

# Improving “Bois Noir” Identification with Interpretable Black Box Models

Blaž Škrli<sup>1,2</sup>, Jan Kralj<sup>2</sup>, Nada Lavrač<sup>2</sup>, Marina Dermastia<sup>3</sup>

<sup>1</sup>Jožef Stefan International Postgraduate School, Ljubljana, Slovenia

<sup>2</sup>Jožef Stefan Institute, Ljubljana, Slovenia

<sup>3</sup>National Institute of Biology, Ljubljana, Slovenia

blaz.skrli@ijs.si

**Abstract**—Construction of predictive models for the identification of “bois noir” disease in grapevine is becoming possible due to increasing amount of available data. Machine learning methodology enables prediction and explanation of such data with accuracy, many times superior to conventional predictive models. In this work, we investigate whether state-of-the-art black box machine learning algorithms improve “bois noir” classification performance, and further, whether we can explain how they learn and which genes play key roles in the learning process. On recent empirical data it was demonstrated that current predictive models can be improved by using black box models such as for example neural networks and gradient boosting machines. Finally it was also demonstrated that individual black box-based predictions can be locally approximated by symbolic classifiers, and thus made interpretable in terms of gene expression profiles.

**Key words**—“bois noir”, machine learning, gene marker identification, black box models

## I. INTRODUCTION

“Bois noir” is a grapevine disease associated with the presence of ‘*Candidatus Phytoplasma solani*’ (Quaglino et al. 2013), and remains an active area of research. In plants, phytoplasmas are limited to the phloem, where they thrive and multiply. Typical symptoms include discoloration and necrosis of leaf veins and leaf blades, downward curling of leaves, lack or incomplete lignification of shoots, stunting and necrosis of shoots, abortion of inflorescences and shrivelling of berries (Dermastia et al. 2017). At the cellular level, modifications include sieve-element necrosis, callose depositions at the sieve plates and starch accumulation. Currently, there are no phytoplasma-immune grapevine cultivars.

One of the methods for preventing the disease’s spread is early detection. Modern gene expression detection methodology offers a possibility to measure expression of selected gene markers, based on which the plant’s sanitary status can be inferred (Hren et al. 2009; Prezelj et al. 2016; Rotter et al. 2018). There exist many different approaches for associating a plant’s state with its expression profiles. These approaches can be divided into two main groups; symbolic and non-symbolic, or black box models. Symbolic approaches include standard statistical association tests,

such as Fisher’s exact test, linear and multivariate regression, decision trees (Rotter et al. 2018), association rules, and similar. These approaches are interpretable, fast and scalable, yet not necessarily optimal in terms of predictive power. On the contrary, black box models in majority of the scenarios outperform symbolic models, yet are harder to interpret. One of the currently open research questions revolves around development of methodology for explanation of the black box models (Robnik-Šikonja 2018).

Statistical aspects of “bois noir” spread have been analyzed in a recent study by Rotter et al. (2018). Multiple grapevine plants were observed over the course of 6 years, from 2004 to 2010. Gene expression levels for 21 different genes were estimated by qPCR each year and studied with regard to ‘*Ca. P. solani*’ estimated titer and severity of symptoms. Using the obtained experimental information, for the first time a pathogen-environment-host organism triangle was proposed. Further, a plant’s states in terms of pathogenicity were classified using a symbolic (interpretable) decision tree model. The model used weather-, host gene expression- and pathogen-based features and could predict a pathogen’s state with >75% accuracy.

In this work black box models were investigated about how potentially improve the current state-of-the-art in ‘*Ca. P. solani*’ detection, whilst maintaining interpretability – it can be observed which genetic markers influence a classifier’s decision the most. This paper is structured as follows: i) brief overview of the experimental data used in this study; ii) assessment of some of the current state-of-the-art methods for black box model explanation; iii) demonstration that there exist explainable black box models, which outperform symbolic models, currently used for “bois noir” disease estimation.

## II. METHODOLOGY

The proposed two step approach for improving the current state-of-the-art “bois noir” predictive models is described here. In the first step, it is investigated whether existing black box models can be used to improve the classification performance. In order to test different models, the data needs to be appropriately preprocessed. In this work, only a single measurement per year

was considered. As there were two measurements for the year 2008, the spring measurement was discarded. For each year, a set of instances, corresponding to individual plants were obtained, described with 21 different gene markers (described in detail in Rotter et al. 2018). The gene markers correspond to individual features, whereas the target feature corresponds to a plant’s state. There are two possible states assigned to each plant: healthy and infected. As the data are time-stamped, all the years up to 2008 (without spring measurement) were considered for training the models, and the year 2009 for testing the models. In a second experiment, it was also trained the models on the spring 2008 measurement and try to predict autumn 2008 measurement, i.e. a single time point (spring) was used to predict plants’ states in autumn.

The models considered in this work are the following:

- Random Forests (RF). This algorithm is considered as a non-symbolic variation of decision trees. Here, sets of trees are constructed for different subsets of the input dataset. Their predictions are combined to obtain final predictions. A common approach to combining predictions is by majority voting. Here, prediction of the majority is considered as final prediction.
- Gradient Boosting Machines (GBM). This algorithm similarly constructs trees, yet in this case, boosting is the technique for model improvements. Here, multiple weak learners are combined to form stronger learner.
- Feedforward Neural Networks (NN). Neural networks are becoming increasingly popular methodology for predictive modeling. A neural network consists of multiple layers of neurons – computational units, updated during model training. The training is performed by updating weights corresponding to individual layers in a process termed backpropagation. Weight changes corresponding to correct classes are propagated from the output nodes of the network towards the input nodes. The most common optimization procedure used to train NNs is stochastic gradient descent, an efficient numeric method for finding local optima.
- Support Vector Machines (SVMs). SVMs remain one of the most robust machine learning algorithms. They perform classification via a “kernel trick” - training data is mapped into a multidimensional space where individual classes can be more easily separated. Some of the common kernels used are the rbf and polynomial kernels.
- Decision Trees (DT). This is the baseline model, as used in previous research. Here, a single tree is constructed based on the input data. Branching is determined by a heuristic, such as for example information gain.

- Linear Discriminant Analysis (LDA). A known algorithm which tries to discriminate between individual classes by selecting subsets of the train set.
- Quadratic Discriminant Analysis (QDA). A quadratic variation of the LDA algorithm – better at approximating non-linear relationships.

The parameters used for individual models are as follows. The SVM was parameterized with C parameter, which was set to 4. 100 trees were used for the RF and GBM models. Two hidden layers were considered with the NN model. All models were implemented using Scikit-Learn machine learning library (Pedregosa et al. 2011).

Current state-of-the-art machine learning approaches are not concerned only with construction of black box models, but also their explanation. Explaining otherwise non-symbolic models is not a trivial task. There exist many approaches, which can be broadly split into two groups. The ones explaining individual predictions and the ones explaining the set of all predictions. In a common machine learning setting, an explanation algorithm is given as input a trained model, train data and a predicted instance. It tries to approximate a black box model’s prediction using local feature perturbations, learned with symbolic models. There also exist approaches based on game theory.

In this work we leverage LIME (Ribeiro et al. 2016), a method recently introduced in the deep learning community for explanation of image classifications. Lime operates by approximating individual predictions using the lasso classifier. We consider two different uses of the LIME algorithm. For correctly predicted instances, it was first explored how different black box models behave in terms of features, important for the observed prediction. Next, it was averaged all feature importance over correctly predicted instances to obtain a more general representation of possible causal patterns, picked up by the selected black box classifier.

### III. RESULTS AND DISCUSSION

The results corresponding to predictions of the last year based on all previous data are firstly reported. Classifier performances can be summarized as in Table 1.

Table 1. Classification results for different algorithms.

Classifier	F1 score	Accuracy
RF	0.40	0.77
MLP	0.30	0.65
GBM	0.50	0.92
DT	0.17	0.61
SVM	0.25	0.30
LDA	0	0.88
QDA	0.25	0.30

It can be observed from Table 1, that the Gradient Boosting Machines classifier outperforms other alternatives. As this is a black box algorithm – meaning a user can not directly understand how it learned from the data, we further present analysis of its predictions. As discussed in the methodology section, we leveraged the LIME algorithm for black box model explanation. The main output of this algorithm are contributions of individual features (genes) to a specific prediction. An example result is visualized in Fig. 1.

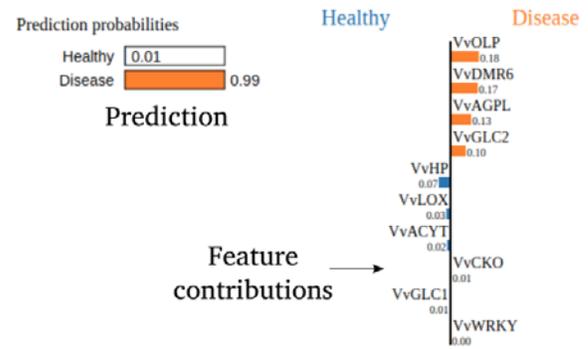


Figure 1. Individual feature contributions for a plant.

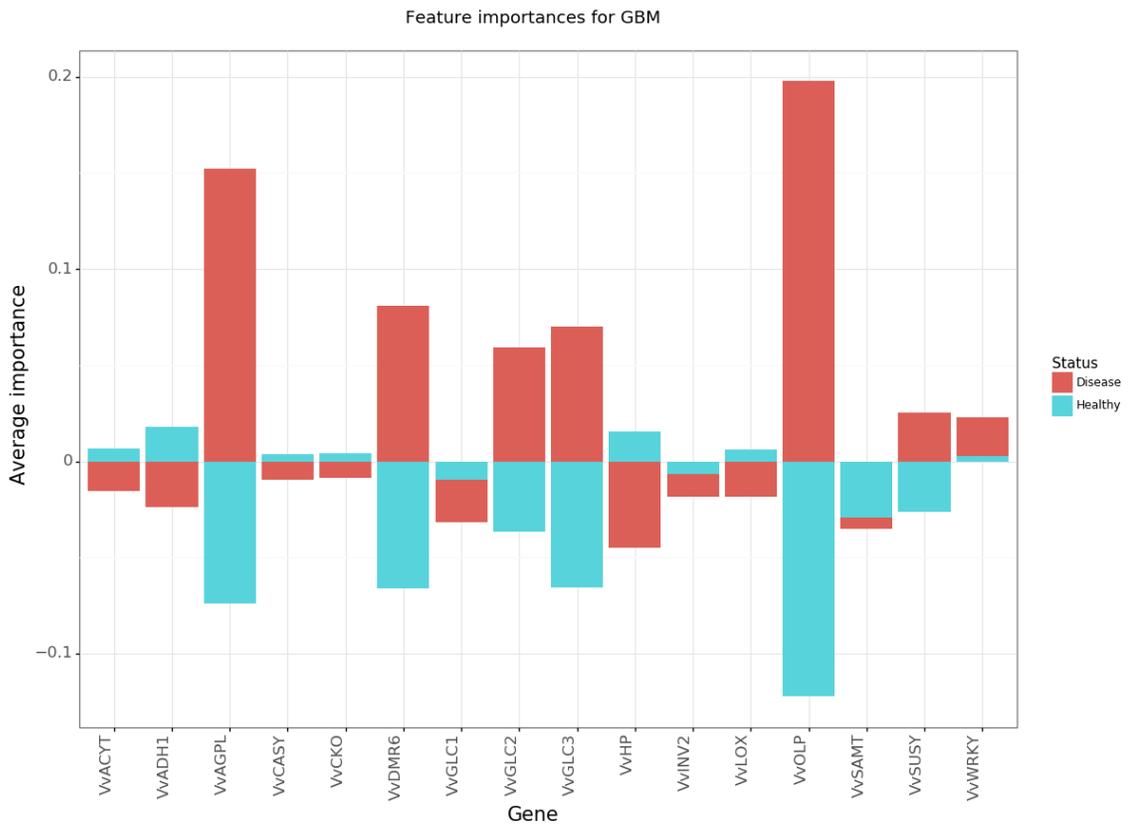


Figure 1. Individual feature contributions with regard to different sanitary status.

It was possible to observe that genes, similar to the ones in the original publication (Rotter et al. 2018) emerged as relevant for disease-related prediction. One of the problems of such explanations is, they only represent interpretations for individual instances (plants). As feature contributions can be obtained for all target instances (plants in the last year), the importance values over all target instances were then used. This visualization is presented in Fig. 2.

The *VvOLP*, *VvGLC3* and *VvAGPL*, *VvDMR6* genes emerged as the most relevant for the sanitary status prediction (plants with “bois noir” disease). The four genes were previously identified as some of the key factors involved in “bois noir” infection process. Next, it was performed similar analysis using only spring 2008 plants to predict sanitary status observed in autumn. Results of the performed predictions are reported in Table 2.

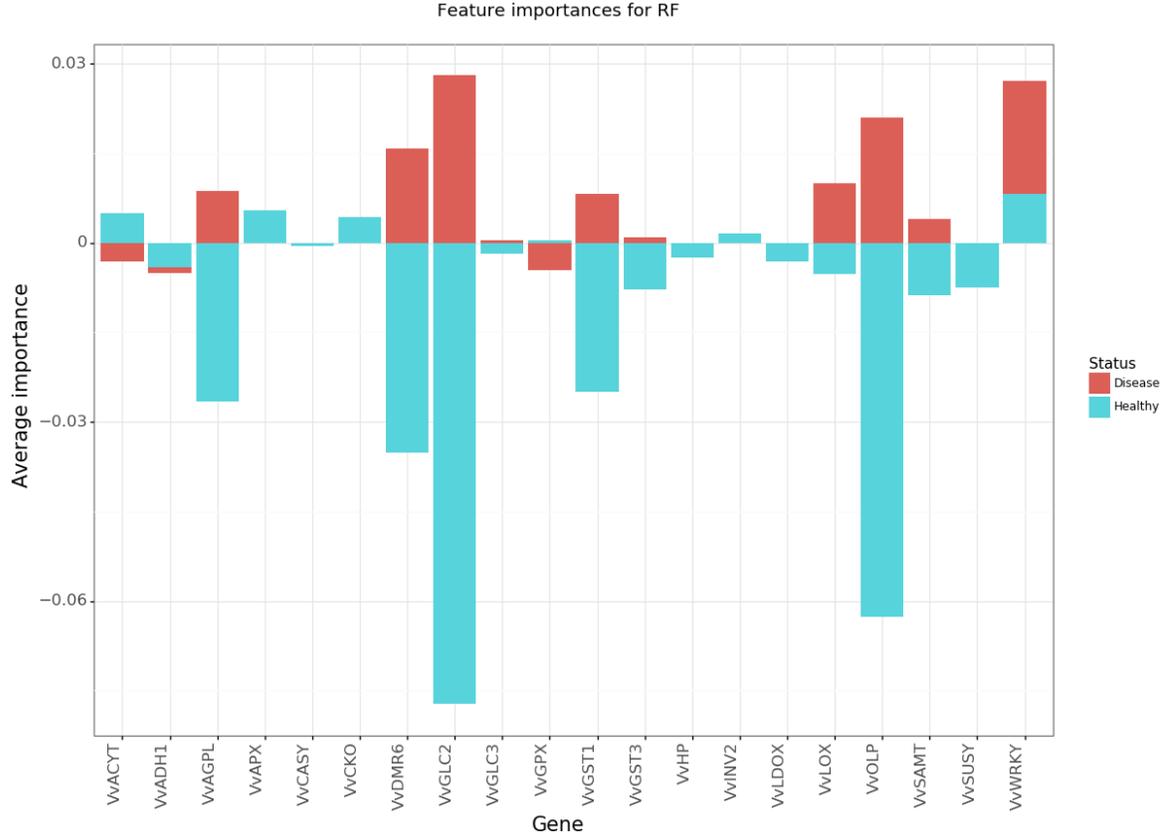


Figure 2. Feature contributions for the spring-autumn prediction. The predictions were obtained using random forests.

Table 2. Spring to autumn predictions.

Classifier	F1 score	Accuracy
RF	0.54	0.88
MLP	0.33	0.80
GBM	0.33	0.84
DT	0.0	0.84
SVM	0.22	0.80
LDA	0.33	0.84
QDA	0.33	0.69

Table 2 similarly indicates the existence of better classifiers, even though the baseline decision tree (DT) classifier performs reasonably well. The best performing classifier in this part of the study is the Random Forest classifier (RF). Its overall feature importance are shown in Fig. 3.

Similarly to the GBM, which overall performs best, RF in this case identified VVOLL and VVGLC2 as important factors, associated with observed sanitary status.

## I. CONCLUSIONS

In this work it was investigated how state-of-the-art black box model explanations can be used to improve understanding, as well as predictive performance of algorithms used to identify “bois noir” occurrence. The results indicate that there indeed exist black box classifiers, which improve predictive accuracy. It was demonstrated how individual predictions, as well as their aggregations over all test instances can be used to explain the classifier’s behavior.

## ACKNOWLEDGMENT

This work was funded by the Slovenian Research Agency (ARRS) grant J1-7151.

## REFERENCES

Dermastia, M., Bertaccini, A., Constable, F., Mehle, N. (2017). Grapevine yellows diseases and their phytoplasma agents: Biology and detection. Springer.

Hren, M., Nikolić, P., Rotter, A., Blejec, A., Terrier, N., Ravnikar, M., Gruden, K., Dermastia, M., Gruden, K. (2009). "Bois noir" phytoplasma induces significant reprogramming of the leaf transcriptome in the field grown grapevine. *BMC Genomics*, 10 (1), 460.

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. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Prezelj, N., Covington, E., Roitsch, T., Gruden, K., Fragner, L., Weckwerth, W., Dermastia, M. (2016). Metabolic consequences of infection of grapevine (*Vitis vinifera* L.) cv. Modra frankinja with "flavescence dorée" phytoplasma. *Frontiers in Plant Science*, 7. <https://doi.org/10.3389/fpls.2016.00711>

Quaglino, F., Zhao, Y., Casati, P., Bulgari, D., Bianco, P. A., Wei, W., Davis, R. E. (2013). 'Candidatus Phytoplasma solani', a novel taxon associated with "stolbur"- and "bois noir"-related diseases of plants. *International Journal of Systematic and Evolutionary Microbiology*, 63, 2879–2894.

Ribeiro, M. T., Singh, S., Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier (pp. 1135–1144). Presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp. 1135–1144.

Robnik-Šikonja, M. (2018). Explanation of prediction models with explain prediction. *Informatica*, 42 (1).

Rotter, A., Nikolić, P., Turnšek, N., Kogovšek, P., Blejec, A., Gruden, K., Dermastia, M. (2018). Statistical modeling of long-term grapevine response to 'Candidatus Phytoplasma solani' infection in the field. *European Journal of Plant Pathology*, 150, 653–668.