13

AI-Driven Strategies to Implement a Grapevine Downy Mildew Warning System

Luiz Angelo Steffenel[1], Axel Langlet[1], Lilian Hollard[1], Lucas Mohimont[1], Nathalie Gaveau[1], Marcello Copola[2], Clément Pierlot[3], and Marine Rondeau[3]

[1]Université de Reims Champagne Ardenne, France
[2]STMicroelectronics, France
[3]Vranken-Pommery Monopole, France

Abstract

In this paper, we assess the usage of machine learning techniques to predict the infection events of Downy Mildew. Every year, Champagne vineyards are exposed to grapevine diseases that affect the plants and fruits, most caused by fungi. Using data from an agro-meteorological station, we compare machine learning performances against traditional prediction methods for Downy Mildew (*Plasmopara viticola*) infections. Indeed, depending on the year, we obtain 82 to 97% accuracy for primary infections and 98% for secondary infections. These results may guide the development of Edge AI applications integrated to meteorological stations and agricultural sensors,and help winegrowers to rationalize the vine's treatment, limiting the damages and the usage of fungicide or chemical products.

Keywords: artificial intelligence, Downy Mildew, CNN, random forest, SVM.

13.1 Introduction

Every year, Champagne vineyards are exposed to grapevine fungal diseases that affect the plants and fruits. Black rot (*Guignardia bidwellii*), Downy mildew (*Plasmopara viticola*), Powdery mildew (*Erysiphe necator*), and Graymold (*Botrytis cinerea*) are examples of diseases that can affect grape quality and hinder the productivity. Each fungus develops under certain environmental conditions and detecting favourable conditions for the spread of the diseases may lead to proactive actions to prevent its dissemination.

In the specific case of the Downy Mildew caused by *Plasmopara viticola*, there are two cycles of infestation that affect the grapevine. The first one is caused by sexual spores (called *primary infections*) and the second one by the dissemination of asexual (*secondary infections*) [4].

The mechanical identification of the fungus development cycles and their forecast has already been the subject of several works, including [8][5] or [7]. Indeed, several of these works define algorithms to identify the primary or secondary infection events using a combination of weather and ground observed variables, which led to the creation of decision-support systems for the vine-growers. However, these algorithms are limited to strict input parameters, which are not always available, and do not explore the potential of hidden correlations with other data variables such as dew point, cloud coverage or vapor pressure deficit.

Artificial intelligence, on the other side,relies only on the dataset rather than on models. It uses computing power to expand the search for patterns and correlations among a broader and richer dataset, often reaching similar or better results than existing models.

Despite its potential, artificial intelligence has been rarely used to identify Downy Mildew infections. Among the precursor works, we can cite Chen et al. [3], which applied several regression models as well as random forest and gradient boost to predict severe infection events in the Bordeaux vineyard. Volpi et al. [9] also use decision trees and random forests to identify different diseases in Tuscany, Italy, but relying on meteorological data from ERA5-Land instead of in-site sensors.

Interestingly, artificial intelligence is more used to monitor crops through image systems rather than weather sensors. For instance, [1][2] use image recognition techniques to identify the intensity of the infections on watermelon or squash crops using hyperspectral images from aerial views. Another work [6] uses Convolutional Neural Networks to detect *Plasmopara viticola spores* in microscopic images.

In this paper, we explore the interest of using machine learning techniques to identify Downy Mildew infections using datasets obtained from regular agro-meteorological sensors. Our aim is both to identify the most efficient and robust methods and to prepare the path to their implementation on Edge AI devices deployed directly on the vineyards.

The remainder of this paper is organized as follows: Section 13.2 presents the datasets and research methodology used in this work. Section 13.3 introduces the different machine learning techniques used in this work, as well as their implementation specifications. In Section 13.4 we present a comparative study of machine learning strategies, aiming at their accuracy as well as their robustness over the years. Section 13.5 goes beyond the simple results by discussing the impact of AI-based algorithms on the monitoring of crops. Finally, Section 13.6 concludes this work.

13.2 Research Material and Methodology

13.2.1 Datasets

The data used in this paper was obtained from a Promété AGRI-300 weather station installed at "Moulin de la Housse" vineyard from Vranken-Pommery group in Reims[1]. This station provides hourly readings from several features of interest:

- Wind speed [Km/h] (max, average)
- Wind gust [Km/h] (max)
- Relative humidity [%] (max, min, average)
- Pluviometry [l/m^2]
- Leaf wetting duration [min]
- Dew point [C] (min, average)
- Solar radiation [W/m^2] (average)
- Air temperature [C] (max, min, average)
- Vapor press deficit [kPa] (min, average)

More than 20k entries were recorded for each feature from 2019 to 2021, except for the Leaf wetting duration that could only be recorded in 2019/2020 as the sensor stop working in February 2021.

The presented machine learning approaches are implemented, optimized and evaluated on a Nvidia DGX1 server that includes eight Tesla V100 GPUs connected through an NVlink network supporting up to 40 GB/s bidirectional bandwidth. Regarding programming tools, we have implemented our approaches using the Python language with scikit-learn, Tensorflow and Keras libraries.

[1] Data could be provided upon request

13.2.2 Labelling Methodology

To train machine learning models to identify Mildew favourable situations, we adopted a supervised learning approach. To label the training dataset, we applied the algorithms proposed by [7]. Two different Mildew infection alert situations are identified in that work, each one with strict requirements. Hence, primary infections are related to the conditions for winter spores' germination, which may occur when the average daily temperature exceeds 10 °C and the precipitation within the last 48h reaches 10 mm (called "3-10" flag). If rainfall or gentle breeze (i.e., wind of speed greater than 3.4m/s) occurs at night within the following 48h, primary infection has presumably occurred, causing the start of the incubation period of *Plasmopora viticola*. Figure 13.1 schematizes this algorithm.

Second mildew infections may happen when the incubation period from the first infection has been completed. It depends on favourable night conditions (FNCs) conditions where the weather is humid (relative humidity (RH) $>$80%), and the temperature is higher than 12°C for at least 2h. In such case, the secondary infection warning is raised if we also observe more than 2h of uninterrupted leaf wetness (LW) and average temperature (T) above 10°C, with precipitation or strong wind that can increase spore spread. Figure 13.2 schematizes this algorithm.

Thanks to these two algorithms, we create two binary labels, one for primary alert and the other for secondary alert, used in independent classification models. These labels are only used during the training phase, as our objective is to obtain accurate predictions based on the raw input data from the weather station sensors.

13.3 Machine Learning Models

This section presents different strategies to model the Downy Mildew warning system using machine-learning techniques. As presented in Section 13.2, our dataset covers three years (2019-2021) and includes several features directly related to the algorithms from [7] such as temperature, relative humidity, pluviometry, wind speed or leaf wetness. Other algorithms variables were adapted from existing data, so the absence of solar radiation (provided by the weather station) was used as an indicator for night time instead of a calculation based solely on the date.

We deliberately kept other variables not cited in the original algorithms, such as the dew point and the vapor press deficit. As stated before, our aim

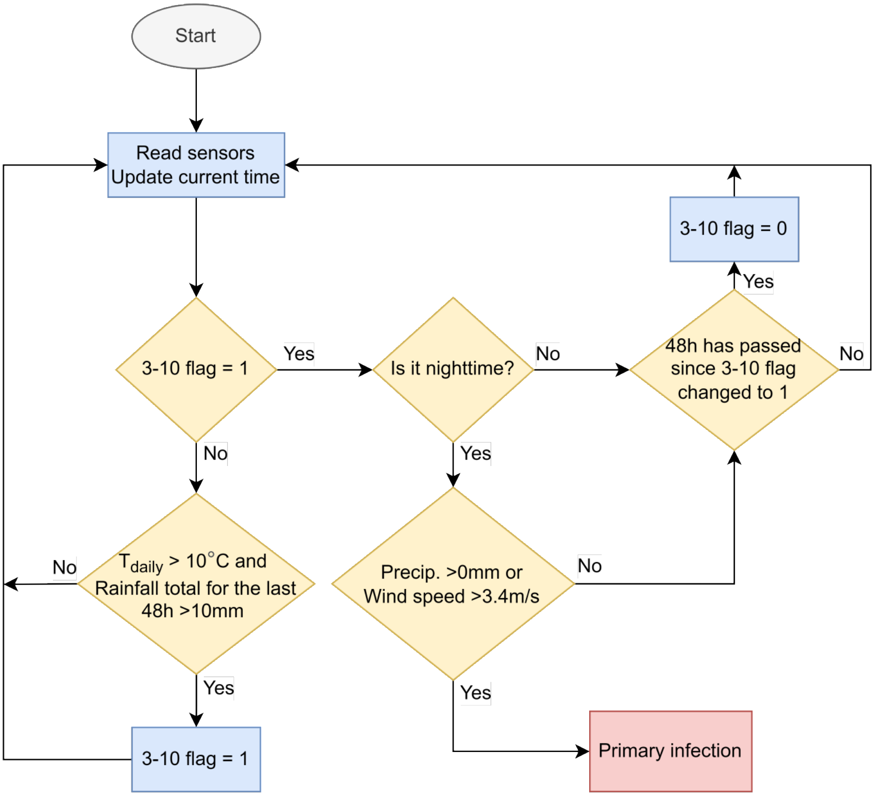

Figure 13.1 Algorithm for primary infection alarms [7]

is to explore potential correlations with additional variables. Similarly, we do not compare the accuracy with the real risks in the vineyard but only with the expected labels. Performing such comparison requires on-site evaluation and a separate tagging from a human operator, which is part of our future works.

Another point to consider is how to enter the dataset as alerts depend on historical events from at least the last 48h. Instead of using mode complex time-series models such as LSTM or GRU, we chose to feed the algorithms with a concatenation of the features recorded in the last 48h. This approach allows us to express the problem in a simpler way that can be approached using a wider range of machine learning techniques, including some best adapted to constrained environments such as those in a Edge AI scenario.

As a result, we model the problem as a binary classification problem, i.e., for each level of infection alert (primary or secondary), we create separated "

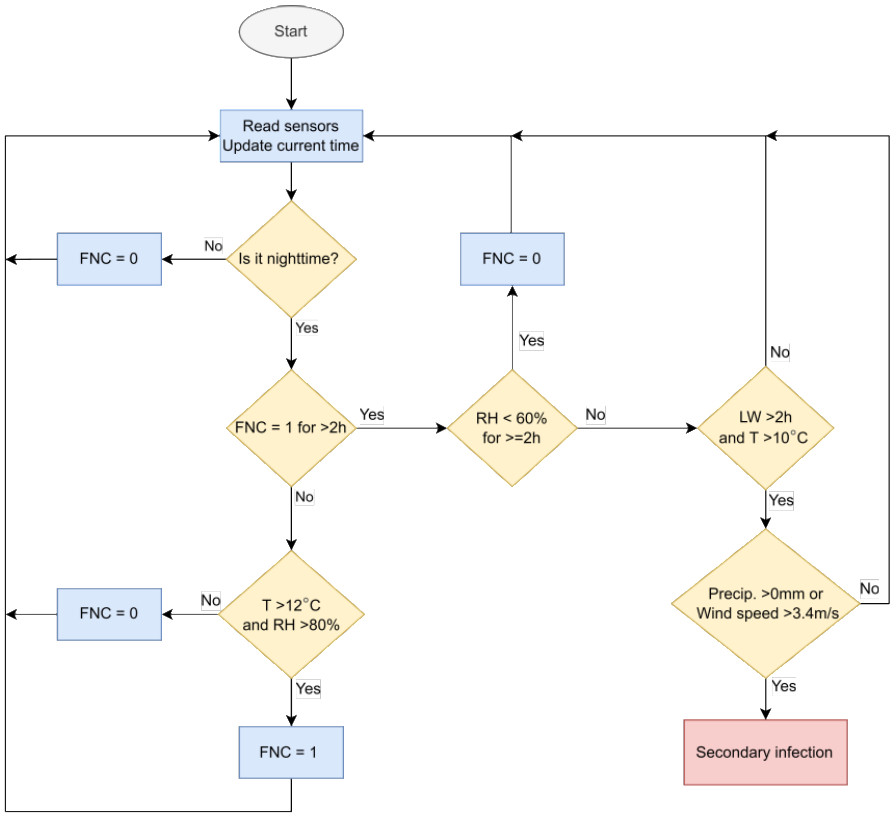

Figure 13.2 Algorithm for secondary infection alarms [7]

alert"/ "not alert" labels. We decided to split it into two binary classification problems instead of a multi-class classification problem to favour each alert type's accuracy. Henceforth, we choose to compare five well-known binary classification techniques:

- Decision trees
- Random forest
- Support Vector Machines (SVM)
- Dense Neural Networks (DNN)
- Convolutional Neural Networks (CNN)

Decision Trees and Support Vector Machine predictors use the basic *scikit-learn* implementation (*DecisionTreeClassifier* and *SVC*, respectively) with no additional optimisation. Random Forests (*RandomForestClassifier*) were trained with the parameter *n iterators = 1000*.

The Dense Neural Network implemented in Keras using seven Dense layers with respectively 200, 100, 100, 50, 50,10, and 2 outputs. Activator *ReLU* was used in all but the last layer (*None*); the model was compiled with *Binary Crossentropy (from logits=True)* loss, Adam optimizer, and *Binary Accuracy* metrics.

Finally, the Convolutional Neural Network was implemented in Keras, using at the input two conv2D layers (32 and 64 outputs, respectively), with 3x3 padding, 0.2 dropout and ReLU activation. Once flatted, a Dense network with 100 outputs, 0.5 dropout and ReLU activation sits just before a final Dense network with 2 outputs (Sigmoid activation).

As the dataset only covers three years, we adopted a cross-validation approach where, for each technique, we generated a different model for each respective year (2019, 2020 or 2021). Therefore, each model was trained only with the data from its own year, split into 90% training and 10% testing parts (randomly shuffled) and later submitted to cross-validation against the other years. Not only the cross-validation helps identifying the most robust model but also allows to investigate the impact of the 2021 weather profile, which differed from the two previous ones due to several climatic events (early crop freeze, rainy weather) that favoured the spread of diseases and led to a massive reduction in crop production and quality.

13.4 Results

13.4.1 Primary Mildew Infection Alerts

As stated above, we create three different training-validation datasets, one for each year. Therefore, Table 13.1 compares the accuracy score from the 2019's model when applied to 2020 and 2021. The best scores are presented in bold, showing that two techniques detach from the others: CNN and SVM. CNN shows slight better scores in the 2021 dataset but is closely followed by SVM.

In the case of the 2020's model, Random Forest and SVM perform well for the 2019 case, and almost all techniques (except simple Decision Tree) present similar results for the 2021 case (see Table 13.2). Finally, the 2021's model Random Forest seems the best technique for the 2019 dataset, while SVM is better in the case of the 2020 dataset (Table 13.3). We can, however, point out that Random Forest achieves good results in this latter case, even if not as good as the SVM scores. If the“ best” technique varies from year to

Table 13.1 Accuracy of 2019 Primary Infection Models

	2019	2020	2021
Decision Tree	-	0.607	0.597
Random Forest	-	0.841	0.743
Support Vector Machines	-	**0.978**	0.821
Dense Neural Network	-	0.909	0.815
Convolutional Neural Network	-	**0.978**	**0.822**

Table 13.2 Accuracy of 2020 Primary Infection Models

	2019	2020	2021
Decision Tree	0.925	-	0.797
Random Forest	**0.935**	-	**0.821**
Support Vector Machines	**0.935**	-	**0.821**
Dense Neural Network	0.925	-	**0.821**
Convolutional Neural Network	0.932	-	**0.821**

Table 13.3 Accuracy of 2021 Primary Infection Models

	2019	2020	2021
Decision Tree	0.921	0.881	-
Random Forest	**0.938**	0.961	-
Support Vector Machines	0.935	**0.974**	-
Dense Neural Network	0.910	0.849	-
Convolutional Neural Network	0.915	0.895	-

year, both SVM and CNNs show robust results, closely followed by Random Forest.The choice reposes therefore in the computing capabilities available to the devices.

We can also see that 2021 was different from the previous ones. If models from 2019 or 2020 achieve lower scores when predicting 2021 alerts, we can also say that models trained with 2021 data are among the best ones when predicting alerts for the previous years. This was somehow expected, as 2021 was rich in favourable events for spreading diseases in the vineyard.

13.4.2 Secondary Mildew Infection Alerts

As the meteorological station stopped recording leaf wetness in February 2021, we could not tag Secondary Mildew Infections on the 2021 dataset. Nonetheless, we compare both 2019 and 2021 models in cross-validation, as we previously did for the Primary Mildew Infection.

Hence, Table 13.4 condenses the results from all machine learning techniques when cross validating each year's models. Secondary Infection alerts

Table 13.4 Accuracy of 2021 Primary Infection Models

	2019 (model 2020)	2020 (model 2019)
Decision Tree	0.960	0.895
Random Forest	0.979	0.988
Support Vector Machines	0.979	**0.991**
Dense Neural Network	0.932	**0.991**
Convolutional Neural Network	**0.980**	**0.991**

seem much easier to identify, with higher accuracy scores. Unfortunately, the absence of a 2021 dataset does not allow a broader comparison under different weather conditions (2021 presented the lowest accuracy in the Primary Alert experiments).

Once again, CNN presents the highest accuracy scores, closely followed by SVM and Random Forest. Indeed, we shall point-out that SVN and Random Forest are good candidates when considering the implementation on environments with performance restrictions, such as in the case of IoT / Edge AI.

13.5 Discussion

The results obtained here are encouraging but shall be considered in the context of the reduced span of the dataset gathered from a single agro-meteorological station installed since 2019. A deeper analysis would require several years of data, as performed by [3] or [9].

However, our main objective was to conceive a proof of concept inscribed in the efforts of the European project AI4DI to develop and disseminate an environmental monitoring system based on different industrial sensors (e.g., TEROS, Bosch BME68x, ST Microelectronics) connected to STM32WL enhanced by a machine learning core. These sensors are expected to enable continuous monitoring of the environment, the soil, meteorological conditions, and/or plant performances.

Besides implementing AI models on the STM32WL, some sensors can also be enriched with a machine learning core. This is the case of the LSM6D SOX sensor from ST Microelectronics, which comprises a set of configurable parameters and decision trees able to run AI algorithms in the sensor itself. Hence, this environment would benefit from simpler models such as random forest and SVM, rather than CNN.

Today, while many agricultural weather meteorological stations are available on the market, innovation comes from implementing Edge AI directly on the sensors or, in some cases, in the gateways. Therefore, the current work represents a primary effort to identify good and robust models that could be deployed in an edge AI environment.

13.6 Conclusion

Every year, Champagne vineyards are exposed to grapevine diseases that affect the plants and fruits, and the Downy Mildew, caused by *Plasmopara viticola* is a common disease. Forecasting the infection events of Downy Mildew may help vine growers to rationalize the treatment of the vine, limiting the damages and the usage of fungicide or chemical products.

In this paper, we compare the accuracy of several machine learning techniques when applied to datasets from the Champagne region. By creating multiple models and using cross-validation across different years, we were able to identify three candidate techniques with close results, namely Convolutional Neural Networks, Support Vector Machines and Random Forest.

If CNN seems to be more robust across different years, the accuracy difference is minimal,and the other techniques present an interest in the case of deployment over an Edge AI infrastructure. Indeed, we aim to prepare the path to the implementation of Downy Mildew forecast models on Edge AI sensing devices that will be deployed directly on the vineyards to closely monitor the crops.

Acknowledgements

This work has been performed in the project AI4DI: Artificial Intelligence for Digitizing Industry, under grant agreement No 826060. The project is cofunded by grants from Germany, Austria, Finland, France, Norway, Latvia, Belgium, Italy, Switzerland, and the Czech Republic and - Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU).

We want to thank Vranken-Pommery Monopole for providing the datasets for the training. We also thank the ROMEO Computing Center[2] of Université de Reims Champagne Ardenne, whose Nvidia DGX-1 server allowed us to accelerate the training steps and compare several model approaches.

[2]https://romeo.univ-reims.fr

References

[1] J. Abdulridha, Y. Ampatzidis, J. Qureshi, and P. Roberts. Identification and classification of downy mildew severity stages in watermelon utilizing aerial and ground remote sensing and machine learning. *Frontiers in Plant Science*, 13, 2022.

[2] J. Abdulridha, Y. Ampatzidis, P. Roberts, S. C. Kakarla. Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence. *Biosystems Engineering*, 197:135–148, 2020.

[3] M. Chen, F. Brun, M. Raynal, and D. Makowski. Forecasting severe grape downy mildew attacks using machine learning. *PLOS ONE*, 15:1–20, 03 2020.

[4] C. Gessler, I. Pertot, and M. Perazzolli. Plasmopara viticola: A review of knowledge on downy mildew of grapevine and effective disease management. *PhytopathologiaMediterranea*, 50:3–44, 04 2011.

[5] E. Gonzalez-Domínguez, T. Caffi, N. Ciliberti, and V. Rossi. A mechanistic model of botrytis cinerea on grapevines that includes weather, vine growth stage, and the main infection pathways. *PLOS ONE*, 10(10):1–23, 10 2015.

[6] I. Hernández, S. Gutiérrez, S. Ceballos, R. Iñíguez, I. Barrio, and J. Tardaguila. Artificial intelligence and novel sensing technologies for assessing downy mildew in grapevine. *Horticulturae*, 7(5), 2021.

[7] I. Mezei, M. Lukic, L. Berbakov, B. Pavkovic, and B. Radovanovic. Grapevine downy mildew warning system based on nb-iot and energy harvesting technology. *Electronics*, 11(3), 2022.

[8] V. Rossi, T. Caffi, S. Giosue, and R. Bugiani. A mechanistic model' simulating primary infections of downy mildew in grapevine. *Ecological Modelling*, 212(3):480–491, 2008.

[9] I. Volpi, D. Guidotti, M. Mammini, and S. Marchi. Predicting symptoms of downy mildew, powdery mildew, and graymold diseases of grapevine through machine learning. *Italian Journal of Agrometeorology*, (2):57–69, Dec. 2021.