USDA/FAS/IPAD) global agricultural monitoring project. The mission of this project is to provide an objective, timely, and regular assessment of global agricultural production prospects and conditions affecting global food security.

The choice of this data for our study stems from the need to access precise and near real-time information on crop health. The data’s scientific quality, coupled with atmospheric correction and cloud filtering, offers us a thorough view of the evolution of wheat crops in Taounate over time. During our NDVI data exploration in Taounate, a critically important observation emerged: on February 18, 2016, an interruption was noted in the data. To ensure continuity in our analysis, we took corrective measures by replacing the missing value with the data average just before and after this specific date. This approach was adopted to minimize the impact of the interruption on the time series consistency, thereby enabling a more reliable analysis of agricultural trends. Temporary data breaks are a common challenge in time series-based studies, and this correction helps maintain the integrity of our wheat crop health assessment in Taounate during the study period.

### 3.2. Production dataset

The production data used in our study represented in Figure 2 by the bar graph, emanate from the high commission for planning (HCP), the main provider of Morocco’s official statistics. Reliable and official source reinforces the credibility of our analysis, relying on information carefully collected by an institution having institutional and professional independence. Wheat production in Taounate, expressed in millions of quintals, experienced notable fluctuations from 2000 to 2020. In 2000, production reached 1423.1 million quintals, followed by a significant decline to 652.7 million quintals in 2001. In 2002, an increase was recorded with production of 1723.5 million quintals, but in 2003, production rose to 1757.2 million quintals, marking a year of growth. In 2004, production decreased to 978.5 million quintals, but in 2005, a sharp increase was observed with a production of 2321.6 million quintals.

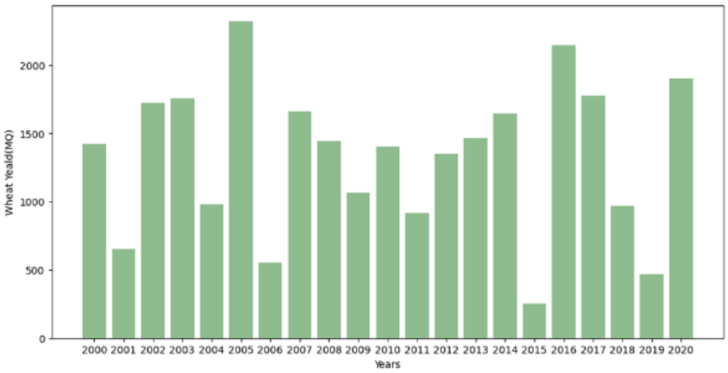


Figure 2. Annual wheat yield (2000–2020)

In 2006, production fell to 553.3 million quintals, only to experience a significant increase in 2007, reaching 1658 million quintals. The following years were marked by variations, with high points in 2009 (1063.3 million quintals) and 2016 (2146.5 million quintals). The year 2015 was characterized by exceptionally low production, recording only 251.8 million quintals. In 2017, production rebounded, reaching 1,776 million quintals. The years 2018 and 2019 saw a decrease in production, with respective figures of 969.5 and 470.2 million quintals. In 2020, production increased to 1904.4 million quintals. This data shows the influence of factors such as weather conditions, agricultural practices, and other environmental variables.

## 3.3. Dataset correlation study

The study of the correlation between NDVI values derived from MODIS images (every 8 days) and annual wheat production provides a decisive perspective to shed light on the relationship between vegetation and wheat grain yield. By checking the correlation between these measurements and wheat yield, we seek to determine the degree to which plant vigor, as captured by NDVI, impacts productivity. This method makes it possible to determine critical phases of the wheat growth cycle where the correlation provides very high values, hence providing information on the critical moments. A positive correlation value indicates that optimal plant conditions are linked to very high production, while a negative correlation value suggests adverse factors for growth.

By visualizing the correlation graph presented in Figure 3, we notice a significantly high correlation between the NDVI-MODIS values and the annual wheat production during the period (T13-T25) which corresponds to the period from January 17 to April 30. During this time-lapse, a sustained positive correlation between NDVI and wheat yield hence suggests that crop health determined by the vegetation index has a notable effect on wheat production.


Figure 3. Highlight of the correlation period between NDVI and wheat yield

The data choice for the regression is justified by the fact that this period (Jan 17-Apr 30) seems particularly sensitive to variations in NDVI, unlike the other moments. Thus, it seems reasonable to suggest that the wheat harvest is significantly impacted by environmental and agricultural conditions during these weeks. By focusing on this phase, the relationship between the vegetation index and wheat production can be captured in a more precise way, therefore leading to very reliable and specific predictions at this so-called critical temporal range of the crop wheat growth cycle.

### 3.4. K-means clustering with SVR modeling

In our study, we seek to implement a clustering (K-means) and SVR approach to agricultural data. The objective is to model the relationship between the average NDVI for each year (2000 to 2020) and the corresponding wheat production. The process starts by applying the K-means algorithm to group the years into k clusters, then obtaining the cluster centers μ (mean NDVI) for each cluster j, and finally obtaining the cluster assignment indices $c^{(i)}$ for each year *i*.

Mathematically, this could be expressed as follows:

$$J(c,\mu) = \sum_{i=1}^{m}\sum_{j=1}^{k}\left\|x^{(i)} - \mu\right\|^2 \cdot I\left(c^{(i)} = j\right) \tag{1}$$

such that:
$x^{(i)}$: the average of the NDVI for year *i*
μ: the center of cluster j
$c^{(i)}$: the index of the cluster to which year i is assigned
I (·): the indicator function

Then, for each cluster identified by K-means, an SVR model is trained. For each cluster j, we must extract the years i belonging to cluster j, then apply the SVR on this subset of data ($X_j$, $Y_j$), and finally obtain the weights $w_j$, the bias $b_j$, and the predictions $Y'_j$ for this cluster. Mathematically, this could be expressed as:

$$\min_{w_j, b_j, \xi_j \cdot \xi_j^*} \frac{1}{2} w_j^T \cdot w_j + C\sum_{i=1}^{m}\left(\xi_{j,i} + \xi_{j,i}^*\right) \tag{2}$$

for a specific cluster j [14]:

Under the constraints:

$$\begin{aligned} y_{j,i} - w_j^T \cdot x_{j,i} - b_j &\leq \varepsilon + \xi_{j,i} \\ w_j^T \cdot x_{j,i} + b_j - y_{j,i} &\leq \varepsilon + \xi_{j,i}^* \end{aligned} \tag{3}$$

where:
$x_{j,i}$: the average of NDVI for year i in cluster j
$y_{j,i}$: the wheat production for year i in cluster j
$\xi_{j,i}$ and $\xi^*_{j,i}$: the deviation variables
"C": the regularization parameter, and ε is a tolerable error parameter.

This represents a modeling approach where K-means is used to identify clusters of similar behavior in terms of mean NDVI, and then SVR is applied to each cluster to model the relationship between NDVI and wheat production within each group.

Our solution optimizes data management by using clustering to divide data according to its dispersion, followed by precise application of SVR. We have integrated two key methods: clustering and normalization: we start by grouping data into homogeneous clusters via K-means and then normalize the data to ensure a fair and balanced comparison. Hyperparameter search and cross-validation: for each cluster, we employ RandomizedSearchCV to explore the SVR hyperparameter space efficiently, adjusting the values of *C*, gamma, epsilon, and kernel. We use LeaveOneOut for small clusters to adjust the number of folds for larger clusters, ensuring proper cross-validation and avoiding overfitting. This approach guarantees robust predictions adapted to the specific characteristics of each cluster.

## 4. RESULTS

The data was split into two training and testing sets, with ratios of 80% and 20%, respectively. The distribution between training and testing parts significantly influences the result prediction accuracy. To ensure consistency between datasets of different sizes, we maintained the same ratio. Although the choice of this ratio depends on personal preference, opting for a higher proportion in the training set tends to improve the accuracy of the results. The ratio we opted for is the most recommended one [15].

As part of our study on wheat production in Taounate, we adopted a varied approach to analyze the data [16], taking into account in particular the NDVI. First, we explored linear trends using LR by assessing the correlation between NDVI and wheat production. Next, we applied SVR to account for potential non-linear relationships and also incorporated NDVI into our model. For in-depth understanding, we also used the DT algorithm to capture complex structures in the data, integrating NDVI as a key variable in the decision-making process. Subsequently, we refined our analysis by integrating the K-means algorithm in conjunction with the SVR, while continuing to consider the NDVI as a determining factor in the cluster formation.

Now, we are ready to present the results visually in graph form. These graphical representations including NDVI will allow us to clearly illustrate trends, clustering patterns, and relationships between variables, providing an in-depth visual understanding of the dynamics of wheat production in Taounate over time. This multidimensional approach will allow us to explore the specific contribution of NDVI to variations in wheat production.

## 4.1. Linear regression

LR allows you to build a model attempting to find a linear relationship between the independent variables (X) and the predicted dependent variable (Y). From a visual observation, we try to create a line so that the points are as close as possible, compared to other possible lines [4], [17]. The LR algorithm was applied to dates with a better correlation between MODIS-NDVI and wheat grain yield to be able to predict productivity based on the derived NDVI.

The regression line slope shown in Figure 4 indicates a positive relationship between NDVI values and wheat yield. In other words, as the NDVI value increases, wheat grain production tends to increase, this suggests a strong correlation between plant health, represented by NDVI, and agricultural production [18]. However, it is important to note that this correlation does not necessarily justify direct causation, since other factors can greatly influence production such as extreme weather factors. This is seen on the graph and is reflected by points a little away from the line.

The determination coefficient of 0.93 is a strong indication that 93% of the variability in production is mainly linked to changes in the NDVI value. This shows the ability of the model to predict wheat grain yield based on the NDVI value derived from the study area satellite images. The RMSE of 0.155 fortifies the regression results confirming the link between NDIV and wheat production.

## 4.2. The decision trees

DT have shown their superiority in modeling non-linear relationships. They perform successive divisions of the data based on the predictors, which allows them to create non-linear separation zones. Thus, if the dependent and the independent variable change in a non-linear manner, the tree can capture these changes, hence the use of this algorithm to model the non-linear relationship between the average NDVI and wheat production [19].

We observe that the predicted and observed value points are closely superimposed, this means that our model is pertinent. This is also confirmed by the $R^2$ of 0.80 which is generally considered a good fit (Figure 5). Plotting the graph (Figure 6) allowed us to optimally visualize the strong correlation between the predicted values and the collected values. We based ourselves on the peaks of the plot superposition. We count 14 peaks that overlap perfectly.

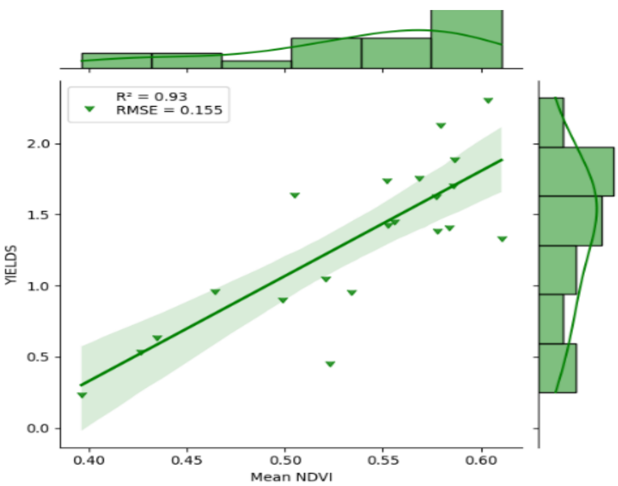


Figure 4. NDVI-wheat production LR (2000–2020)

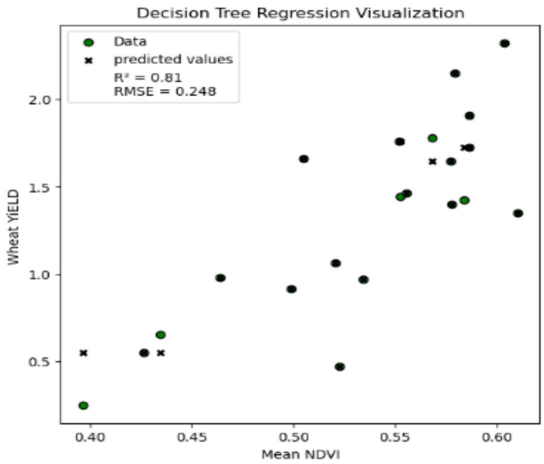


Figure 5. Wheat production DT representation

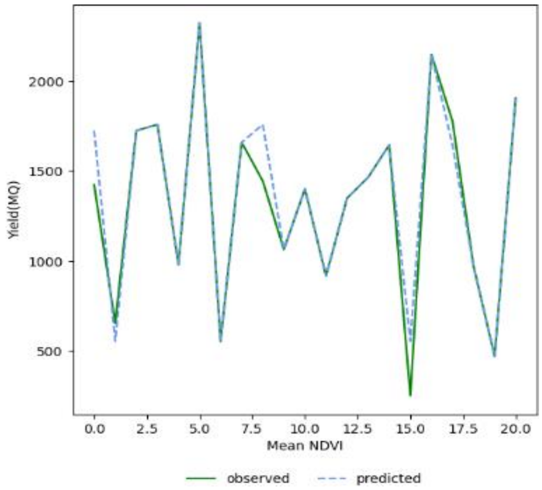


Figure 6. Predicted and observed wheat production

### 4.3. Vector regression support

Support vector machines (SVM) are machine learning algorithms widely used for classification [20], [21] and regression tasks [22]-[24]. Their popularity stems from their ability to find optimal decision boundaries in multidimensional spaces, making them particularly suited to complex problems in several fields: medical [25], and education [26]. SVR becomes a reliable choice when dealing with sparse datasets [27]. Its ability to model complex non-linear relationships, particularly through the use of kernels such as radial basis function (RBF), makes it possible to capture subtle patterns present in scattered data. Unlike some linear models, which can perform poorly in this situation, SVR is powerful to identify trends. Strategically positioned support vectors contribute to the robustness of the model, providing greater flexibility to adapt to changes inherent in sparse data. Therefore, the choice of SVR is justified due to its intrinsic ability to cope with the challenges of data dispersion by providing reliable and accurate results even in situations where other models may struggle.

Looking at the graph in Figure 7, we perceive the exact superposition of the support vector points with the observed data points. This phenomenon demonstrates the ability of the model to effectively identify and exploit the most critical points in the construction of the model. More remarkably, the points predicted by SVR show perfect proximity to the real data, even in the presence of significant dispersion. This behavior highlights the model's ability to generalize and strongly capture underlying patterns despite the inherent variability in the data. The high $R^2$ of 0.96 indicates that the model interprets the variance in the data well, thus strengthening our explanation.

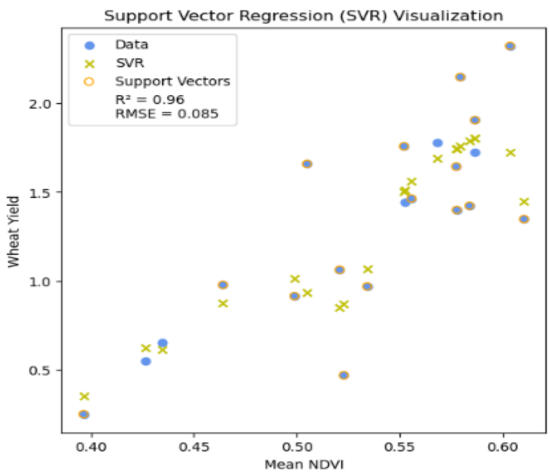


Figure 7. SVR graph

### 4.4. Implementation of K-means clustering

The transition from a model solely based on SVR to a model based jointly on K-means and SVR represents a strategic evolution in the modeling approach. Initially, SVR stands out for its ability to seize intricate relationships in data, providing accurate predictions even in contexts of dispersion. However, by introducing K-means, we aim to better understand the underlying structure of the data by identifying homogeneous clusters. Each then becomes the substrate for a dedicated SVR model, allowing finer adaptation. This cascade approach focuses on optimizing the accuracy and interpretability of the model by recognizing and processing local variations. Initially, we must visualize the data to define the need for clustering; here we use the same dataset, namely the annual average NDVI as the independent variable (X) and the wheat grain production as the variable target on which we intend to predict (Figure 8).

The data points do not follow a distribution that can be encapsulated in the hyperplane. So, using the K-means clustering algorithm is in order, but one must first determine the optimal cluster number to accurately retain the data points within the identified cluster number. To that end, we could use the Silhouette coefficient (Figure 9). If the latter is high, this implies a grouping of better quality; a score that reaches the value 1 reveals spaced and distinguished clusters while a score of 0 suggests an insignificant distance between clusters. When the score is -1, it indicates an incorrect cluster assignment.

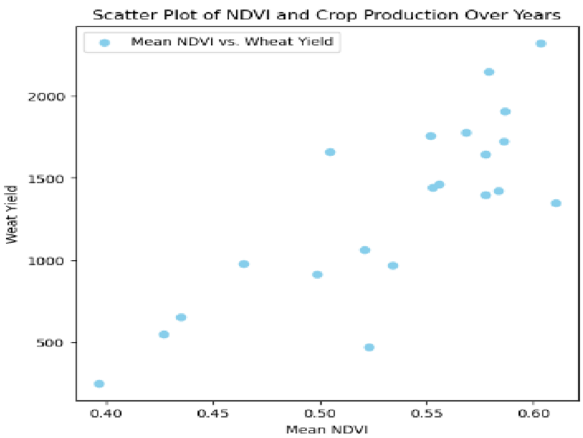


Figure 8. NDVI and wheat production relationship (2000–2020)

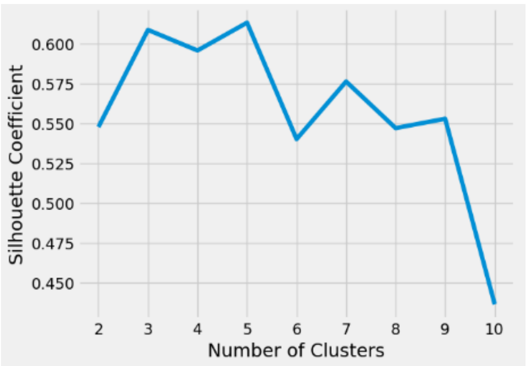


Figure 9. Silhouette coefficient scores of different cluster numbers

The Silhouette coefficient has a crucial impact on decision-making regarding the cluster number. After its evaluation, we found two promising options: either three or five clusters. To dispel this uncertainty between the two alternatives, it is essential to go through a silhouette analysis; this is the visual indicator allowing the coherence evaluation and the segmentation stability based on the silhouette outline thickness.

In Figure 10, we could see significant stability and clear separation between the clusters. As for Silhouette values, they are uniform and high, signifying better intracluster cohesion and intercluster distinction. Notwithstanding, in the five clusters case (Figure 11), the plot thickness displays a clear resemblance which could lead to over-segmentation and excessive complexity, since certain clusters could have a strong similarity with their peers.

### 4.5. SVR creation for clusters

The second step consists of developing SVM for regression (SVR) for each of the established clusters as seen in Figure 12. This makes it possible to implement an individualized approach that aims to construct regression models specific to each subgroup of data, in this way we will obtain a solution adapted to the complexity inherent in the segmented dataset subject of our study.

Creating separate SVR models for each cluster allows the optimization of overall predictive performance by capturing the nuances and variations specific to each subgroup as CL0 and CL2 has an $R^2$=0.98 and CL1 has an $R^2$=0.85. Thus, the choice of SVR turns out to be judicious, it excels in modeling non-linear relationships while minimizing the prediction error, hence the low RMSE for the three clusters.

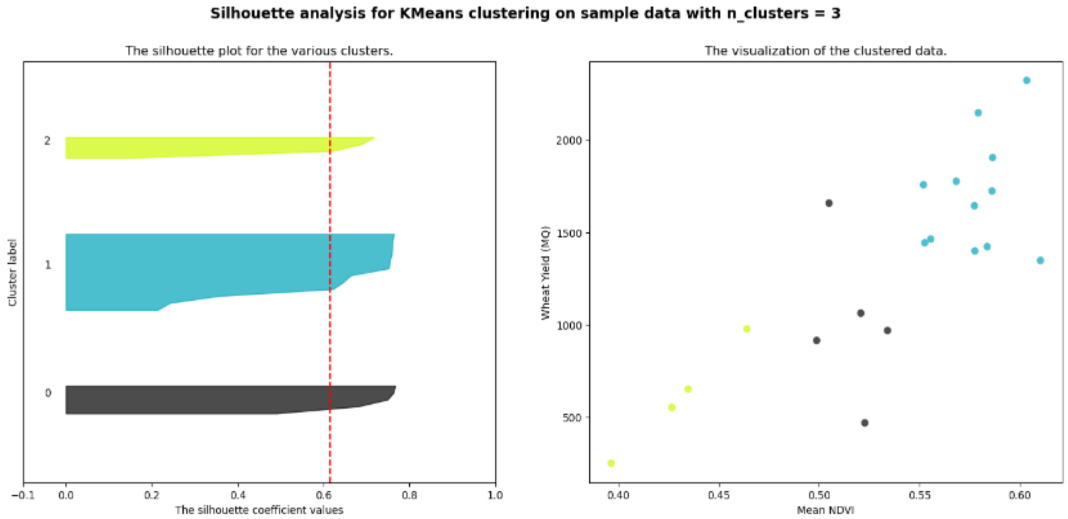


Figure 10. Silhouette analysis for K-means (k=3)

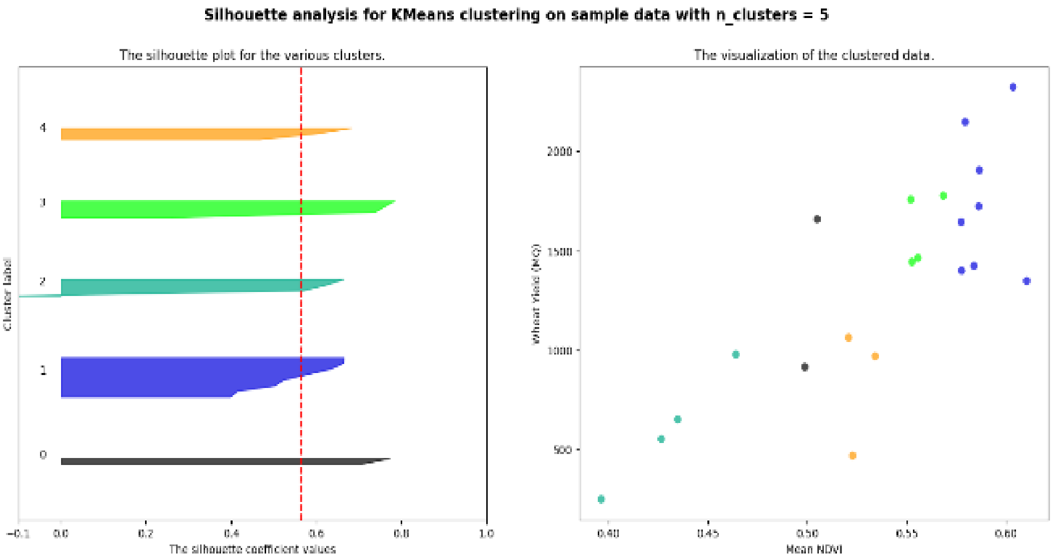


Figure 11. Silhouette analysis for K-means (k=5)

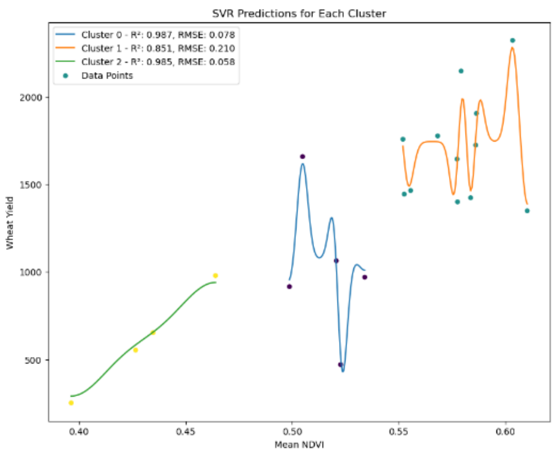


Figure 12. Combined K-means/SVR results graph

## 5. DISCUSSION

Our study exceeds the results of previous research in terms of the performance of ML algorithms, as shown in Table 1. While previous studies mainly used global approaches, we demonstrated that our algorithms, notably SVR, deliver results superiors. To further improve this performance, we combined K-means with SVR, which allowed us to better manage data heterogeneity and obtain more accurate yield forecasts.

Our results show a strong correlation between NDVI and wheat production. LR yielded an $R^2$ of 0.93 and an RMSE of 0.015, indicating a strong but limited match in capturing non-linear dynamics between variables. In contrast, SVR demonstrated increased ability to model these more complex relationships, with an $R^2$ of 0.96 and an RMSE of 0.085, outperforming LR in terms of accuracy and robustness. The DT model, although having an ability to handle non-linear relationships, presented lower performance with an $R^2$ of 0.81 and an RMSE of 0.247, which highlights its limitations in the face of the complexity of the data.

To improve the modeling of data heterogeneity, we segmented the dataset into three clusters using the K-means algorithm. This segmentation helped reveal underlying structures in the data, with a first cluster primarily composed of linear relationships, while the other two clusters represented more complex non-linear behaviors. For non-linear clusters, we applied SVR with a RBF kernel, which is better suited to capturing nonlinear relationships. This approach significantly improved the performance of the model, reaching an $R^2$ of 0.98 for clusters 0 and 2, and an $R^2$ of 0.85 for cluster 1.

These results confirm that the clustering approach followed by regression adapted makes it possible to better model the complexity of the interactions between NDVI and wheat production, compared to global methods. However, like any study, ours has limitations. Data interruptions in February 2016, although managed by the inclusion of averages of neighboring data, may have introduced bias. Additionally, the spatial and temporal variability of NDVI data and the complexity of environmental interactions may limit the generalizability of the results. The results obtained in this study pave the way for future improvements in agricultural yield modeling. A more refined approach, integrating unmanned aerial vehicle (UAV) imagery data combined with satellite imagery, could considerably improve the spatial resolution and accuracy of forecasts. Additionally, including other environmental and climatic variables in models could help capture the complex multifactorial interactions that influence agricultural yields. Another avenue would be to test deep learning approaches, such as convolutional neural networks (CNN), to improve the models' ability to handle data variability and complexities [28].

Table 1. Previous studies' synoptic table

| Author | Location | RS images | ML algorithms | $R^2$ | RMSE |
|---|---|---|---|---|---|
| Shafiee *et al.* [3] | Vollebekk research farm, at the NMBU, South-Eastern Norway | UAV | SVR | 0.90 | 0.479 |
| | | | LASSO | 0.90 | 0.500 |
| Ashourloo *et al.* [4] | Hamedan, Iran | Sentinel-2 | GPR | 0.73 | 0.228 |
| | | | RF | 0.71 | 0.237 |
| | | | SVR | 0.68 | 0.249 |
| | | | DT | 0.58 | 0.283 |
| Belmahi *et al.* [12] | Fez-Meknes region, Morocco | MODIS-NDVI | LR | [0.58-0.79] | [2.12-4.96] q/ha |
| Ashfaq *et al.* [5] | Punjab Province, Pakistan | Landsat 8 | SVM | 0.88 | 0.05 |
| | | | RF | 0.74 | 0.03 |
| | | | LASSO | 0.80 | 0.60 |
| Our present study | Taounate Province, Fez-Meknes region, Morocco | MODIS-NDVI | LR | 0.93 | 0.155 |
| | | | DT | 0.81 | 0.247 |
| | | | SVR | 0.96 | 0.085 |

## 6. CONCLUSION AND PERSPECTIVES

This study demonstrated the effectiveness of MODIS spectral data, and in particular the NDVI index, for forecasting wheat yields. The results revealed a significant relationship between NDVI and wheat production, validating its use as a reliable yield indicator. The application of various ML algorithms, including LR, DT, and SVR, has helped develop robust predictive models. The integration of clustering with K-means further made it possible to personalize forecasts by segment, thereby improving the accuracy of the estimates. A notable advancement of this research is the ability to make yield forecasts since the beginning of the growing season. By using NDVI at the well-defined period mentioned above, it is possible to provide an estimate of yields early enough to allow growers to anticipate challenges and implement solutions to optimize production. This approach not only makes it possible to better plan the necessary interventions, but also to adapt agricultural practices based on forecasts, thus increasing yields and the agricultural systems sustainability. In conclusion, this research highlights the importance of remote sensing technologies and advanced analytical techniques to improve agricultural management and build resilience to environmental variations.

In addition, an interesting perspective would be to couple these forecasting techniques with an IoT system. This could monitor nutrient levels and soil moisture in real-time, providing more accurate data. This integrated approach would allow a better understanding of the specific factors influencing yields and to more effective response to the identified problems, thereby improving crop management and the sustainability of agricultural systems.

## REFERENCES


[1] M. U. Ahmed and I. Hussain, "Prediction of wheat production using machine learning algorithms in northern areas of Pakistan," *Telecommunications Policy*, vol. 46, no. 6, p. 102370, Jul. 2022, doi: 10.1016/j.telpol.2022.102370.

[2] D. Almbaidin and W. Etaiwi, "Predicting wheat supply and demand: a systematic literature review of machine learning approaches," in *International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2023*, Jul. 2023, pp. 1–8, doi: 10.1109/ICECCME57830.2023.10252963.

[3] S. Shafiee, L. M. Lied, I. Burud, J. A. Dieseth, M. Alsheikh, and M. Lillemo, "Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery," *Computers and Electronics in Agriculture*, vol. 183, p. 106036, Apr. 2021, doi: 10.1016/j.compag.2021.106036.

[4] D. Ashourloo, M. Manafifard, M. Behifar, and M. Kohandel, "Wheat yield prediction based on Sentinel-2, regression, and machine learning models in Hamedan, Iran," *Scientia Iranica*, vol. 29, no. 6 D, pp. 3230–3243, Jun. 2022, doi: 10.24200/SCI.2022.57809.5429.

[5] M. Ashfaq, I. Khan, A. Alzahrani, M. U. Tariq, H. Khan, and A. Ghani, "Accurate wheat yield prediction using machine learning and climate-NDVI data fusion," *IEEE Access*, vol. 12, pp. 40947–40961, 2024, doi: 10.1109/ACCESS.2024.3376735.

[6] K. Choudhary, W. Shi, M. S. Boori, and S. Corgne, "Agriculture phenology monitoring using NDVI time series based on remote sensing satellites: a case study of Guangdong, China," *Optical Memory and Neural Networks (Information Optics)*, vol. 28, no. 3, pp. 204–214, Jul. 2019, doi: 10.3103/S1060992X19030093.

[7] S. Bouskour, L. Bahatti, and M. H. Zaggaf, "The use of NDVI to improve cereals agriculture: a review," in *2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET 2023*, May 2023, pp. 1–7, doi: 10.1109/IRASET57153.2023.10153054.

[8] T. Benabdelouahab, Y. Lebrini, A. Boudhar, R. Hadria, A. Htitiou, and H. Lionboui, "Monitoring spatial variability and trends of wheat grain yield over the main cereal regions in Morocco: a remote-based tool for planning and adjusting policies," *Geocarto International*, vol. 36, no. 20, pp. 2303–2322, Dec. 2021, doi: 10.1080/10106049.2019.1695960.

[9] C. Vaiphasa, S. Piamduaytham, T. Vaiphasa, and A. K. Skidmore, "A normalized difference vegetation index (NDVI) time-series of idle agriculture lands: A preliminary study," *Engineering Journal*, vol. 15, no. 1, pp. 9–16, Jan. 2011, doi: 10.4186/ej.2011.15.1.9.

[10] E. Ouharba, Z. E. Triqui, and R. Moussadek, "Climate change impact on cereal culture in Morocco. Case study: Rommani (Rabat Region), Bouregreg Watershed Centre," *ScienceLib*, vol. 7, no. 150211, pp. 1–21, 2015.

[11] S. Tkatek, S. Amassmir, A. Belmzoukia, and J. Abouchabaka, "Predictive fertilization models for potato crops using machine learning techniques in Moroccan Gharb region," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 5, pp. 5942–5950, Oct. 2023, doi: 10.11591/ijece.v13i5.pp5942-5950.

[12] M. Belmahi *et al.*, "Analysis of relationship between grain yield and NDVI from MODIS in the Fez-Meknes Region, Morocco," *Remote Sensing*, vol. 15, no. 11, p. 2707, May 2023, doi: 10.3390/rs15112707.

[13] NASA and USDA, "GIMMS global agricultural monitoring," *Accessed Online 16 Jun 22 [https://glam1.gsfc.nasa.gov/api/doc/about]*. https://glam1.gsfc.nasa.gov/.

[14] E. Cheng *et al.*, "Wheat yield estimation using remote sensing data based on machine learning approaches," *Frontiers in Plant Science*, vol. 13, Dec. 2022, doi: 10.3389/fpls.2022.1090970.

[15] R. Karim, M. K. Alam, and M. R. Hossain, "Stock market analysis using linear regression and decision tree regression," in *2021 1st International Conference on Emerging Smart Technologies and Applications, eSmarTA 2021*, Aug. 2021, pp. 1–6, doi: 10.1109/eSmarTA52612.2021.9515762.

[16] J. Han *et al.*, "Prediction of winter wheat yield based on multi-source data and machine learning in China," *Remote Sensing*, vol. 12, no. 2, p. 236, Jan. 2020, doi: 10.3390/rs12020236.

[17] H. S. E. Imanni, A. E. Harti, and L. E. Iysaouy, "Wheat yield estimation using remote sensing indices derived from sentinel-2 time series and google earth engine in a highly fragmented and heterogeneous agricultural region," *Agronomy*, vol. 12, no. 11, p. 2853, Nov. 2022, doi: 10.3390/agronomy12112853.

[18] A. Idrissi, S. Nadem, A. Boudhar, and T. Benabdelouahab, "Review of wheat yield estimating methods in Morocco," *African Journal on Land Policy and Geospatial Sciences*, vol. 5, no. 4, pp. 2657–2664, 2022, doi: 10.48346/imist.prsm/ajlp-gs.v5i4.33050.

[19] F. Maselli, M. Moriondo, L. Angeli, L. Fibbi, and M. Bindi, "Estimation of wheat production by the integration of MODIS and ground data," *International Journal of Remote Sensing*, vol. 32, no. 4, pp. 1105–1123, Feb. 2011, doi: 10.1080/01431160903510799.

[20] L. Bahatti, O. Bouattane, M. E. H. Ech-chhibat, and M. H. Zaggaf, "An efficient audio classification approach based on support vector machines," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, 2016, doi: 10.14569/ijacsa.2016.070530.

[21] H. Ibrahim, S. A. Anwar, and M. I. Ahmad, "Classification of imbalanced data using support vector machine and rough set theory: A review," *Journal of Physics: Conference Series*, vol. 1878, no. 1, p. 012054, May 2021, doi: 10.1088/1742-6596/1878/1/012054.

[22] A. Fentis, L. Bahatti, M. Tabaa, B. Chouri, and M. Mestari, "Data driven model for short term PV power forecasting using least square support vector regression," in *Proceedings of 2017 International Renewable and Sustainable Energy Conference, IRSEC 2017*, Dec. 2018, pp. 1–6, doi: 10.1109/IRSEC.2017.8477401.

[23] M. Panahi, N. Sadhasivam, H. R. Pourghasemi, F. Rezaie, and S. Lee, "Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR)," *Journal of Hydrology*, vol. 588, p. 125033, Sep. 2020, doi: 10.1016/j.jhydrol.2020.125033.

[24] M. S. Ahmad, S. M. Adnan, S. Zaidi, and P. Bhargava, "A novel support vector regression (SVR) model for the prediction of splice strength of the unconfined beam specimens," *Construction and Building Materials*, vol. 248, p. 118475, Jul. 2020, doi: 10.1016/j.conbuildmat.2020.118475.

[25] N. M. Ali, N. A. A. Aziz, and R. Besar, "Comparison of microarray breast cancer classification using support vector machine and logistic regression with LASSO and boruta feature selection," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 2, pp. 712–719, Nov. 2020, doi: 10.11591/ijeecs.v20.i2.pp712-719.
[26] O. Chamorro-Atalaya *et al.*, "Quadratic vector support machine algorithm, applied to prediction of university student satisfaction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 139–148, Jul. 2022, doi: 10.11591/ijeecs.v27.i1.pp139-148.
[27] K. H. Suradiradja, I. S. Sitanggang, L. Abdullah, and I. Hermadi, "Estimation of biomass of forage sorghum (sorghum bicolor) Cv. Samurai-2 using support vector regression," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 3, pp. 1786–1794, Jun. 2023, doi: 10.11591/ijeecs.v30.i3.pp1786-1794.
[28] S. Bouskour, M. H. Zaggaf, and L. Bahatti, "Deep learning recognition of wheat leaf disease using MobileNetV2 model," in *2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET 2024*, May 2024, pp. 1–6, doi: 10.1109/IRASET60544.2024.10548207.

## BIOGRAPHIES OF AUTHORS

**Sara Bouskour** is a Ph.D. student at the Electrical Engineering and Intelligent Systems Laboratory, ENSET Mohammedia, specializing in smart agriculture. Holding a Master's degree in Computer Science, her research focuses on leveraging intelligent systems and data analytics to enhance efficiency in agriculture. Passionate about interdisciplinary collaboration, she actively contributes to conferences and workshops, sharing her insights on integrating technology into sustainable agricultural practices. As a dedicated Ph.D. student, she aspires to shape the future of Smart Agriculture. For any professional inquiries, contact Sara Bouskour through the Electrical Engineering and Intelligent Systems Laboratory, ENSET Mohammedia. She can be contacted at email: sara.bouskour@gmail.com.

**Mohamed Hicham Zaggaf** received the Ph.D. degree in electric engineering from the Hassan II University of Casablanca, Morocco, in 2019. Currently, he is an Assistant Professor at the Normal School of Technical Education (ENSET), Hassan II University of Casablanca, Morocco. His research interests include optimization and nonlinear control of AC machines and energy. He can be contacted at email: zaggaf@enset-media.ac.ma.

**Lhoussain Bahatti** received the Aggregation Electr. Engin. degree in 1995 from the ENS CACHAN France and the DEA diploma in information processing in 1997. In 2013, he had his Ph.D. degree in Signal and information processing from the Faculty of Science and Technology of Mohammedia (FSTM) and qualified to Direct Research in 2016. Published more than 60 research publications in various, national and international conference proceedings and Journals on signal and image processing, machine learning, classification, and control, also has experience in teaching it since 1995. He was the coordinator of the "Electrical Engineering and Industrial Systems Control" section at ENSET (2015–2018) and head of the "Parallel Architectures, Image and Signal Processing" research team (APTSI) at "Disturbed Systems Signals and Artificial Intelligence" laboratory (SSDIA) (2016–2020). Since 2018, Head of Electrical Engineering at ENSET and member of the" Electrical Engineering and Intelligent Systems" (IESI) laboratory and head of the "Signals, Images, and Intelligent Systems" (SISI) team since 2020. His research is also focused on the control of renewable energy production systems and smart agriculture. He can be contacted at email: bahatti.enset@gmail.com.

**Zayrit Soumaya** received the Master degree in Electronics, Electrotechnics, Automatic, and Industrial Computing from Faculty of Science Ain Chock. University Hassan II-Casablanca, in 2023 she obtained a Ph.D. in the Research Laboratory in Industrial Engineering, Information Processing and Logistics (GITIL). Faculty of Science Ain Chock, University Hassan II-Casablanca, Morocco. Her interests are in speech processing for detecting people with neurological disorders and AI methods. She can be contacted at email: zayritsoumaya@gmail.com.