Development of a Geospatial Predictive System of Crop Yield in Vineyards - A Case Study in Spain

Juan J. Cubillas
Dept. Information and Communication Technologies applied to Education.
International University of La Rioja
Logroño, Spain
e-mail: {juanjose.cubillas}@unir.net
Francisco Feito, Juan M. Jurado,, David Jurado, J.Roberto Jiménez,, Lidia Ortega
Dept. Computer Science.
University of Jaen
Jaen, Spain
e-mail: {ffeito|jjurado|djurado|rjimenez|lidia}@ujaen.es
Carlos Enríquez, Antonio Garrido, M.Isabel Ramos
Dept. Cartographic, Geodetic and Photogrammetric Engineering.
University of Jaen
Jaen, Spain
e-mail: {cenrique|agarrido|miramos}@ujaen.es

***Abstract*—This project aims to develop an Artificial Intelligence (AI)-based system for early crop yield prediction in vineyards. The objective is to provide farmers with a reliable tool that allows them to optimize resource planning, reduce risks, and enhance crop sustainability. The methodology integrates multi-source and multi-scale data, including historical yield information, multispectral satellite images, and climatic variables, such as temperature, humidity and precipitation, obtained from MODIS and ERA5, from Copernicus services. It employs advanced AI techniques, such as image processing and regression models. A key phase is validating and adjusting the model using high-resolution data captured by drones. The expected impact is outstanding accuracy in harvest prediction, which will lead to a significant reduction in uncertainty, greater operational efficiency, and improved grape quality, transforming viticulture into a more predictive and sustainable discipline.**

Keywords*-*Artificial intelligence; agriculture; crop yield prediction; remote sensing.

I. INTRODUCTION

The early estimation of crop yields for a specific crop is essential for all actors involved, including farmers, intermediaries, insurance companies, administrations and, of course, the consumer himself. Since time immemorial, good or bad harvests have brought both prosperity and famine to populations and thus determined their livelihoods and subsistence. Today, they still generate major imbalances in the economies of many families and areas of the planet, mainly because there are still no effective tools to make accurate forecasts sufficiently in advance. In this field, the most significant advances are determined by Information and Communication Technologies (ICT) at the service of Precision Agriculture (PA). This field also includes Remote Sensing for capturing images of the terrain and their advanced processing using Machine Learning techniques to forecast possible problems, such as diseases, and above all those related to crop yields [1] and [2].

It is the focus of most of the scientific community's efforts to try to identify the variables that mainly determine the behaviour of harvests. Undoubtedly, one of the most determining factors is climate [3], [4]. Although the wine sector has a somewhat more stable production than other traditional crops, such as olives, weather conditions are the main reason for the variability between the harvests of 2013 (7,500 tonnes) and 2017 (5,400 tonnes) at regional level [5]. Another important aspect related to climate is the quality of the grapes and, therefore, of the resulting wines [6], .

In order to be able to determine future behaviour, it is usually necessary to know what happened in the past. In the case of crop prediction, it is important to make this correlation between climatological variables and harvest results. The use of satellite data offers great advantages for working with medium and large-scale territories, such as municipalities, provinces or other types of geographical demarcations. However, their greatest capacity is to provide data with a certain frequency, providing historical data [7], [8]. Although they do not provide the same resolution as sensors attached to drones, they can cover large areas of land and provide data from the past that can also be correlated with data from previous harvests. In addition, different satellites provide images of different types: optical, multispectral, hyperspectral, thermal or LiDAR (Light Detection and Ranging), which are widely used in precision agriculture.

Most of the works developed for harvest forecasting differ in methodology depending on the type of crop. The importance of its forecasting in the field of wine production is pointed out in some works to determine the desired quantity and quality of grapes, which is crucial for winemakers [9], [10], [11]. However, the methodology of data capture, data cleaning and pre-processing can be considered a common task. Although each crop needs to adjust a different model based on its specific characteristics, a common methodology can be established for many crop types. In each case, the importance of data collection at different times of the year both at the climatological level and using specific vegetation indices for each case is considered.

Crop yield prediction is definitely one of the challenging problems in precision agriculture; however, as Xu et al. [12] point out, it is not a trivial task. Nowadays, crop yield prediction models can reasonably estimate actual values, but better performance in yield prediction is still desirable [6]. Numerous authors have emphasised the importance of quantitative crop yield prediction for years, considering it as a valuable tool to support farmers [13]. The close relationships between pollen emission and fruit production are extensively studied in this research. However, final fruit production is influenced by various climatic and agronomic conditions both in the pre-flowering period and in the period between flowering and harvest, such as water deficit, temperature extremes and phytopathological problems.

The structure of the paper has 4 sections: Section I is the Introduction where the crucial importance of early yield estimation in vineyards is highlighted for all actors in the sector. Section II, Methodology, proposes the implementation of a geospatial vineyard yield prediction system using AI and remote sensing, by integrating multi-source and multi-scale data. Section III describes the expected results and, finally, Section IV presents the incipient conclusions of this work.

II. Methodology

The implementation of a vintage prediction system for vineyards using AI and remote sensing involves the integration of multi-source and multi-scale data, the design of a geospatial database in the cloud and the creation of a predictive model validated with field data. Success lies in efficient data management and analysis, accuracy of predictions and accessibility for winegrowers, as shown in 1 . A phased implementation is proposed.

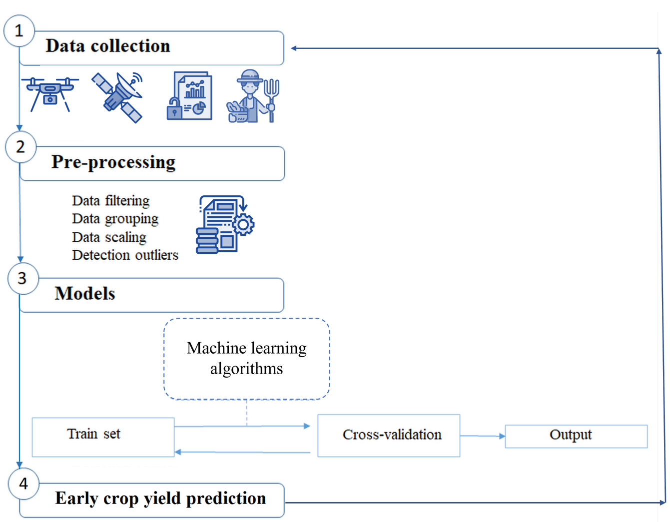

Figure 1. Methodology workflow.

A. System architecture planning and design

The first step will be to design a system architecture that allows data to be managed, processed and analysed in an efficient and scalable way. Three main sources of data will be considered:

1) Public data:: Satellite imagery providing multispectral information on vine cultivation.

2) Project-specific data:: High-resolution images, both satellite and captured from drone-mounted sensors.

3) Meteorological data:: Real-time weather information from local stations and historical bases, as well as products derived from remote sensing.

A geospatial database will be designed to efficiently store and manage geolocated information, and a cloud infrastructure will be implemented to ensure remote access, scalability and data security.

B. Data acquisition and processing

This phase includes the collection of the multi-source data and the processing of the data. The different origin and nature of the data requires a specific treatment of the data, both to be integrated homogeneously in the database without affecting the coherence of the data and to generate the derived products necessary for the implementation of the predictive model itself.

C. Implementation of the geospatial database

The geospatial database will be used to store and manage spatial data, allowing complex queries based on vineyard locations and associated variables. In addition, geospatial visualisation tools will be integrated to provide users with a visual representation of the data and to facilitate the interpretation of the information. Furthermore, being cloud-based, the database will be scalable, allowing new datasets to be incorporated as more data is obtained, without affecting the performance of the system. The cloud will also facilitate collaboration by allowing multiple users to access the system from different locations, which is essential when working with a technology transfer project involving multiple stakeholders.

D. Design and implementation of the predictive system in the cloud

The next step is the design and implementation of the predictive system in the cloud. This system will use advanced Machine Learning (ML) techniques capable of integrating diverse data sources and learning complex patterns that allow early estimation of the harvest. Once the model is trained, it will be implemented on Oracle's cloud platform. This cloud platform should also be accessible from mobile devices, facilitating remote access for users, so that it can also serve as a means of capturing data on harvest quantity (in the first instance) and other information on farming practices to feed back and retrain the predictive system.

E. Design and development of graphical interface for system use

Using Oracle Application Express, a system will be developed that will allow authorised users to visualise the harvest prediction and allow them to analyse the actual harvest and prediction data. This will allow non-expert users and from home to access and use the machine learning models, allowing to interpret and apply predictions in an intuitive and efficient way.

F. Validation and adjustment of the model with drone data

A fundamental part of the implementation of the system is the validation of the predictions generated by the predictive model. For this purpose, data collected directly with drones in the vineyards will be used as a reference point to verify the accuracy of the system's predictions from satellite images. The drone data, due to its high resolution and ability to capture fine details of the vineyard, will allow validation of the harvest predictions and adjustment of the model as needed. In order to ensured statistically robust validation it shall be adopted a sufficient number of sampling points covering a representative range of conditions within the study vineyards. Also, the timing of data collection will be directly related to the vegetative cycle of the vineyard.

This validation process is iterative and will progressively improve the accuracy of the system as more drone data is collected and more experience is gained with the system.

G. Scalability and maintenance of the system

Once the predictive system has been validated and fine-tuned, the focus will be on ensuring its long-term scalability and maintainability. As technology and data will continue to evolve, the system must be flexible and able to adapt to new data sources and predictive algorithms. The cloud platform must have tools that allow for continuous updating of the model, incorporation of new data, and enhancement of the system without interrupting service to users. This also includes the implementation of a monitoring system to ensure optimal performance of the infrastructure, detect possible errors and ensure the accuracy of the system.

H. Knowledge transfer and training

Training programmes will be designed to teach winegrowers how to use the platform, interpret forecasts and make informed harvesting decisions. This training will be crucial to ensure technology adoption and maximise the impact of the system on improving productivity in the vineyards.

III. Preliminary Results and Expectations

In a machine learning study focused on early grape harvest prediction, results are anticipated that will transform vineyard management. The primary goal is to achieve outstanding accuracy in harvest date prediction, minimising the discrepancy between model estimate and reality. Regression algorithms, trained on historical data, climatological data and multispectral images, are expected to reveal complex and non-linear patterns, overcoming the limitations of traditional methods. This accuracy would translate into more efficient harvest planning, allowing growers to optimise resource allocation and coordinate labour in advance.

The model is expected to reveal the relative importance of the variables analysed, from climatic fluctuations to vegetation indices captured by satellites and drones. This information will allow winegrowers to better understand the influence of various factors on their vineyards, adapting to the particularities of each vintage and mitigating the effects of climate change.

Rigorous validation of the model is crucial to ensure its robustness and applicability in different scenarios. The integration of drone data, with its high spatial resolution, is expected to complement satellite information, refining predictions and allowing accurate assessment at the plot scale. In terms of metrics, high R^2 values, close to 1, and low RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) values are aspired, reflecting the high accuracy and low error of the predictions.

IV. Conclusions

The implementation of machine learning models for early grape harvest prediction represents a significant advance in precision viticulture. The expected results, based on the integration of multi-source data and regression algorithms, promise not only to improve the accuracy of predictions, but also to deepen our understanding of the factors influencing grapevine phenology. The ability to accurately anticipate harvest yields months in advance will allow grape growers to optimise the planning of their activities, from resource allocation to grape quality management. In addition, the identification of the most influential variables, such as climatic conditions and vegetation indices, will provide valuable information for informed decision-making.

Ultimately, this approach has the potential to transform viticulture into a more predictive and sustainable discipline. Rigorous validation of the models, using high-resolution drone and satellite data, will ensure their robustness and applicability in different contexts. Quantification of model performance through metrics, such as R^2, RMSE and MAE will provide an objective basis for assessing their accuracy and reliability. The implementation of these models is expected to lead to a significant reduction of uncertainty in wine crop management, resulting in increased efficiency and improved grape quality. In addition, the ability to capture and analyse complex patterns in the data will allow researchers and viticulturists to gain new insights into grapevine physiology and its response to environmental conditions.

References

[1] N. Bali and A. Singla, "Emerging Trends in Machine Learning to Predict Crop Yield and Study Its Influential Factors: A Survey," *Archives of Computational Methods in Engineering*, vol. 29, no. 1, pp. 95–112, Jan. 2022, ISSN: 1886-1784. DOI: 10.1007/s11831-021-09569-8.

[2] P. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," *Computers and Electronics in Agriculture*, vol. 163, p. 104 859, Aug. 2019, ISSN: 0168-1699. DOI: 10.1016/j.compag.2019.104859.

[3] S. Quiroga and A. Iglesias, "A comparison of the climate risks of cereal, citrus, grapevine and olive production in Spain," en, *Agricultural Systems*, vol. 101, no. 1, pp. 91–100, Jun. 2009, ISSN: 0308-521X. DOI: 10.1016/j.agsy.2009.03. 006.

[4] A. Barriguinha, B. Jardim, M. de Castro Neto, and A. Gil, "Using NDVI, climate data and machine learning to estimate yield in the Douro wine region," *International Journal of Applied Earth Observation and Geoinformation*, vol. 114, 2022, DOI: 10.1016/j.jag.2022.103069.

[5] S. Miovska, C. M. Bande, and N. Stojkovik, "Predicting Wine Properties Based on Weather Conditions Using Machine Learning Techniques," in *2024 47th ICT and Electronics Convention, MIPRO 2024 - Proceedings*, 2024, pp. 140–145. DOI: 10.1109/MIPRO60963.2024. 10569756.

[6] G. Canavera *et al.*, "A sensorless, Big Data based approach for phenology and meteorological drought forecasting in vineyards," *Scientific Reports*, vol. 13, no. 1, 2023. DOI: 10.1038/s41598-023-44019-4.

[7] M. F. Aslan, K. Sabanci, and B. Aslan, "Artificial Intelligence Techniques in Crop Yield Estimation Based on Sentinel-2 Data: A Comprehensive Survey," *Sustainability (Switzer-land)*, vol. 16, no. 18, 2024. DOI: 10.3390/su16188277.

[8] S. Pancholi and A. Kumar, "Investigating the Capability of DOVE Satellite Temporal Data for Mapping Harvest Dates of Sugarcane Crop Types Using Fuzzy Model," *Journal of the Indian Society of Remote Sensing*, vol. 52, no. 10, pp. 2127–2142, 2024, Type: DOI: 10.1007/s12524-024-01927-w.

[9] F. Palacios, M. P. Diago, P. Melo-Pinto, and J. Tardaguila, "Early yield prediction in different grapevine varieties using computer vision and machine learning," *Precision Agriculture*, vol. 24, no. 2, pp. 407–435, 2023. DOI: 10.1007/s11119-022-09950-y.

[10] C. Laurent *et al.*, "A review of the issues, methods and perspectives for yield estimation, prediction and forecasting in viticulture," *European Journal of Agronomy*, vol. 130, p. 126 339, 2021, ISSN: 1161-0301. DOI: https://doi.org/10.1016/j.eja.2021.126339.

[11] J. A. Taylor, B. Tisseyre, and C. Leroux, "A simple index to determine if within-field spatial production variation exhibits potential management effects: Application in vineyards using yield monitor data," *Precision Agriculture*, vol. 20, no. 5, pp. 880–895, Oct. 2019, ISSN: 1573-1618. DOI: 10.1007/s11119-018-9620-3.

[12] X. Xu *et al.*, "Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China," en, *Ecological Indicators*, vol. 101, pp. 943–953, Jun. 2019, ISSN: 1470-160X. DOI: 10.1016/j.ecolind.2019.01.059.

[13] O. Fabio *et al.*, "Yield modelling in a Mediterranean species utilizing cause–effect relationships between temperature forcing and biological processes," en, *Scientia Horticulturae*, vol. 123, no. 3, pp. 412–417, Jan. 2010, Number: 3, ISSN: 0304-4238. DOI: 10.1016/j.scienta.2009.09.015.

 ISBN: 978-1-68558-269-2