

# Data Driven Predictive Models Based on Artificial Intelligence to Anticipate the Presence of *Plasmopara Viticola* and *Uncinula Necator* in Southern European Winegrowing Regions

Marta OTERO <sup>a,1</sup>, Luisa Fernanda VELASQUEZ <sup>a</sup>, Boris BASILE <sup>b</sup>, Jordi Ricard ONRUBIA <sup>a</sup>, Alex Josep PUJOL <sup>a</sup>, Josep PIJUAN <sup>a</sup>

<sup>a</sup> *Eurecat, Centre Tecnològic de Catalunya, Unit of Applied Artificial Intelligence, Science and Technology Park of Lleida, Building H3, 25003 Lleida, Spain*

<sup>b</sup> *Department of Agricultural Sciences, University of Naples Federico II, 80055 Naples, Italy*

**Abstract.** Downy and powdery mildews are two of the main diseases threatening grapevine cultivation worldwide caused by the phytopathogens *Plasmopara viticola* and *Uncinula necator*, respectively. These diseases may cause severe damage to grapevines by inducing wilting of plant organs, including bunches, especially when vines are untreated. This fact, together with the widespread of these pathogens due to the large extensions of land dedicated to grapevine monoculture, makes necessary to develop new predictive modeling tools that allow anticipating disease appearance in the vineyard, minimizing the losses in fruit yield and quality, and helping farmers in defining appropriate and more sustainable disease management strategies (fungicides applied at the right time and dose). For this purpose, farms located in three countries (Portugal, Spain, and Italy) were selected to study the relationship between the microclimatic characteristics of the plots, the phenological stage of the plants throughout the annual cycle, and the presence of both pathogens using different Machine and Deep Learning classification algorithms: Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, K-Nearest Neighbors, Naïve Bayes, Support Vector Machines, and Deep Neural Networks. The results showed that, after an entire annual grapevine cycle, the best performing models were Support Vector Machines for downy mildew and Random Forest for powdery mildew, providing a prediction accuracy of more than 90% for the infection risk and more than 80% for the treatment recommendation. These models will be fine-tuned during two additional vegetative seasons to ensure their robustness and will receive short- and medium-term climatological and phenological forecasts to make recommendations. The preliminary results obtained show that these models are a promising tool in the field of plant disease prevention and resource saving.

**Keywords.** disease anticipation, resource saving, fungicide management, prevent economic losses, Machine learning.

## 1. Introduction

Europe has always had a strong wine culture for thousands of years, making many of its viticultural regions world leaders in wine or fresh fruit production. Although new countries have been joining vine production for self-consumption and export, the

---

<sup>1</sup> Corresponding author: Marta Otero, Unit of Applied Artificial Intelligence, Eurecat, Centre Tecnològic de Catalunya, Science and Technology Park of Lleida, Building H3, Lleida, Spain. E-mail: marta.otero@eurecat.org

European Union continues to be the pioneer in world production, reaching figures of more than 60% [1]. France, Italy, and Spain are among the countries with the highest wine production in the European Union. Grapevines (*Vitis vinifera* L.) are usually located in slopy areas or on flat terrains and typically grow under Mediterranean climatic conditions. Due to the characteristics of the farms, monoculture is usually encouraged, and often a single variety is grown. This cultivation model favors the appearance of diseases and can cause catastrophic consequences in terms of production [2]. Two of the main diseases affecting grapevine production are downy and powdery mildews caused by *Plasmopara viticola* and *Uncinula necator*, respectively. In recent years, several authors have tried to approach plant disease detection with Artificial Intelligence [3,4] obtaining promising results from different perspectives. However, these studies were limited to very specific climatic zones and cultivars without taking into account neither the evolution of the disease in different grapevine varieties and climatic conditions nor the interaction between fungal development and grapevine phenology. Therefore, the objective of the present work was to monitor the epidemiology and management of *P. viticola* and *U. necator* in different growing areas and grapevine varieties to understand and predict the behavior of both phytopathogens. This will allow anticipating their appearance and designing an efficient management protocol of the vineyards, reducing the use of resources such as water and fungicides by applying the treatments at an appropriate dose and date and preventing economic losses.

## 2. Materials and Methods

### 2.1. Selection of sampling plots, data collection and preprocessing

The study areas are centered in the Mediterranean and Atlantic region of vine production in the European Union: Quinta do Ataíde (Portugal), L'Aranyó (Spain), and Mirabella Eclano (Italy). In each of the experimental plots the following disease management treatments were compared: i) an unsprayed control, ii) a treatment where to apply the recommendations generated by the models and iii) a treatment where diseases were managed conventionally. Each treatment consisted of a total of 20 plants divided into 4 blocks. Disease monitoring during the first growing season was conducted during the months of March-October 2021, making field visits every 7 days. On each measuring date, a total of 100 leaves per treatment were visually inspected to measure the percentage of affected leaves and the percentage of affected leaf blade. With respect to the treatments applied in the field to control the diseases, the historical data of product applications and doses supplied during 2021 was collected. Depending on the vine phenological stage, an adjustment was made to the predictions in terms of the volume of mixture applied in the field based on previous knowledge, optimizing the use of water used to apply treatments.

The independent features used in the model were climatology and the evolution of phenology throughout the year. The former comprises daily data obtained from *in situ* sensors that measure the microclimatic plot conditions, while the latter is based on in-field crop observations by the farmers. Climatological differences between plots were determined by Repeated Measures Analysis of Variance using the Least Significance Difference as a post hoc test. Finally, different feature engineering techniques were used to obtain the best possible data quality. In addition, quantitative disease occurrence data were translated into qualitative data to provide low-, medium- and high-risk alerts and propose treatment dosages.

## 2.2. Initial model selection and effectiveness evaluation

Once the data were ready to be introduced into a model, eight different classification Machine and Deep Learning models (Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, K-Nearest Neighbors, Naïve Bayes, Support Vector Machines and Deep Neural Networks) were studied using the scikit-learn library [5] in Python3. For this purpose, a battery of different hyperparameters was prepared for each of the models, and class compensation by stratification was performed during the separation of the train and test datasets. All models were trained with a 10-fold cross validation and, when necessary, the features were standardized. The degree of success in predicting the models was expressed in terms of accuracy.

## 3. Results and discussion

Regarding the characterization of the study regions, Table 1 shows the average relative air humidity, air temperature, and precipitation for the period 2017-2020 and only for the year 2021. The results indicate that (a) the average air temperature was significantly different in the three regions, (b) L'Aranyó is also the site with less rainfall, and (c) in 2021 there was a slight increase in relative air humidity with respect to the previous four years, which *a priori* could favor the appearance of phytopathogens.

**Table 1.** Characterization of the most important environmental features for the growth of phytopathogens in the three study regions from 2017 to 2020 (top) and in 2021 (bottom): daily mean humidity, mean temperature, and precipitation. The values represented correspond to the mean of the values and their standard deviation. The asterisk indicates significant differences between the corresponding regions for each feature.

Period	Region	Air humidity (%)	Air temperature (°C)	Precipitation (mm)
2017 - 2020	L'Aranyó	69.643 ± 14.220*	15.288 ± 7.492*	1.182 ± 5.925*
	Mirabella Eclano	70.055 ± 14.671*	14.910 ± 6.893*	2.424 ± 6.618
	Quinta do Ataíde	66.908 ± 15.759*	16.729 ± 6.788*	1.433 ± 4.496
2021	L'Aranyó	71.145 ± 14.408	14.792 ± 7.442*	0.751 ± 2.998*
	Mirabella Eclano	72.568 ± 24.620	18.009 ± 6.908*	2.520 ± 6.315
	Quinta do Ataíde	67.630 ± 13.762*	16.185 ± 6.348*	1.288 ± 3.950


Considering the predictive reliability offered by the models, it is possible to observe that, in general terms, a prediction accuracy of over 90% was achieved for the different levels of disease and over 80% for the treatment recommendation (Table 2). To select the most appropriate algorithm, disease classification models were prioritized, seeking not only high degrees of accuracy, but also a minimum difference between the train and test sets. Therefore, the Support Vector Machines (SVM) model for downy mildew and the Random Forest (RF) model for powdery mildew were finally selected. The coefficients of the feature importance shown by these models agreed with those described in the literature, since the SVM model showed the minimum/average air temperature and the degree of humidity as determining features for classification, which is consistent with the oospores being completely dependent on environmental and foliar humidity for their penetration into the plant [6,7]. In addition, the RF model showed that the most important feature for classification was air temperature, which is also in agreement with all the literature studied, in which temperature is established as a critical feature for secondary infection [8,9]. Both models, but especially the SVM, emphasized the importance of the phenological stage of the plant when classifying the risk of infection.

**Table 2.** Average of the accuracy values of the train-validation/test sets for each of the models developed.  $A_d$  means accuracy of the disease model and  $A_t$  means accuracy of the treatment model.

Model	Downy mildew		Powdery mildew	
	$A_d$ train/test	$A_t$ train/test	$A_d$ train/test	$A_t$ train/test
Logistic regression	0.970 / 0.964	0.874 / 0.856	0.978 / 0.956	0.894 / 0.837
Decision tree	0.973 / 0.950	0.859 / 0.878	0.976 / 0.926	0.898 / 0.875
Random forest	0.977 / 0.957	0.870 / 0.878	0.978 / 0.963	0.911 / 0.881
Gradient Boosting	0.975 / 0.965	0.875 / 0.863	0.981 / 0.941	0.913 / 0.889
K-nearest neighbors	0.965 / 0.958	0.868 / 0.870	0.976 / 0.949	0.885 / 0.881
Support Vector Machines	0.965 / 0.964	0.882 / 0.863	0.978 / 0.956	0.889 / 0.859
Naïve Bayes	0.934 / 0.901	0.816 / 0.814	0.976 / 0.933	0.819 / 0.815
Deep Neural Network	0.965 / 0.944	0.868 / 0.798	0.969 / 0.926	0.882 / 0.837

However, despite the promising results, it is advisable to continue improving the models in the following years with the incorporation of more data from subsequent seasons that provide greater variance to the sample, as well as data from regions with different environmental conditions (e.g., regions with cold climates) to those already present for model validation. Finally, thanks to the predictions obtained from mature climatic and phenological models, which will serve as new inputs to the models, it will be possible to make short- and medium-term forecasts and treatment recommendations for growers.

## Acknowledgements

We would especially like to thank Familia Torres Wines, Mastroberardino Società Agricola Srl, and Symington Family Estates for providing the vineyards for these experiments. We would also like to thank Mr. Federico Oldani from LINKS Foundation for facilitating access to the historical data in an automated way. L.F. Velasquez is a fellow of Eurecat's "Vicente López" PhD grant program. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869565. 

## References

- [1] Pink M. The sustainable wine market in Europe - introduction to a market trend and its issues. *Acta Scientiarum Polonorum Oeconomia*. 2015;14(2).
- [2] Shipton PJ. Monoculture and Soilborne Plant Pathogens. *Annu. Rev. Phytopathol.* 1977;15(1): 387-407.
- [3] Shruithi U, Nagaveni V, Raghavendra BK. A Review on Machine Learning Classification Techniques for Plant Disease Detection. 5th International Conference on Advanced Computing & Communication Systems (ICACCS); 2019 March 15-16; Coimbatore, India: IEEE; c2019. p. 281-84.
- [4] Chen M, Brun F, Raynal M, Makowski D. Forecasting severe grape downy mildew attacks using machine learning. *PLOS ONE*. 2020;15(3): e0230254.
- [5] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 2011;12: 2825-30.
- [6] Lalancette N, Madden LV, Ellis MA. A quantitative model for describing the sporulation of *Plasmopara viticola* on grape leaves. *Phytopathology*. 1988;78:1316-21.
- [7] Orlandini S, Massetti L, Dalla Marta A. An agrometeorological approach for the simulation of *Plasmopara viticola*. *Comput. Electron. Agric.* 2008;64:149-61.
- [8] Chellemi DO, Marois JJ. Development of a demographic growth model for *Uncinula necator* by using a microcomputes spreadsheet program. *Phytopathology*. 1991;81:250-54.
- [9] Lu W, Newlands NK, Carisse O, Atkinson DE, Cannon AJ. Disease Risk Forecasting with Bayesian Learning Networks: Application to Grape Powdery Mildew (*Erysiphe necator*) in Vineyards. *Agronomy*. 2020;10:622-50.