FEDERICO JOSÉ GONZÁLEZ VILLASANTI

MACHINE LEARNING FOR TOMATO LATE BLIGHT OUTBREAK AND PROGRESS FORECAST IN THE ESPÍRITO SANTO REGION, BRAZIL

Dissertation submitted to the Plant Pathology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Adviser: Eduardo Seiti Gomide Mizubuti

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

G643m
2023

González Villasanti, Federico José, 1991-

Machine learning for tomato tate blight outbreak and progress forecast in the Espírito Santo region, Brazil / Federico José González Villasanti. – Viçosa, MG, 2023.

1 dissertação digital (38 f.): il. (algumas color.).

Orientador: Eduardo Seiti Gomide Mizubuti.

Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Fitopatologia, 2023.

Referências bibliográficas: f.33-38.

DOI: https://doi.org/10.47328/ufvbbt.2024.157

Modo de acesso: World Wide Web.

1. Tomate - Resistência a doenças e pragas - Controle. 2. *Phytopththora infestans*. 3. *Solanum lycopersicum:* Inteligência artificial. 4. Predição. I. Mizubuti, Eduardo Seiti Gomide, 1966-. II. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-graduação em Fitopatologia. III. Título.

CDD 22. ed. 635.642

Bibliotecário(a) responsável: Advania Elza da Silva CRB-6/3263

VIÇOSA - MINAS GERAIS
2023

FEDERICO JOSÉ GONZÁLEZ VILLASANTI

MACHINE LEARNING FOR TOMATO LATE BLIGHT OUTBREAK AND PROGRESS FORECAST IN THE ESPÍRITO SANTO REGION, BRAZIL

Dissertation submitted to the Plant Pathology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Documento assinado digitalmente
FEDERICO JOSE GONZALEZ VILLASANTI
Data: 10/06/2024 10:15:23-0300
Verifique em https://validar.iti.gov.br

Federico José González Villsanti

Author

Documento assinado digitalmente
EDUARDO SEITI GOMIDE MIZUBUTI
Data: 17/06/2024 09:27:49-0300
Verifique em https://validar.iti.gov.br

Eduardo Seiti Gomide Mizubuti

Adviser

AGRADECIMENTOS

Ao meu orientador, Professor Eduardo Mizubuti, pela oportunidade, suporte e paciência.

Aos membros do laboratório de Biologia de Populações, pelo suporte e ajuda.

À Universidade Federal de Viçosa, pela oportunidade de realizar a pós-graduação.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

ABSTRACT

VILLASANTI, Federico José González, M.Sc. candidate, Universidade Federal de Viçosa, August, 2023. **Machine learning for tomato late blight outbreak and progress forecast in the Espírito Santo region, Brazil**. Adviser: Eduardo Seiti Gomide Mizubuti.

Tomato late blight (TLB) caused by *Phytophthora infestans* (Mont.) de Bary is one of the most destructive diseases of tomato crops (*Solanum lycopersicum*). Due to its economic importance, several integrated management tools were developed to improve its control, including disease forecast models. Available models in the market rely mostly on weather-based risk alerts and empirical approaches, while recent technologies such as machine learning provide new capabilities for modeling and forecasting. Six field trials in two years were conducted to gather disease measurements. Each trial had a hyperlocal weather station installed to record meteorological data. A Support Vector Machine (SVM) model was used to forecast disease onset with an accuracy of 95%. Two machine learning models constructed to forecast TLB progress were tested and compared: Random Forest Regressor (RF) and an Extreme Gradient Boosting Regressor (XGBR). The XGBR returned a lower symmetric mean absolute percentage error when compared to the RF for the exponential stage of the epidemics and a similar error for the asymptote stage. The weather variables that affected TLB progress were related to water availability. ML models can predict the onset and development of TLB, despite clear limitations regarding a small disease dataset. Machine learning models can be used to forecast disease and become part of a disease support system aimed at improving TLB management.

Keywords: *Phytophthora infestans*; *Solanum lycopersicum*; Disease Support Systems; Predictive Models; Artificial intelligence.

RESUMO

VILLASANTI, Federico José González, M.Sc. candidate, Universidade Federal de Viçosa, agosto de 2023. **Machine learning for tomato late blight outbreak and progress forecast in the Espírito Santo region, Brazil**. Orientador: Eduardo Seiti Gomide Mizubuti.

A requeima do tomateiro, causada por *Phytophthora infestans (*Mont.) de Bary, é uma das doenças com maior potencial destrutivo para a cultura do tomateiro (*Solanum lycopersicum*). Diversas ferramentas de manejo integrado são desenvolvidas para melhorar a eficiência de controle. Modelos preditivos disponíveis atualmente fornecem alertas fitossanitários baseados em dados meteorológicos ou modelos empíricos, que poderiam se beneficiar de ferramentas como o *machine learning*. Seis ensaios de campo foram conduzidos em dois anos para gerar dados da severidade da requeima ao longo do tempo e registrar variáveis meteorológicas. Um modelo *Support Vector Machine* (SVM) foi utilizado para predizer o risco da doença com 95% de acurácia. Dois modelos de machine learning foram testados e comparados para selecionar as variáveis micrometeorológicas que influenciam o progresso da doença: o *Random Forest Regressor* (RF) e o *Extreme Gradient Boosting Regressor* (XGBR). O XGBR foi comparativamente melhor que o RF para o primeiro estágio da doença e similar no segundo, quando medido pelo erro médio absoluto percentual. As variáveis micrometeorológicas que mais se correlacionaram com o desenvolvimento da doença foi a disponibilidade de água. Modelos de ML podem prever o início e desenvolvimento da doença e podem ser utilizados como parte de um sistema de suporte à decisão para o manejo eficiente da requeima do tomateiro.

Palavras-chave: *Phytophthora infestans*; *Solanum lycopersicum*; Modelos Preditivos; Inteligência artificial.

LIST OF ILLUSTRATIONS

LIST OF TABLES

SUMMARY

1. INTRODUCTION

Tomato *(Solanum lycopersicum* L.) is the second most important vegetable crop after potato (*Solanum tuberosum* L.) (FAO, 2022). In Brazil, the annual tomato production reaches 1.9 million tons, generating BRL 3.5 billion revenue per year. Approximately 24,000 hectares are planted to fresh tomatoes and this acreage is distributed in approximately 49,000 farms across all Brazilian regions (CEPEA, 2020).

Among the many challenges of tomato production disease management ranks at the topmost positions. Late blight caused by *Phytophthora infestans* (Mont.) de Bary is regarded as one of the most destructive diseases of potato and tomato crops (Mizubuti & Fry, 2006; Fry et al., 2015). Several factors contribute to the damage caused by *P. infestans* to host crops, such as the short latent period of the disease (Maziero et al., 2009), the abundant sporulation of the pathogen, the effective dispersal of inoculum by wind, and the pathogen capabilities to evade plant defenses through effectors that suppress the host defenses (Leesutthiphonchai et al., 2018).

Tomato late blight (TLB) outbreaks occur frequently in the south and southeastern parts of Brazil. Isolates of *P. infestans* of the US-1 clonal lineage and A1 mating type were associated with TLB in Brazil and all published records to date point to a clonal population of the pathogen causing TLB (Reis et al., 2003; Miranda et al., 2010). This clonal lineage is considered an old population that was distributed worldwide (Goodwin et al., 1994). The asexual reproductive cycle of *P. infestans* can occur in less than four days at optimal conditions and is responsible for the fast increase in the blighted plant tissue. The shortest incubation and latent periods for isolates of the US-1 lineage were recorded when inoculated tomato plants were kept at 22 °C (Maziero et al., 2009). Frequent outbreaks of TLB are more common during the summer, likely due to the frequent rains during this season and the tolerance of US-1 isolates to higher temperatures (Maziero et al., 2009).

Due to its short life cycle, massive sporulation and fast lesion expansion, large portions of plant tissues can become blighted very quickly if control measures are not properly implemented. TLB control requires intensive use of chemical compounds and may require up to 20 sprays per season (Lee et al., 1999). In fact, management of TLB in Brazil is practically solely dependent on chemical control which is highly effective when properly deployed. However, chemical control is often used inadequately and this can contribute to yield losses, increase the chances of developing populations of *P. infestans* resistant to antimicrobial compounds besides the negative environmental impact caused by the irrational use of fungicides and oomicides (chemical compounds effective against oomycetes Govers, 2001).

Fresh tomato production in Espírito Santo state amounts to approximately 150 thousand tons per season with an average yield of 59 tons per hectare (IBGE, 2021), but there is high usage of chemical compounds to avoid losses due to late blight. The main tomato producing region of the state, located on the hilly side also known as "Capixaba Mountains", is known for the typical high pressure of late blight during the summer. For the past 30 years, the average temperature for this region for the December-May period is 21.8 °C, while the cumulative precipitation is 675 mm (Meteoblue, 2023).

Late blight management in the region is based on chemical control regularly scheduled. The main products used in the region for late blight control are Ridomil Gold® (mefenoxam and mancozeb, 250 mL/100L, Syngenta©), Manzate 800® (mancozeb, 3.0 kg/ha, UPL©), Bravonil 720® (chlorothalonil, 200 mL/100L, Syngenta©), Forum® (dimethomorph, 150 mL/100L, BASF©) and Curzate® (cymoxanil and mancozeb, 2 kg/ha, Corteva©) and the average number of sprays to control the diseases, including late blight, is estimated to be 36 (Spark, 2022). Sprays for late blight control frequently start right after transplant even when environmental conditions are not favorable, reinforcing the need of a decision support tool that can help tomato growers in the region to efficiently manage the disease.

Currently, there are 22 different active ingredients registered for TLB in Brazil (Agrofit, 2023). Both, multi-site fungicides, which disrupt several metabolic processes of the fungal cell, and site-specific fungicides and oomicides, which typically target a single site are available for TLB management. The most used chemical compounds for TLB control are the multi-sites mancozeb and chlorothalonil, the single-site oomicides mefenoxam, mandipropamid, benthiavalicarb and propamocarb hydrochloride. Additionally, cymoxanil, an oomicide with unknown mode of action, is also heavily used in TLB management programs (Spark, 2022). The use of single-site oomicides such as mefenoxam, propamocarb hydrochloride, and more recently mandipropamid (Abuley et al., 2023) represent a challenge for late blight control due to the risk of developing resistant populations of *P. infestans*. Such concern has been put forth for decades since the first report of resistance of *P. infestans* to metalaxyl (an old isomer of mefenoxam) (Davidse et al., 1981). Strict measures aimed at TLB resistance management are imperative and growers must follow the recommendations such as those outlined by the Fungicide Resistance Action Committee (FRAC, 2023).

Disease support systems (DSS) are important tools to optimize the efficacy of control measures. Most DSS are forecasting models for disease outbreaks or increase in severity. Since Van Everdingen (1926) proposed his model for potato late blight, the objective of several studies is to understand disease epidemics and how weather variables affect their development. DSS integrate and organize information required to adopt timely management practices to control a disease, regarding information on the pathogen, the influence of weather on the disease and fungicide efficacy (Small et al. 2015).

Several empirical approaches have been proposed for addressing late blight control on potatoes and tomatoes. Some of the most notable forecast systems were proposed by Smith (1956), Wallin (1962), Krause (1975), Fry (1983), Small et al. (2015) and Dancey et al. (2017). Most models that depend on weather and disease observations were built using data from experiments conducted under controlled conditions. Such models are somewhat site-specific and generally fail to adapt to different environments. Therefore, models that address epidemics at a landscape scale may improve late blight management in areas of distinct environmental conditions from those in which the site-specific models were based upon (Skelsey, 2021). The landscape-scaled models are data-intensive and benefit from using multiple data sources as inputs for disease prediction (Fenu & Malloci, 2021). It should, however, be noted that most forecast systems were developed to support decision making for the management of potato late blight and there is very limited support for TLB. There is only one DSS available to farmers that was specifically developed for TLB management on a regional scale: BlightPro (Small et al., 2015). This system is adopted by the USA Blight Program (USA Blight, 2022) and allegedly results in an added value to farmers that ranges from $496 to $1714 dollars per acre (Liu et al., 2018).

Technologies such as remote sensing and the Internet of Things provide useful and accurate complementary data required to improve agricultural techniques, including pest and disease management (Pierce, 2019). The large amounts of data generated by these tools require powerful computational techniques and those based on artificial intelligence (AI) algorithms, such as machine learning (ML) and deep learning (DL) are quickly expanding in several research areas in agriculture (Kamilaris et al., 2017). The field of image processing for automatic identification of weeds, diseases and pests has attracted attention and several publications address this topic, mainly due to the advancements and popularity of DL algorithms for image processing in other domains outside agriculture (Skelsey, 2021). While automation of disease identification is useful for decision making, its contribution for optimization of control measures such as fungicide sprayings is questionable since most active ingredients must be applied preventatively. The control of high-progress rate epidemics such as late blight must be based on avoiding the establishment of infection (Mizubuti & Fry, 2006).

Plant epidemics are complex processes that involve multidimensional ecological interactions. Artificial intelligence tools excel in the daunting task of finding patterns and general functions that can represent a process of interest. However, to date, most applications of disease forecasting models rely on a binary scenario of risk of disease development (Skelsey, 2021). Nevertheless, a quantitative assessment of risk or probability of disease outbreaks or epidemics growth is needed to provide better decision support. In this regard, AI resources are anticipated to play a major role in the development of robust risk assessment tools.

Several approaches can be used for prediction tasks, but no algorithm works best for all datasets (Skelsey, 2021). Alves et al. (2016) used an artificial neural network (ANN) to predict tomato late blight AUCPD based solely on disease assessments from a controlled environment experiment. They proposed a new methodology for experiments that screen resistant genotypes on tomato breeding programs, reducing the number of assessments with the calibrated ANN. Nevertheless, this ANN-based system was not oriented to support decisions regarding TLB management. To date, there is no system designed to help schedule chemical sprays and to optimize TLB control.

Forecasting TLB epidemics should take two components into account: 1. The start of TLB outbreak; and 2. The increase in TBL severity over time depending on prevailing weather conditions. The forecast of TLB outbreak can be considered as a classification problem. In this regard, a Support Vector Machine (SVM) method can be used for supervised predictions of TLB outbreak. The SVM transforms data with kernel functions and is effective on high dimensional spaces (Burges, 1998). SVM is a powerful algorithm that can handle high-dimensional data and is particularly useful when the number of features is much larger than the number of samples (Ghaddar & Naoum-Sawaya, 2018), such as the case for forecasting disease onset from weather variables. According to Fenu & Malloci (2021), SVMs were mostly used for weather-based risk assessment models. The same authors in 2020 used a SVM and an ANN to predict potato late blight onset. The SVM outperformed the ANN for two of three scenarios tested, and both models were implemented in a disease support system to help farmers manage late blight in the Sardinia region, Italy. Singh et al. (2019) also used a SVM for potato late blight prediction and a factor analysis model to select the most important weather variables. Another example of use of SVMs with satisfactory results is on tomato powdery mildew (Bhatia et al., 2020). A hybrid model with logistic regression outperformed both algorithms when used individually and reached a 92% accuracy. Malicdem & Fernandez (2015) compared SVM and ANN to predict the onset of rice blast. The SVM outperformed ANN on the binary classification task. These examples show that ML models are suitable for disease prediction, but there are some limitations as well.

One of the main challenges when building machine learning forecasting models to predict increase in disease severity values is the relatively small datasets which prevent the usage of certain algorithms. This problem is addressed by choosing less data-intensive algorithms that can perform well in those conditions as well as preprocessing and resampling techniques (Fenu & Malloci, 2021). Almost all research on machine learning models for late blight forecasting are focused on potatoes (Gu et al., 2016; Singh et al., 2019; Fenu & Malloci, 2020; Skelsey, 2021). Among the AI methods that can be used with relatively small training datasets, Random Forest Regressor (RF) and Extreme Gradient Boost Regressor (XGBR) methods are worth considering (Fenu & Malloci, 2021). The RF is an ensemble learning method that combines the recursive partitioning of decision trees with the bagging procedure, which involves creating multiple subsets of the training data by randomly sampling with replacement and training a decision tree on each subset (Breiman, 1996). The final prediction is then made by averaging the predictions of all the trees to generate multiple predictions for a dataset, outputting the class that is the mode of the classes in problems of classification, or an average prediction of the individual trees for problems of regression (Breiman, 1996). The RF method performs well in time series forecasting with multivariate dataset by creating lag variables and seasonal component variables manually. The XGBR is also an ensemble learning method that uses the boosting procedure, which involves training multiple decision trees sequentially, with each subsequent tree learning from the errors of the previous tree and the final prediction is made by combining the predictions of all the trees, with more weight given to the trees with higher accuracy (Chen et al., 2015). Both models include different regularization penalties to avoid overfitting, which is a risk when working with small datasets. Penalty regularizations produce successful training so the models can generalize adequately. Chen et al. (2020) compared RF with a gradient boosting (GB) algorithm to predict high disease severity and incidence of downy mildew in grapes. The GB performed better, and the authors estimated that the implementation of the model could reduce the use of chemical treatments by at least 50%. Alves et al. (2018) proposed a system to predict occurrence of several diseases of olive and grape in Portugal, using a RF model with mixed results.

1.1. OBJECTIVES

1.2. General objective

To implement machine learning algorithms to forecast tomato late blight.

1.3. Specific objectives

- To develop a machine learning-based forecast system for tomato late blight in Espírito Santo state, Brazil.
- To validate the forecast system developed for tomato late blight using data collected under field conditions.

2. MATERIAL AND METHODS

2.1. Field trials

Field experiments were carried during 2021 and 2022. In both seasons, trials were installed at different locations in the Capixaba Mountains region, Brazil. All trials were conducted between December and May, with different transplanting dates and different tomato hybrids susceptible to TLB (Table 1). For all sites, tomato transplants were planted on a 1 m between rows x 1 m between plants in a row spacing leading to a total of 10,000 plants/ha.

Table 1. General characteristics of the experimental plots: Location, elevation, transplanting dates, and hybrids.

Season	Trial name	Coordinates	Elevation (m)	Tomato hybrids	Transplanting date	Municipality
2021	Julimar	-20.220556, -41.052333	1028	Landall	20 Jan 2021	Afonso Cláudio
	Adenilson	-20.386722, -41.095389	1092	Paron Oxy	09 Feb 2021	Venda Nova do Imigrante
	Gobbi	-20.247056, -41.075750	1028	Caniatti	28 Jan 2021	Afonso Cláudio
2022	Elmo	-20.163472, -41.012556	1092	Moriá	08 Feb 2022	Santa Maria do Jetibá
	Adenilson	-20.393722, -41.096333	1060	Fusion	27 Jan 2022	Venda Nova do Imigrante
	Gobbi	-20.253978, -41.054686	934	Pizzadoro	26 Dec2021	Afonso Cláudio

2.2. Experimental design

In the 2021 season, the observational unit consisted of three rows each of 10 tomato plants totaling 30 plants. The experiment was laid out in an incomplete block design, with five treatments and three replicates. In the 2022 season, the observational units were comprised of three rows of 5 plants each, totaling 15 plants per plots. The experiment was set in randomized complete block design with four blocks.

2.3. Disease severity assessments

Natural inoculation was allowed in the trials and assessments started immediately after the transplant. Assessments of late blight severity were made in the central row of each plot using the simplified leaf area diagram proposed by Corrêa et al. (2009). Whole plant severity was visually assessed for six and three plants in the first and second seasons, respectively. Assessments were conducted at every seven days beginning after the detection of the first symptoms in the experimental area, for a total of six assessments during the crop cycle.

2.4. Control of pests and diseases

To control pest and diseases other than late blight, weekly sprays with crop protection products started five days after transplant. Insecticides Ampligo® (clorantraniliprole and lambda-cyalothrin, dose 30 ml/100L, Syngenta ©), Match (lufenuron, dose 80 ml/100L, Syngenta ©), Vertimec® (abamectin, dose 100 ml/100L, Syngenta ©), Trigard® (ciromazin, 15 g/100L, Syngenta ©) and Minecto Pro® (cyantraniliprole and abamectin, 60 ml/100L, Syngenta ©) were used in rotation to control pests. Fungicide Score® (difenoconazole, 25 ml/100L, Syngenta ©) was used to control fungal diseases. A Yamaho© LS-937 motorized backpack sprayer was used to treat plots with crop protection products, volume ranging from 100 liters per hectare right after transplant to 1000 liters per hectare with fully grown tomato plants, operating at 250 lb/pol^2 and equipped with four TT11004® conic nozzles (Teejet©) to guarantee optimal spray coverage.

2.5. Meteorological data

At each experimental site, an Arable™ Mark 2 automatic weather station was installed between 20 and 30 days after transplanting. The weather station was equipped with sensors to measure air temperature, relative humidity, precipitation, leaf wetness, wind speed and direction, vapor pressure, solar radiation and dew point at an hourly or daily basis depending on the variable of interest. Hourly data of air temperature and relative humidity were grouped in periods of 12h. The "day" period comprised data collected from 6 am to 6 pm and the "night" period ran from 6:01 pm to 5:59 am.

Moving averages were calculated for the day and night periods for precipitation, relative humidity, air temperature, and leaf wetness, for "windows" of three, five, and seven days prior to a given assessment. The sum of precipitation for those same windows was calculated as well.

Daily weather data, day and night weather data and moving averages for the described variables and windows were considered as inputs for the models, totaling 53 different weather-related parameters described in Table 2. All weather data was processed with the software Python version 3.7.13.

Table 2. Weather variables used as input for the models.

Weather variable	Description
Rh[1]	Relative humidity
Max Temp[1]	Maximum temperature
Min Temp[1]	Minimum temperature
Precip[1]	Precipitation
Cumulative Precip[1]	Cumulative precipitation
Vapor Pressure Deficit[1]	Vapor Pressure Deficit
Min Rh[1]	Minimum relative humidity
Et[1]	Reference evapotranspiration
Air Temp[1]	Air temperature
Leaf Wetness[1]	Leaf Wetness
Pardw[1]	Photosynthetically active radiation
Swdw[1]	Shortwave downwelling radiation
Swuw[1]	Shortwave upwelling radiation
Dew Temp[1]	Dew temperature
Max Dew Temp[1]	Maximum dew temperature
Rh At Max Temp[1]	Relative humidity at max temperature
Wind Speed[1]	Wind Speed
Day Air Temp[2]	Air temperature for the "day" period
Day Rh[2]	Relative humidity for the "day" period
Night Air Temp[2]	Air temperature for the "night" period
Night Rh[2]	Relative humidity for the "night" period

Precip Rolling Mean 3, 5 or 7[3]	Average of precipitation for three, five or seven days days prior to the assessment
Air Temp Rolling Mean 3, 5 or 7[3]	Average of air temperature for three, five or seven days prior to the assessment
Max Temp Rolling Mean 3, 5 or 7[3]	Average of maximum temperature for three, five or seven days prior to the assessment
Min Temp Rolling Mean 3, 5 or 7[3]	Average of minimum temperature for three, five or seven days prior to the assessment
Rh Rolling Mean 3, 5 or 7[3]	Average of relative humidity for three, five or seven days prior to the assessment
Leaf Wetness Rolling Mean 3, 5 or 7[3]	Average of leaf wetness for three, five or seven days prior to the assessment
Precip Rolling Sum 3, 5 or 7[3]	Sum of daily precipitation for three, five or seven days prior to the assessment
Day Temp Rolling Mean 3, 5 or 7[3]	Average of air temperature during the "day" period for three, five or seven days prior to the assessment
Night Temp Rolling Mean 3, 5 or 7[3]	Average of air temperature during the "night" period for three, five or seven nights prior to the assessment
Day Rh Rolling Mean 3, 5 or 7[3]	Average of relative humidity during the "day" period for three, five or seven days prior to the assessment
Night Rh Rolling Mean 3, 5 or 7[3]	Average of relative humidity during the "night" period for three, five or seven nights prior to the assessment

[1] Directly measured weather variable

[2] Day period: 6:00am – 6:00pm. Night period: 6:01pm – 5:59am

[3] Calculated Moving Averages

2.6. Disease parameters and progress model

The area under the disease progress curve (AUDPC) was calculated for all treatments using the method proposed by Madden et al. (2007). For the two trials, Julimar 2021 and Gobbi 2022, in which the final late blight severity was higher than 20%, a logistic regression was calculated from disease assessments using the logistic model equation (Madden et al., 2007). For those trials, the rate of the disease progress (r) was estimated after fitting the logistic model, transforming the observed disease (y) to the natural logarithm of y/(1-y) against time.

The disease severity curve was divided into two stages: the exponential stage, characterized by an increase in the rate immediately after infection up to the inflection point of the curve, and the second stage of slowdown characterized by decreased rates of progress, from the inflection point to the final assessment.

All calculations were made using the software Python version 3.7.13.

2.7. Data analysis with machine learning algorithms

A regression approach was used to assess the most important weather variables and to forecast TLB progress. Two models were evaluated: Random Forest Regressor (RF) (Breiman, 1996) and Extreme Gradient Boost Regressor (XGBR) (Chen et al., 2015).

Two trials, Julimar 2021 and Gobbi 2022, that had final TLB severity higher than 20% were used for model training and testing. The models were trained with the weather variables described in Table 2 as input and the disease logistic regression curve described in Figure 2, for the different stages described above (exponential and the slowdown), excluding the last three days of each stage that were dedicated for testing the models. The most important hyperparameters for random forests (Probst et al., 2019) were: the number of estimators, or number of trees generated by the model, set to 100; Max depth refers to a hyperparameter that controls the maximum depth of individual decision trees within the forest, or the maximum distance between the root node and the leaf nodes of each tree. Max depth was not limited, meaning that the trees stopped growing only when all samples in a leaf node belonged to the same class. The bootstrap hyperparameter was set to “True”, meaning that the training data was randomly sampled with replacement to create multiple subsets of the data, which are then used to train individual decision trees.

A Support Vector Machine (SVM) (Burges, 1998) was used for supervised predictions of TLB outbreak. All six trials were used for model training and validation. Data were classified into two categories: the “Asymptomatic” class, corresponding to the period before the first lesions were detected in the field, and the “Symptomatic” class corresponding to the period after the first symptoms were detected in the area. For both classes, different trials were used for training and validation. A critical window of four days was established for the model prediction. Critical hyperparameters for SVMs (Karatzoglou et al., 2006) were: the kernel function, which is used to transform the input into a higher-dimensional space separated by a hyperplane, as a linear function; the C parameter, which controls the trade-off between maximizing the margin and minimizing the classification error, was set to one, which results in wider margins and reduces the chance of overfitting; and the gamma parameter that controls the shape of the decision boundary and was set to scale, to calculate the value based on the number of parameters and the data variance.

All models were programmed using the software Python version 3.7.13.

2.8. Performance metrics

To compare the RF and XGBR performance, the main indicator was the symmetric mean absolute percentage error (sMAPE), an accuracy measure based on percentage of relative errors, defined by the absolute error divided by the magnitude of the exact value (Thieu Nguyen et al., 2019). For the SVM, performance was measured by accuracy, precision and recall. Accuracy is defined by (TP+TN)/(P+N) where TP are true positives, TN are true negatives, P the total number of positives and N the total number of negatives. Precision is defined by TP/(TP+TN). Recall is defined by TP/P (Zhu, 2004).

3. RESULTS

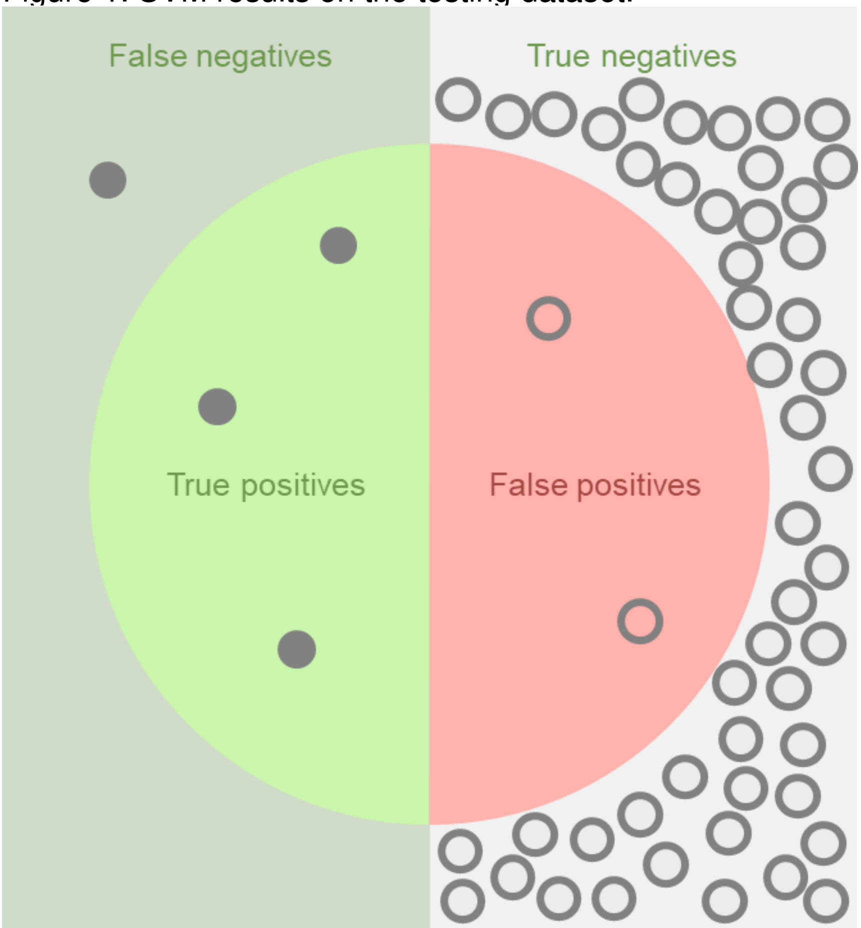
In five of the six trials initial symptoms of TLB were detected at different crop stages. In all trials, ITLB epidemics started after 40 days after transplant and severity was higher where the disease started earlier in the crop cycle (Table 3).

Table 3. Date of detection of late blight measured as days after transplant and the area under the disease progress curve (AUDPC) for all trials.

Trial	First positive assessment of late blight in days after transplant	AUDPC
2021 Adenilson	71	1.13
2021 Julimar	42	10.92
2021 Gobbi	-	-
2022 Adenilson	68	0.01
2022 Elmo	52	0.04
2022 Gobbi	44	8.27

Disease onset forecast: The proposed data split resulted in a training set of 213 days and a testing set of 56 days for the SVM model trained to forecast disease onset. Of the 56 scenarios tested, 50 results were true negatives, three results were true positives, two were false positives and one false negative (Figure 1). The accuracy, precision, and recall calculated from those results were 95%, 60%, and 75%, respectively. The high accuracy value determines that the trained model is performing well when estimating disease onset, whereas the low precision values mean that the model is overestimating the number of situations where disease is likely to occur. Recall is also an important metric for disease forecasting, since the false negatives, defined by situations where disease is likely to occur, but the model failed to predict its occurrence, have a major impact on real applications.

Figure 1. SVM results on the testing dataset.



Disease and weather conditions: Disease severity alongside weather conditions such as daily precipitation, mean temperature and relative humidity for each trial are shown on Figure 1. During the experimental period on both years, the mean temperature was 20.1 °C. The average minimum and maximum temperatures were 15.9 °C and 26.8 °C, respectively. The average relative humidity during daytime was 75.9% and during the night was 95.1%. Rainfall was relatively frequent at the beginning of the cycle and got scarcer with time. The cumulative precipitation during the experiments ranged from 247 mm in trial 2022 Elmo to 473 mm in trial 2022 Adenilson. In the three trials with disease severity above 5%, the epidemics lasted until the final assessment. In two of the three trials disease severity reached an average higher than 60% in plots (Figure 2). The trials where late blight started earlier in the crop cycle and had higher final severity (2021 Julimar and 2022 Gobbi) were planted on January 20th, 2021 and December 26th, 2021, respectively. All other trials were planted later, coinciding with the period with lower precipitation and delayed onset of disease.

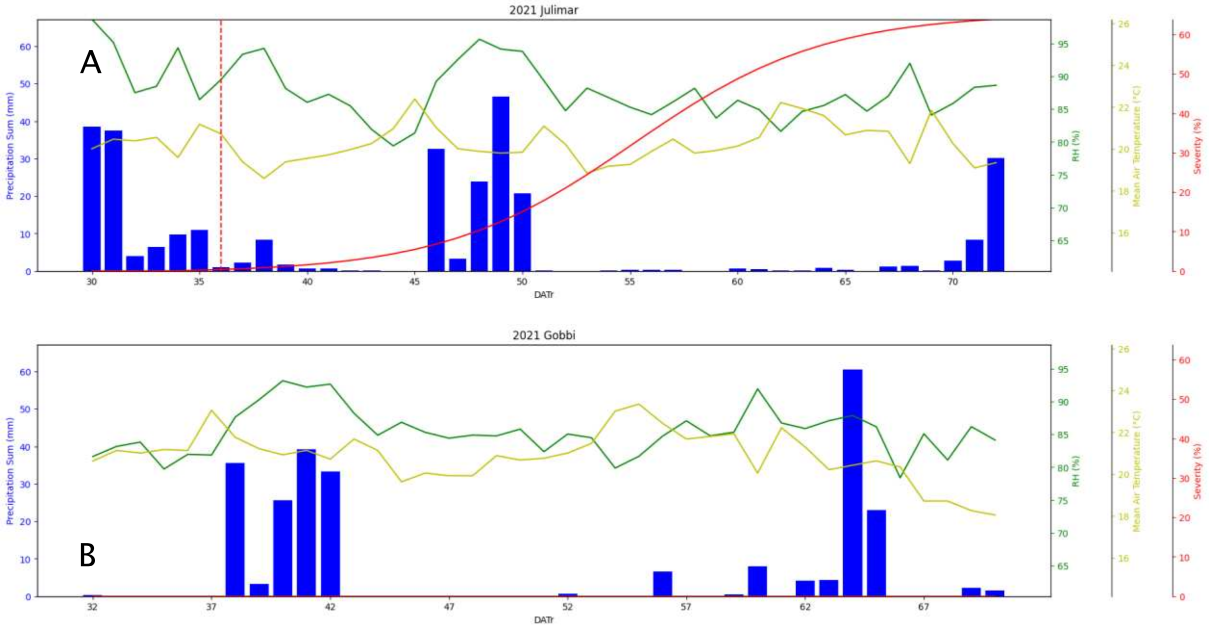


Figure 2. Tomato late blight severity, daily precipitation, mean temperature and relative humidity at different days after transplant (DATr) of the trials: A) 2021 Julimar; B) 2021 Gobbi; C) 2021 Adenilson; D) 2022 Gobbi; E) 2022 Elmo and F) 2022 Adenilson. Vertical dashed line represent severity above 0.5%.

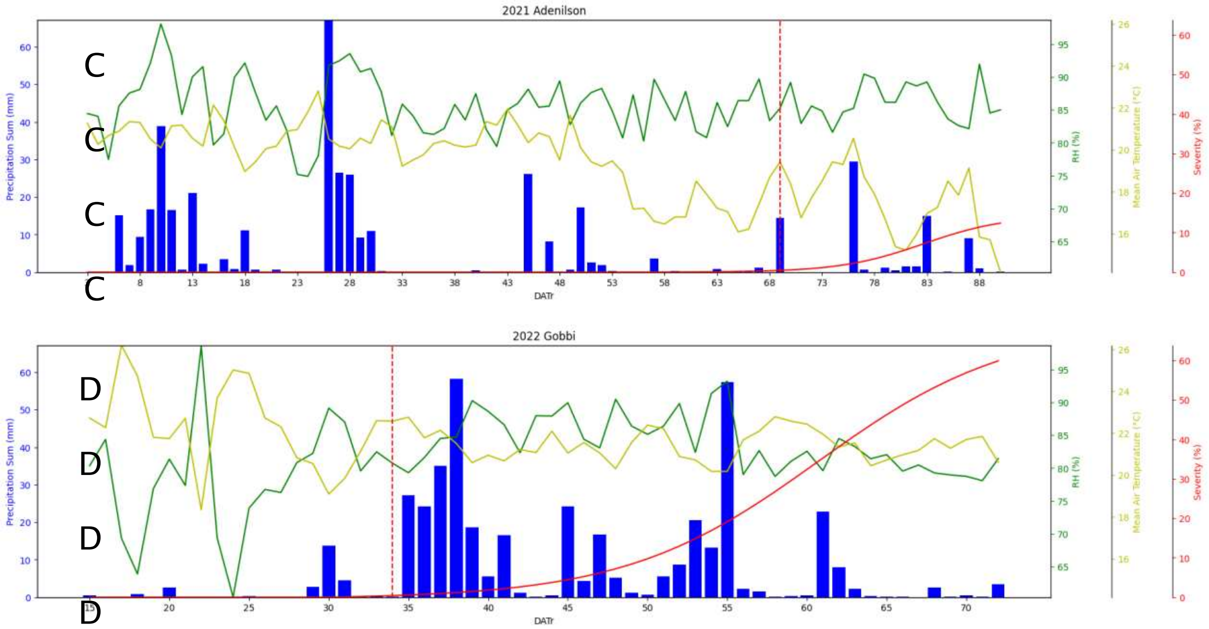


Figure 2. Tomato late blight severity, daily precipitation, mean temperature and relative humidity at different days after transplant (DATr) of the trials: A) 2021 Julimar; B) 2021 Gobbi; C) 2021 Adenilson; D) 2022 Gobbi; E) 2022 Elmo and F) 2022 Adenilson. Vertical dashed line represent severity above 0.5%.

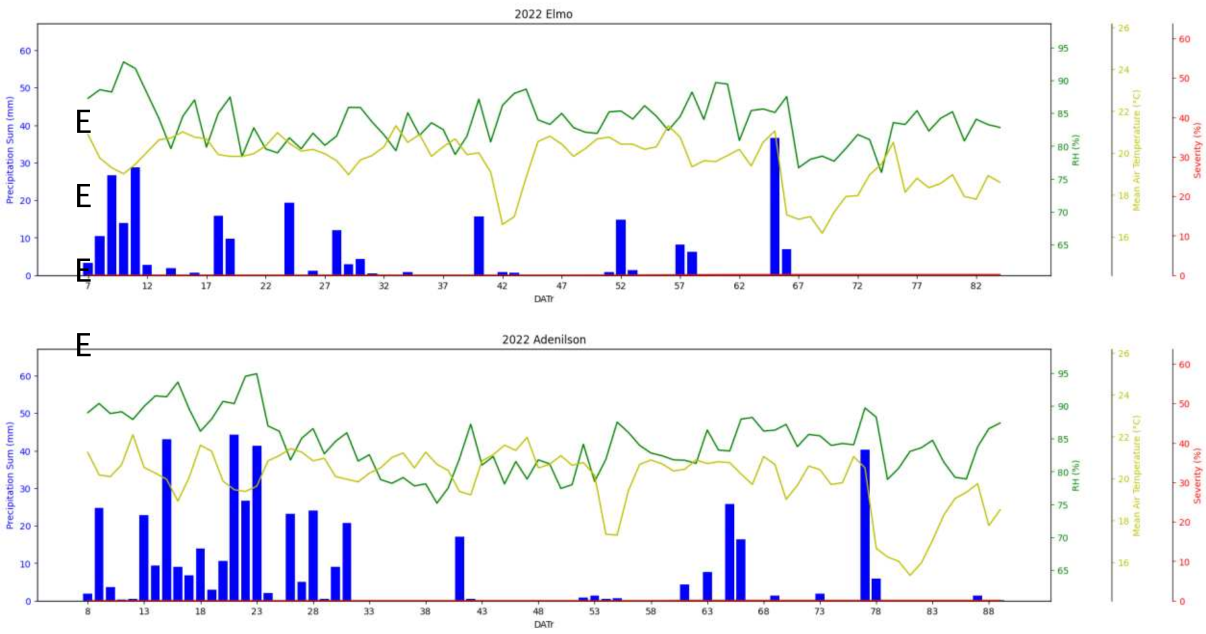


Figure 2. Tomato late blight severity, daily precipitation, mean temperature and relative humidity at different days after transplant (DATr) of the trials: A) 2021 Julimar; B) 2021 Gobbi; C) 2021 Adenilson; D) 2022 Gobbi; E) 2022 Elmo and F) 2022 Adenilson. Vertical dashed line represent severity above 0.5%.

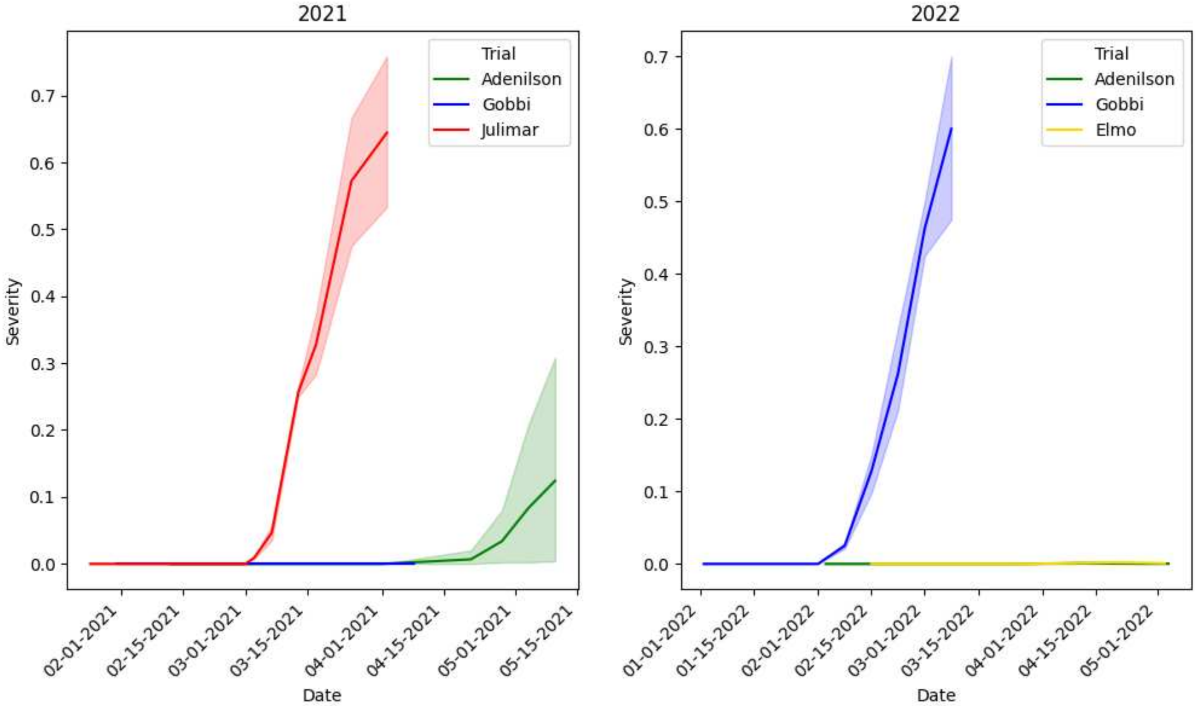


Figure 3. Tomato late blight progress curves in six trials conducted during 2021 and 2022. The solid line represents the average of replicates on time. The shaded band associated with the curves represents the maximum and minimum limits of replicates.

Model fitting: The logistic regressions estimated from disease assessments for 2021 Julimar and 2022 Gobbi (Figure 3). The curve pattern follows that of a typical polycyclic disease with good fit by the logistic model. The disease progress rates (r) estimated for 2021 Julimar and 2022 Gobbi were 0.14 and 0.13, respectively.

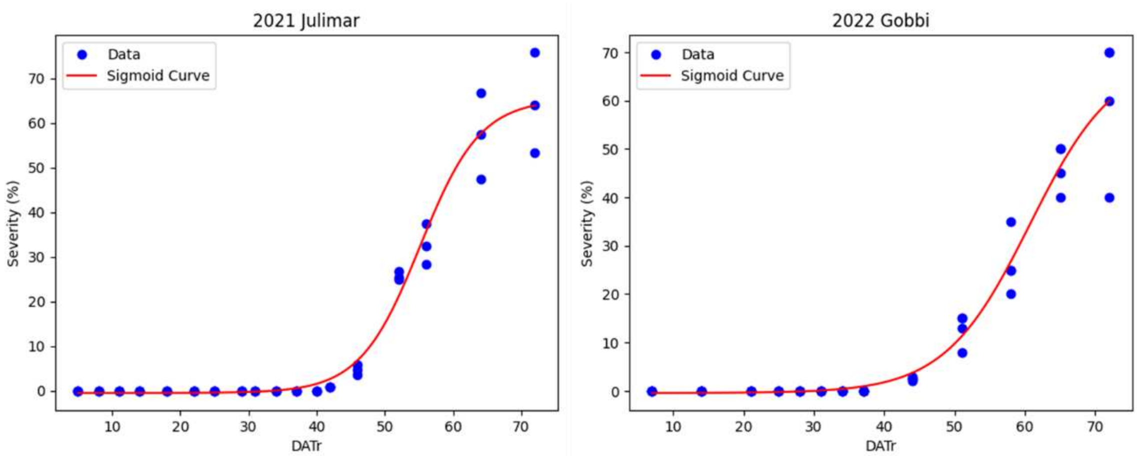


Figure 4. Logistic regression model fitted to tomato late blight severity data in two trials: A) 2021 Julimar and B) 2022 Gobbi.

Feature selection for disease progression and predicted severity: For the two stages of the disease, "exponential" and "slowdown", the RF Regressor and XGBR model were used for feature selection of weather variables from all trials, and the top 10 weather features selected by each model for each stage are displayed in Figure 4. Additionally, the last three days of each stage of the disease progression curve were left out of the training dataset of both models for testing. The comparison between the logistic regression model fitted and the predicted severity of both models for both trials is displayed on Figure 5 and 6. The relative error measured by the sMAPE metric for stages 1 and 2 according to RF were 32.2 and 11.8, respectively. For XGBR, the sMAPE were 17.4 and 12.2, respectively.

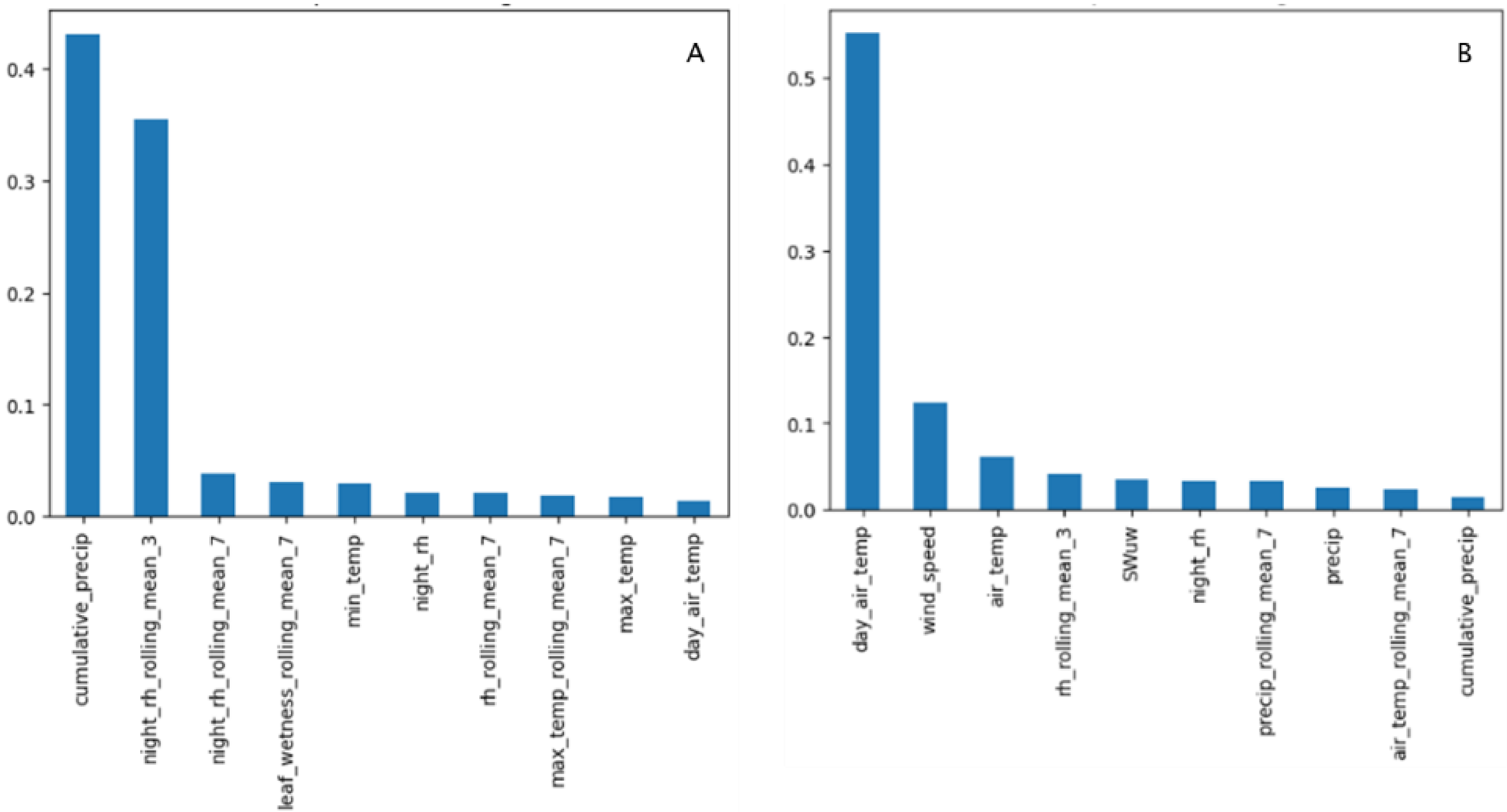


Figure 5. Feature importance selection of weather variables for the XGBR model. Top ten features for each stage. A) Stage 1 - Exponential; B) Stage 2 – Slowdown

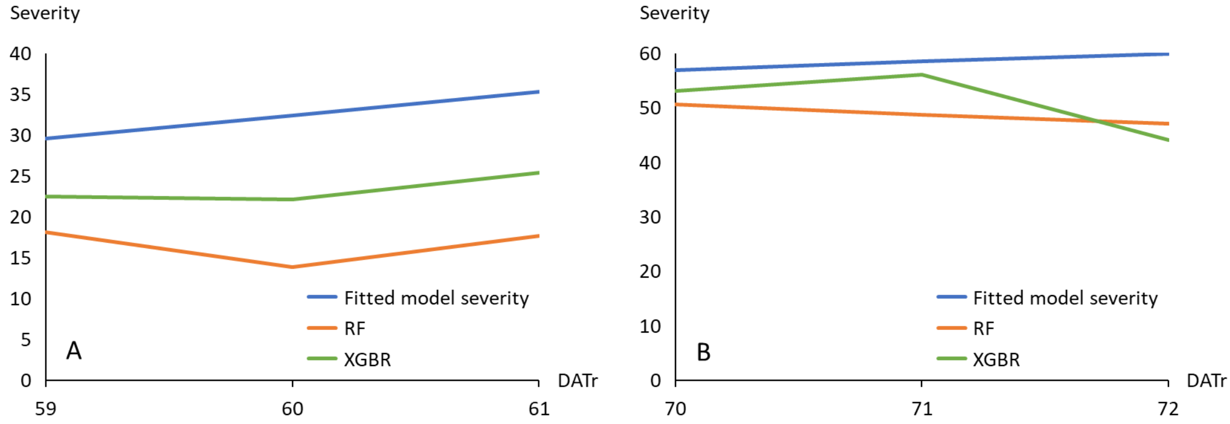


Figure 6. Comparison of the predicted severity of the logistic regression model fitted, the RF and the XGBR models in days after transplant (DATr) for 2022 Gobbi. A) Exponential stage. B) Slowdown stage.

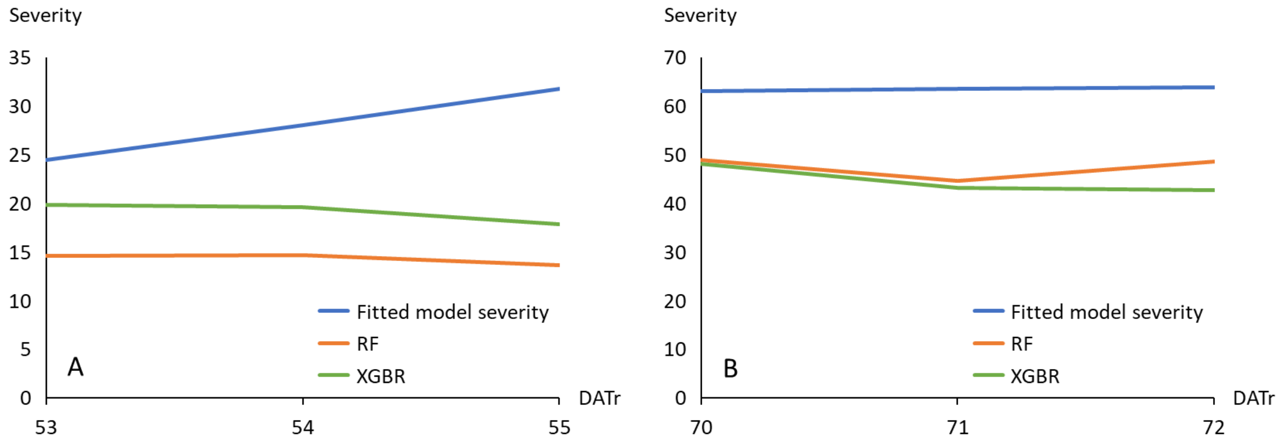


Figure 7. Comparison of the predicted severity of the logistic regression model fitted, the RF and the XGBR models in days after transplant (DATr) for 2021 Julimar. A) Exponential stage. B) Slowdown stage.

4. DISCUSSION

Late blight epidemics have been classically used as an example of a disease that is highly responsive to weather conditions. Therefore, since early in the 20th century, plant pathologists have been trying to construct and validate models that forecast outbreaks and the increase in severity of late blight as a function of weather variables such as relative humidity, temperature, leaf wetness, and precipitation (Smith 1956, Wallin 1962, Krause 1975, Fry 1983, Small et al., 2015; and Dancey et al., 2017). Nevertheless, those models were developed for temperate-climate regions, usually in the European Union (EU) and in the United States of America (USA), where weather patterns are different from Brazil. While validation of those models in local conditions is useful (Batista et al., 2006), the need for local development is imperative considering that these models must be reliable to be trusted and adopted by farmers.

Another key aspect related to the responsiveness of the model to weather data is the ecological characteristics of the extant populations of *P. infestans*. Populations differ across regions and may respond differentially to variables such as temperature (Maziero et al., 2009). The most recent data from Euroblight (2022) shows a predominance of clonal lineages EU_36_A2 in 2019 and EU_23_A1 in 2020 affecting tomatoes in Europe. In the USA, the clonal lineage US-23 of mating type A2 is dominant for both tomato and potato late blight (USABlight, 2021). In Brazil, isolates of clonal lineage US-1 and A1 mating type were associated with TLB (Reis et al., 2003; Miranda et al., 2010). Additionally, there was no evidence of recombination in this clonal lineage (Miranda et al., 2010), a different situation from some regions in the EU (Brurberg et al., 2011). Nevertheless, the *P. infestans* population associated with tomato has changed and A2 isolates are found in some tomato producing regions, including in Espírito Santo state with two isolates from Afonso Cláudio and one isolate from Santa Maria de Jetibá (in the Garrafão area) (Guimarães Silva & Mizubuti, unpublished data). These A2 isolates were collected in the experimental area and this is the first time A2 mating type isolates are reported affecting tomato in Espírito Santo. Thus, there is evidence of a major shift in the pathogen population which up to 2019 was dominated by the US-1 lineage of the A1 mating - type. The "new" A2 population may also respond differently to ecological factors such as temperature and moisture. However, this remains to be studied in the future.

Forecasting models that predict the likelihood of TLB occurrence consider favorable weather conditions on a period and reflect that in diverse units, such as late blight favorable days (Hyre, 1954), severity values (Wallin, 1962) or other metrics. The machine learning models are based on a similar process, but instead of having fixed weather parameters, the model uses all parameters available in the training set to forecast the risk for disease. In this case, the SVM model produced accurate predictions and robust recall. Considering the limited dataset used for training and validation, TLB prediction was considered reasonable. The lower percentage of precision indicates that the model is overestimating the number of positive occurrences of disease, which is preferable to underestimating the risk and returning more false negatives. Further validation and testing are needed to deploy a similar model on a larger, preferably, commercial scale, but this result indicates that SVMs can be used as forecasting models for TLB occurrence.

To better understand conducive conditions for disease development in this region, a RF and a XGBR were trained with disease and weather data to select the parameters that influenced disease development the most. On both stages, precipitation, relative humidity, and temperature were more likely to interfere in TLB development, in line with previous research (Maziero et al., 2009).

Cumulative precipitation and the three-day moving average relative humidity for the night period were clearly the two most important features for stage 1 (Figure 4A). Notably, the first parameter for temperature appears in fifth place. This could relate to the fact that average temperatures during the crop cycle were close to the optimal conditions for US-1 clonal lineage determined by Maziero et al. (2009), making water availability in the environment the most important driver for disease development in this region. It is speculated that the "new" lineage may respond similarly to previous US-1.

The second stage stage represents the slowdown of the rate of disease progress. Since this stage can have major influence of the reducing number of healthy individuals and/or healthy plant tissue, the interpretation of the feature selection should consider this missing factor in the model. Whereas for stage 1 temperature parameters were of secondary importance, for stage 2, day temperature seems to influence disease development (Figure 4B). The next parameters selected by the model are wind speed and radiation but at a much lower frequency than day temperature, which indicates that there are no clear secondary factors influencing the second stage. It is important to note that disease severity keeps increasing during this period, but the rate is reduced.

The XGBR algorithm outperformed RF Regressor for the first stage and had a similar performance for the second stage, suggesting that when using ensemble algorithms, boosting is a more adequate technique over bagging for disease forecasting on a time series. Despite clear differentiation between weather parameters for each stage, the main limitation to achieve lower relative error is the small dataset for disease severity, since these models were created to thrive on data-intensive applications. Data augmentation or synthetic data could be alternatives for this scenario, but invariably bigger datasets for disease parameters are needed and represent the main challenge for the widespread adoption of ML models for disease management. The limited availability of field data is mainly due to the dependency on human assessments and consequently the cost of labor. So far, several ML approaches with imagery techniques for late blight detection have been developed with success, such as late blight recognition on leaves (Gao et al., 2021) or with hyperspectral imagery (Duarte-Carvajalino et al., 2018), but none solve the problem of continuously monitoring disease onset and progression in commercial fields to this day, be it due to dependency of humans to get the samples on the former example, or expensive equipment on the latter.

Another critical issue pertains to weather data. While hyperlocal weather stations are a desirable source of detailed data, there is a limitation for scaling model implementation as well, since few farmers have a hyperlocal meteorological station or can afford the financial investment required to buy such equipment. A regional weather station can scale up implementation and provide useful regional alerts such as those provided by the DSS of the USA Blight program (Small et al., 2015), but sacrifices accuracy of weather-based models at the field scale (Skelsey et al., 2010), especially on regions with high spatial heterogeneity such as the one where this work was developed.

Despite encouraging results on predicting disease onset and progress, a bigger dataset should be prioritized to retrain and test these models before deploying it to forecast TLB at commercial fields. Models that work with small datasets are at a higher risk of overfitting, since the number of features in the model are relatively high compared to the degrees of freedom in the dataset, potentially resulting in inaccurate predictions with new data.

This work shows that ML models are useful for TLB forecast and can constitute an additional tool for integrated disease management. Also, there are a series of enhancements that could increase robustness of forecasting models, such as coupling with crop growth models, pathogen mechanistic models such as the Nærstad model (Hjelkrem et al., 2021), spore detection and disease recognition with images. Creating tailored models for each disease instead of implementing the "off the shelf" ML algorithms can also be a way forward. Integrating all these approaches may be the next stage of research in terms of disease forecasting and prevention.

5. CONCLUSION

ML models can predict the onset and development of TLB, despite clear limitations regarding a small disease dataset. A SVM had a high accuracy for forecasting TLB risk, while XGBR outperformed RF Regressor predicting TLB disease progression. The most important weather variables that impacted disease progression were related to water availability in the environment for different windows and stages of disease development. Further research on this topic should focus on capturing more data and validate this approach with additional field trials, to reduce the risk of overfitting and increase reliability of models applied to this pathosystem. Reliable ML disease forecasting models, packed in easy-to-use applications have the potential of contributing with the value proposed by integrated management practices: environmental and financial sustainability of agricultural practices.

6. BIBLIOGRAPHICAL REFERENCES

ABULEY, I. K., LYNOTT, J. S., HANSEN, J. G., COOKE, D. E. L., & LEES, A. K. The EU43 genotype of Phytophthora infestans displays resistance to mandipropamid. Plant Pathology.

ALVES, D. P., TOMAZ, R. S., LAURINDO, B. S., LAURINDO, R. D. F., SILVA, F. F., CRUZ, C. D., SILVA, D. J. H. D. (2017). Artificial neural network for prediction of the area under the disease progress curve of tomato late blight. Scientia Agricola, 74, 51-59.

ALVES, L., SILVA, R. R., & BERNARDINO, J. (2018). System to Predict Diseases in Vineyards and Olive Groves using Data Mining and Geolocation. In ICSOFT (pp. 713-721).

ARORA, R., SINGH, B., & DHAWAN, A. K. (2017). Theory and practice of integrated pest management. Scientific Publishers.

AGROFIT (2023). Available at https://agrofit.agricultura.gov.br/agrofit_cons.

BATISTA, D. C., LIMA, M. A., HADDAD, F., MAFFIA, L. A., & MIZUBUTI, E. S. G. (2006). Validation of decision support systems for tomato early blight and potato late blight, under Brazilian conditions. Crop Protection, 25(7), 664-670.Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.

BHATIA, A., CHUG, A., & SINGH, A. P. (2020). Hybrid SVM-LR classifier for powdery mildew disease prediction in tomato plant. In 2020 7th International conference on signal processing and integrated networks (SPIN) (pp. 218-223). IEEE.

BURGES, C. J. (1998). A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery, 2(2), 121-167.

BRURBERG, M. B., ELAMEEN, A., Le, V. H., NÆRSTAD, R., HERMANSEN, A., LEHTINEN, A., ... & YUEN, J. (2011). Genetic analysis of Phytophthora infestans populations in the Nordic European countries reveals high genetic variability. Fungal Biology, 115(4-5), 335-342.

CEPEA. Centro de Estudos Avançados em Economia Aplicada. (2020). Especial Tomate. Hortifruti Brasil, 201, 14-20.

CHEN, T., HE, T., BENESTY, M., KHOTILOVICH, V., TANG, Y., CHO, H., & ZHOU, T. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.

CHEN, M., BRUN, F., RAYNAL, M., & MAKOWSKI, D. (2020). Forecasting severe grape downy mildew attacks using machine learning. PLoS One, 15(3), e0230254.

COHEN, Y., RUBIN, A. E., & GALPERIN, M. (2018). Oxathiapiprolin-based fungicides provide enhanced control of tomato late blight induced by mefenoxam-insensitive Phytophthora infestans. PLoS One, 13(9), e0204523.

CORREA, F. M., BUENO FILHO, J. S., & CARMO, M. G. (2009). Comparison of three diagrammatic keys for the quantification of late blight in tomato leaves. Plant Pathology, 58, 1128-1133.

DANCEY, S. R., SKELSEY, P. E. T. E. R., & COOKE, D. E. (2017). The Hutton Criteria: a classification tool for identifying high risk periods for potato late blight disease development in Great Britain. In EuroBlight Workshop (p. 53).

DAVIDSE, L. C., LOOIJEN, D., TURKENSTEEN, L. J., and VANDERWAL, D. (1981).Occurrence of metalaxyl-resistant strains of Phytophthora infestans in Dutch potato fields. Neth. J. Plant Pathol. 87:65-68.

DUARTE-CARVAJALINO, J. M., ALZATE, D. F., RAMIREZ, A. A., SANTA-SEPULVEDA, J. D., FAJARDO-ROJAS, A. E., & SOTO-SUÁREZ, M. (2018). Evaluating late blight severity in potato crops using unmanned aerial vehicles and machine learning algorithms. Remote Sensing, 10(10), 1513.

EUROBLIGHT (2022). Genotype Frequency Chart. Available at https://agro.au.dk/forskning/internationale-platforme/euroblight/pathogen-monitoring/genotype-frequency-chart.

FAO. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. (2021). Plant Health and Food Security; International Plant Protection Convention.
FAO. Food and Agricultural Organization of the United Nations. (2022). Tomato. Available at https://www.fao.org/land-water/databases-and-software/crop-information/tomato/en/.

FENU, G., & MALLOCI, F. M. (2019). An application of machine learning technique in forecasting crop disease. In Proceedings of the 3rd International Conference on Big Data Research (pp. 76-82).

FENU, G., & MALLOCI, F. M. (2020). Artificial intelligence technique in crop disease forecasting: A case study on potato late blight prediction. In International Conference on Intelligent Decision Technologies (pp. 79-89). Springer, Singapore.

FENU, G., & MALLOCI, F. M. (2021). Forecasting plant and crop disease: an explorative study on current algorithms. Big Data and Cognitive Computing, 5(1), 2.
Fry, W. E., Apple, A. E., & Bruhn, J. A. (1983). Evaluation of potato late blight forecasts modified to incorporate host resistance and fungicide weathering. Phytopathology, 73(7), 1054-1059.

H7U, W.E. (2012) Principles of plant disease management. Academic Press.
FRY, W. E., BIRCH, P. R. J., JUDELSON, H. S., GRÜNWALD, N. J., DANIES, G., EVERTS, K. L., ... & SMART, C. D. (2015). Five reasons to consider Phytophthora infestans a reemerging pathogen. Phytopathology, 105(7), 966-981.

FRAC. (2023). Fungicide Resistance Action Committee. Available at https://www.frac.info/.

GAO, J., WESTERGAARD, J. C., SUNDMARK, E. H. R., BAGGE, M., LILJEROTH, E., & ALEXANDERSSON, E. (2021). Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning. Knowledge-Based Systems, 214, 106723.

GHADDAR, B., & NAOUM-SAWAYA, J. (2018). High dimensional data classification and feature selection using support vector machines. European Journal of Operational Research, 265(3), 993-1004.

GOVERS, F. (2001). Misclassification of pest as' fungus' puts vital research on wrong track. Nature, 411(6838), 633-633.

GOODWIN, S. B., COHEN, B. A., & FRY, W. E. (1994). Panglobal distribution of a single clonal lineage of the Irish potato famine fungus. Proceedings of the National Academy of Sciences, 91(24), 11591-11595.

GU, Y. H., YOO, S. J., PARK, C. J., KIM, Y. H., PARK, S. K., KIM, J. S., & LIM, J. H. (2016). BLITE-SVR: New forecasting model for late blight on potato using support-vector regression. Computers and electronics in agriculture, 130, 169-176.
Harrison, J. G. 1992. Effects of the aerial environment on late blight of potato foliage - a review. Plant Pathol. 41:384-416.

HJELKREM, A. G. R., EIKEMO, H., LE, V. H., HERMANSEN, A., & Nærstad, R. (2021). A process-based model to forecast risk of potato late blight in Norway (The Nærstad model): model development, sensitivity analysis and Bayesian calibration. Ecological Modelling, 450, 109565.

HUANG, Y., ZHANG, J., ZHANG, J., YUAN, L., ZHOU, X., XU, X., & YANG, G. (2022). Forecasting Alternaria Leaf Spot in Apple with Spatial-Temporal Meteorological and Mobile Internet-Based Disease Survey Data. Agronomy, 12(3), 679.

HYRE, R. A. (1954). Progress in forecasting late blight of potato and tomato. Plant Disease Reporter, 38(4), 245-253.

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. (2021). Produção Agrícola, Espírito Santo, Tomate. Available at https://cidades.ibge.gov.br/brasil/es/pesquisa/14/10380

KAMILARIS, A., KARTAKOULLIS, A., & PRENAFETA-BOLDÚ, F. X. (2017). A review on the practice of big data analysis in agriculture. Computers and Electronics in Agriculture, 143, 23-37.

KARATZOGLOU, A., MEYER, D., & HORNIK, K. (2006). Support vector machines in R. Journal of statistical software, 15, 1-28.

KRAUSE, R. A., MASSIE, L. B., & HYRE, R. A. (1975). Blitecast: a computerized forecast of potato late blight. Plant Disease Reporter, 59(2), 95-98.

KOKOL, P., KOKOL, M., & ZAGORANSKI, S. (2022). Machine learning on small size samples: A synthetic knowledge synthesis. Science Progress, 105(1).

LEE, T. Y., MIZUBUTI, E., & FRY, W. E. (1999). Genetics of metalaxyl resistance in Phytophthora infestans. Fungal Genetics and Biology, 26(2), 118-130.

LEESUTTHIPHONCHAI, W., VU, A. L., AH-FONG, A. M., & Judelson, H. S. (2018). How does Phytophthora infestans evade control efforts? Modern insight into the late blight disease. Phytopathology, 108(8), 916-924.

MADDEN, L. V., HUGHES, G., & VAN DEN BOSCH, F. (2007). The study of plant disease epidemics.

MALICDEM, A. R., & FERNANDEZ, P. L. (2015). Rice blast disease forecasting for northern Philippines. WSEAS Trans. Inf. Sci. Appl, 12, 120-129.

MCPARTLAND, J.M., et al. (2000) Understanding and using fungicides. Plant Health Progress, 1(1), 1-11.

MAZIERO, J. M. N., MAFFIA, L. A., & MIZUBUTI, E. S. (2009). Effects of temperature on events in the infection cycle of two clonal lineages of Phytophthora infestans causing late blight on tomato and potato in Brazil. Plant Disease, 93(5), 459-466.

METEOBLUE. (2023). Available at ttps://www.meteoblue.com

MIRANDA, B. E. C. D., SUASSUNA, N. D., & REIS, A. (2010). Mating type, mefenoxam sensitivity, and pathotype diversity in Phytophthora infestans isolates from tomato in Brazil. Pesquisa Agropecuária Brasileira, 45, 671-679.

MIAO, J., CAI, M., DONG, X., LIU, L., LIN, D., ZHANG, C., & LIU, X. (2016). Resistance assessment for oxathiapiprolin in Phytophthora capsici and the detection of a point mutation (G769W) in PcORP1 that confers resistance. Frontiers in Microbiology, 7, 615.

MIZUBUTI, E. S., & FRY, W. E. (2006). Potato late blight. In The epidemiology of plant diseases (pp. 445-471). Springer, Dordrecht.

PIERCE, F. J., & NOWAK, P. (1999). Aspects of precision agriculture. Advances in agronomy, 67, 1-85.

PENG, C. Y. J., LEE, K. L., & INGERSOLL, G. M. (2002). An introduction to logistic regression analysis and reporting. The journal of educational research, 96(1), 3-14.

PROBST, P., WRIGHT, M. N., & BOULESTEIX, A. L. (2019). Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: data mining and knowledge discovery, 9(3), e1301.

REIS, A., SMART, C. D., FRY, W. E., MAFFIA, L. A., & MIZUBUTI, E. S. (2003). Characterization of isolates of Phytophthora infestans from southern and southeastern Brazil from 1998 to 2000. Plant Disease, 87(8), 896-900.

SINGH, B. K., SINGH, R. P., TIWARI, P., & KUMAR, N. (2019). Climate based factor analysis and epidemiology prediction for potato late blight using machine learning approaches. In 2019 Women Institute of Technology Conference on Electrical and Computer Engineering (WITCON ECE) (pp. 113-122). IEEE.

SKELSEY, P., ROSSING, W. A., KESSEL, G. J., & VAN DER WERF, W. (2010). Invasion of Phytophthora infestans at the landscape level: how do spatial scale and weather modulate the consequences of spatial heterogeneity in host resistance. Phytopathology, 100(11), 1146-1161.

SKELSEY, P. (2021). Forecasting risk of crop disease with anomaly detection algorithms. Phytopathology®, 111(2), 321-332.

SMALL, I. M., JOSEPH, L., & FRY, W. E. (2015). Development and implementation of the BlightPro decision support system for potato and tomato late blight management. Computers and Electronics in Agriculture, 115, 57-65.

SMITH, L. P. (1956). Potato blight forecasting by 90 per cent humidity criteria. Plant Pathology, 5(3), 83-87.

SPARK. (2022). Crop Protection Business Intelligence Panel – Tomatoes. Available at http://spark-ie.com.br/.

THIEU NGUYEN, T. N., NGUYEN, B. M., & NGUYEN, G. (2019). Efficient time-series forecasting using neural network and opposition-based coral reefs optimization. International Journal of Computational Intelligence Systems, 12(2), 1144-1161.

TYRALIS, H., & PAPACHARALAMPOUS, G. (2017). Variable selection in time series forecasting using random forests. Algorithms, 10(4), 114.

USABLIGHT. (2022). Blight Pro. Available at https://usablight.org/cornell-decision-support-home/.

USABLIGHT (2021). Recent US Genotypes. Available at https://usablight.org/about-late-blight/recent-us-genotypes/.

VAN EVERDINGEN, E. (1926). Het verband tusschen de weersgesteldheid en de aardappelziekte (Phytophthora infestans). Tijdschrift over plantenziekten, 32:129-139.

WALLIN, J. R. (1962). Summary of recent progress in predicting late blight epidemics in United States and Canada. American Potato Journal, 39(8), 306-312.

WEBER-BOYVAT, M., ZHONG, W., YAN, D., & OLKKONEN, V. M. (2013). Oxysterol-binding proteins: functions in cell regulation beyond lipid metabolism. Biochemical pharmacology, 86(1), 89-95.

ZELL, A., & MAMIER, G. (1997). Stuttgart Neural Network Simulator version 4.2 University of Stuttgart.

ZHANG, D., ZHAO, Y., & DU, M. I. N. G. H. U. I. (2013). A new supervised dimensionality reduction algorithm using linear discriminant analysis and locality preserving projection. WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS, E-ISSN, 2224-3402.

ZHU, M. (2004). Recall, precision and average precision. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, 2(30), 6.