Paper: **Crop Pests** **Prediction Method using Regression and Machine learning**

Abstract: With the **advent of data mining, the field of agriculture is also focused on it.** Currently, various **studies, domestic and overseas, are under progress** using machine learning technology, and cases of its utilization are increasing. This paper classifies and **introduces SVM (Support Vector Machine), Multiple Linear Regression, Neural Network, and Bayesian Network** based techniques, and describes some cases of their utilization.

This **paper describes trends in work** **on methods of the prediction of crop pests using machine learning technology.** It briefly introduces methods of using 4 algorithms using SVM (Support Vector Machine), MLR (Multiple Linear Regression), Neural Network, and Bayesian Network, and takes a look at various cases in which they have been used.

**References :**

*[1] Younes Chtioui et al., “A generalized regression neural network and its application for leaf wetness prediction to forecast plant disease”, Chemometrics and Intelligent Laboratory Systems, vol. 48, pp.47–58, 1999*

[*2] Jessica Rutkoski et al., “Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat”, The Plant Genome, vol. 5, pp.51–61, 2012*

[3] D. C. Hooker et al., “Using Weather Variables Pre- and Post- heading to Predict Deoxynivalenol Content in Winter Wheat”, Plant Disease, vol. 86, pp.611–619, 2002

[4] A. K. Tripathy et al., “Data Mining and Wireless Sensor Network for Agriculture Pest/Disease Predictions”, 2011 World Congress on Information and Communication Technologies, pp.1229-1234, 2011

[5*] Cora B. Perez-Ariza et al., “Prediction of Coffee Rust Disease Using Bayesian Networks”, 6th European Workshop on Probabilistic Graphical Models, pp.259-266, 2012*

*[6] Haiguang Wang et al., “Prediction of Wheat Stripe Rust Based on Support Vector Machine”, 2011 Seventh International Conference on Natural Computation, pp.378–382, 2011*

*[7] Wu Dake et al., “The Support Vector Machine (SVM) Based Near-Infrared Spectrum Recognition of Leaves Infected by the Leaf miners”, First International Conference on Innovative Computing, Information and Control, vol. 3, pp.448-451, 2006*

[8] K. Klem et al., “A neural network model for prediction of deoxynivalenol content in wheat grain based on weather data and preceding crop”, Plant Soil and Environment, vol. 53, pp.421-429, 2007 [9] Marie VáĖová et al., “Prediction Model for Deoxynivalenol in Wheat Grain Based on Weather Conditions”, Plant Protection Science, vol. 45, pp.S33–S37, 2009

[10] S. Landschoot et al., “A field-specific web tool for the prediction of Fusarium head blight and deoxynivalenol content in Belgium”, Computers and Electronics in Agriculture, vol. 93, pp.140–148, 2013

Paper: **Early warning system for coffee rust disease** based on **error correcting output codes**: a proposal

Abstract:

**Colombian coffee** producers have had to face the severe consequences of the coffee rust disease since it **was first reported in the country in 1983**. Recently, machine learning researchers have tried to **predict infection *through classifiers such as decision trees, regression Support Vector Machines (SVM), non-deterministic classifiers and Bayesian Networks***, but it has been theoretically and empirically demonstrated that combining multiple classifiers can substantially improve the classification performance of the constituent members. An Early Warning System (EWS) for coffee rust disease was therefore proposed based on Error Correcting Output Codes (ECOC) and SVM to compute the *binary functions* of **Plant Density, Shadow Level, Soil Acidity, Last Night Time Rainfall Intensity and Last Days Relative Humidity.**

INTRODUCTION:

Coffee rust, first reported in 1983 [1], is the most important and severe disease currently affecting the production of Colombian coffee. Resistant varieties have been developed through improvement with genes of Timor Hybrid (a plant that features natural resistance to the disease) as a solution to the rust problem [2], yet **more than 50 percent of the country’s coffee crop is still susceptible in the productive phase**. Studies on coffee rust have concluded that the **spores carrying the infection are spread by climatic elements such as wind and rainfall** [3], wind being the vector for long distance spore transport, while **precipitation droplets are responsible for vertical propagation from infected leaves or soil** [3]. **Once spores make contact with a susceptible leaf, the infection process is improved by high shadow index, high humidity (atmosphere and leaf), soil acidity, high coffee tree density and low soil fertility.**

The **warning system proposed herein** seeks to detect each of the favourable conditions that coffee rust requires to infect the crop and, by taking **prophylactic** measures (biological, chemical and cultural control), thus allow prevention of the onset of the disease. **Having identified the particular conditions favourable to coffee rust,** the system is designed to alert growers to the settled presence of conditions for infection, sending out a graded warning (None, Very Low, Low, Medium, High, Very High) to all coffee growers in the detection area.

REFERENCES:

[1] E. Shieber and G. A. Zentmyer, “Coffee rust in the western hemisphere “in Plant disease, ed, 1984.

[4] M. E. Cintra, C. A. A. Meira, M. C. Monard, H. A. Camargo, and L. H. A. Rodrigues, “The use of fuzzy decision trees for coffee rust warning in Brazilian crops,” in Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on, 2011, pp. 1347-1352.

[5] O. Luaces, L. H. A. Rodrigues, C. A. A. Meira, Jos, #233, R. Quevedo, et al., “Viability of an alarm predictor for coffee rust disease using interval regression,” presented at the Proceedings of the 23rd international conference on Industrial engineering and other applications of applied intelligent systems - Volume Part II, Cordoba, Spain, 2010.

*[6] O. Luaces, L. H. A. Rodrigues, C. A. Alves Meira, and A. Bahamonde, “Using nondeterministic learners to alert on coffee rust disease,” Expert Systems with Applications, vol. 38, pp. 14276-14283, 2011.*

*[7] C. B. Pérez-Ariza, A. E. Nicholson, and M. J. Flores, “Prediction of Coffee Rust Disease Using Bayesian Networks,” presented at the The Sixth European Workshop on Probabilistic Graphical Models, Granada (Spain), 2012.*

[8] L. Li, B. Zou, Q. Hu, X. Wu, and D. Yu, “Dynamic classifier ensemble using classification confidence,” Neurocomputing, vol. 99, pp. 581-591, 2013.

*[9] T. G. Dietterich and G. Bakiri, “****Error-correcting output codes****: a general method for improving multiclass inductive learning programs,” presented at the Proceedings of the ninth National conference on Artificial intelligence - Volume 2, Anaheim, California, 1991.*

[10] Z. H. Zhou, Ensemble Methods: Foundations and Algorithms: Chapman and Hall/CRC;, 2012.

[11] N. J. Nilsson, Learning machines: foundations of trainable pattern-classifying systems. New York: McGraw-Hill, 1965.

[12] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” presented at the Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, 2007.

Paper: **Incidence and severity of coffee diseases** in smallholder plantations in northern Malawi

Abstract:

A **two-year survey (1998, 1999) was conducted in northern Malawi to estimate the occurrence of coffee diseases,** with emphasis on coffee berry disease(CBD) caused by Colletotrichum kahawae, in relation to management practices on smallholder plantations. A total of 113 farms was visited in "ve extension planning areas (EPAs) which grow coffee, and the **prevalence, incidence and severity** of the diseases was recorded. Disease **prevalence was clearly affected by agro-environmental factors**, with no CBD recorded in the SE Mzimba EPA and low levels in the Phoka Hills EPA, but a mean of 95% of farms were affected in the Misuku Hills EPA over the 2 years**. In contrast, coffee rust was found in all areas and was usually more prevalent than CBD.** Disease levels were higher in 1999 than in 1998, and the mean incidence of CBD per affected farm in 1999 was 40%. This ranged, however, from 10 to 56% depending on the EPA. The worst affected farm had 97% of bushes infected with CBD. There was also a difference in disease levels on the two most commonly-grown cultivars. The mean incidence of CBD for cv. Agaro was 23% but only 11% for cv. Geisha in 1998 compared to 59 and 19%, respectively in 1999. **Management practices were correlated with levels of disease in 1999, and pruned or intercropped plantations had lower disease incidence.**

**Paper: The intensity of a coffee rust epidemic is dependent on production situations**

**Abstract:**

A simple **correspondence analysis** was used to **show that a link could be found between certain production situations and the intensity of coffee rust epidemics**. Local characteristics specific to each plantation **were particularly well linked to the intensity of coffee rust epidemics**, whereas **regional factors such as rainfall appeared to be of secondary importance**. **The yield and the number of leaves of the coffee trees were positively linked to epidemic development. Soil pH and fertilisation were negatively associated with epidemic development.** Shade, when it did not limit yield, probably affected the microclimate in such a way that coffee rust incidence increased. Altitude was a serious constraint in disease development.

**Intro:**

We feel that the risk of an epidemic could be assessed simply, by considering that **an epidemic is the outcome of a risk associated with the characteristics of the region, primarily with the climate, but also with the soil that seems to affect coffee rust** (Lamouroux et al., 1995; Avelino, 1999), and of a risk **attributable to local conditions, i.e. a risk primarily linked to the characteristics of the plant, but also to crop management patterns, and especially shade management,** which are known to act on coffee tree plantations microclimate.

Arabica coffee is cultivated in Honduras in a fairly large range of altitudes (mainly from 600 to 1700 m). The **annual rainfall and its pattern are very diverse as a consequence of different oceanic influences.** A substantial Pacific influence is felt in South East, marked by a long dry season. A clear Caribbean influence marks the North West along with a short dry season (Zu´niga˜ Andrade, 1990). Consequently, Honduras has a large diversity of production situations (De Wit, 1982). Honduran coffee growing was thus a good sphere for our survey.

**Paper: Prediction of Coﬀee Rust Disease Using Bayesian Networks (USEFUL) [Perez Prediction]**

Cora B. P´erez-Ariza Universidad de Granada, Spain (cora@decsai.ugr.es)

Ann E. Nicholson Monash University, Australia (ann.nicholson@monash.edu)

M. Julia Flores University of Castilla-La Mancha (julia.ﬂores@uclm.es)

**Abstract:**  
In this paper we present an agricultural case study for learning Bayesian networks (BNs), namely prediction of coﬀee rust. Wide-spread in all major production areas, coﬀee rust causes premature defoliation, weakening the plant and reducing subsequent yield. It is typically controlled by use of chemical fungicides which must be applied before symptoms of infection are observed.

Improved prediction would reduce the use of fungicides, producing healthier quality product and decreasing both economic costs and environmental impact.

We use a dataset obtained from an experimental farm in Brazil over 8 years. Our preliminary data analysis informed our pre-processing of the original dataset. We also identiﬁed several structural priors, which our **BN learner**, **CaMML (Causal Minimum Message Length),** incorporates into its scoring metric and hence into its structure learning.

**Intro:**

Coffee rust has provoked intensive losses in many coﬀee-producing countries. **It attacks all coffee species, being more severe in Coﬀea Arabica**. The **disease ﬁrst appeared in 1970** in the occidental hemisphere, **in Brazil**, and was rapidly extended to other major coﬀee producing countries in **Central and South America**, whose **coﬀee trees are more sensitive to rust infection**. Coﬀee rust is primarily caused by a fungus called Hemileia vastatrix. The symptoms of coﬀee rust infection include the **appearance of spots on the upper leaf surface of plants**. Its **incidence can be measured by the percentage of leaves infected by the fungus.**

The prevalence and impact of coﬀee rust have led to many studies about it, including its prediction and how to decide whether to apply fungicides. **Plant pathologists have tried to characterise this infection forecasting (**Kushalappa, 1990), and more recently **machine learning researchers in Brazil have applied decision trees (Meira et al., 2008), regression Support Vector Machines (SVM) (Luaces et al., 2010) and non-deterministic classiﬁers** (Luaces et al., 2011) (described in more detail in Section 2), to the problem; **but it has not yet been solved**.

**Bayesian networks have been applied to many real applications in environmental sciences, in ﬁelds such as farming, water resources, reforestation and ecological modelling**; agricultural examples include fungicide for mildew in wheat (Jensen, 1995) and growing barley without pesticides (Kristensen and Rasmussen, 2002). Aguilera et al. (2011) have recently reviewed BN applications in the environmental sciences published between 1990 and 2010. **This background encouraged us to apply Bayesian networks to this agricultural case study, the prediction of coﬀee rust.**

With the availability of a **dataset, used by the aforementioned Brazilian machine learning (ML) researchers, our approach here is focused on ML.** The dataset comprises **monthly accounts of the incidence of disease on an experimental farm in Brazil over 8 years.** Additionally, the dataset registers the values **of variables known to stimulate the growth of fungus** (Avelino et al., 2006): weather conditions, fruit load of the plantation, and spacing between plants.

**Previous ML approaches:**

1. In (Meira et al., 2008), the authors develop a decision tree with the aim of aiding the understanding of coﬀee rust epidemics. **The class variable was the change in infection rate,** deﬁned as the **percentage of infected leaves, based on the monthly observations. They used three classes to classify the monthly change in the infection: reduction or stagnation (≤ 0%); moderate growth (> 0 & ≤5%); and accelerated growth (> 5%).** They used **14 predictive variables: space, load, and weather measurements related to temperature, rain and relative humidity.** The model **correctly classiﬁed 73% using cross-validation, while the success rates were 88%, 57% and 79%, respectively**, for these infection rate classes (i.e. the model did poorly on the moderate increase in incidence). The most important explanatory variables were found to **be mean temperature during leaf wetness periods, expected yield, mean of maximum temperatures during the incubation period and relative air humidity.**
2. More recently, these researchers presented a more sophisticated approach to the same problem (and dataset), using SVM regression (Luaces et al., 2010); **rather than predicting the change in incidence, they consider actual incidence**. They trained a SVM that provided a **correlation of about 0.94 between predicted and actual incidences.** However, when they try to devise an alarm system for predicting values above a given threshold, **the number of false negatives turned to be too high (i.e. they were missing too many cases).** So, they implemented **what they call nondeterministic regressors**, changing target points for intervals of ﬁxed width. This allowed them to consider three possibilities: alarm, non-alarm and warnings (within a variable window around 4.5% incidence). This lowered the number of false negatives but, naturally, added false warnings.

**Description of the dataset:**

The data used for this work was obtained from an experimental farm3 (Funda¸c˜ao Procaf´e, Varginha, Minas Gerais, Brazil) where the incidence of coﬀee rust was monitored **from October 1998 to October 2006.** The **farm is organized into eight plots, where the rust is observed in diﬀerent combinations of spacing between plants (dense or thin) and diﬀerent fruit load (high or low).** A **meteorological station reports the temperature, humidity, wind speed, solar radiation and amount of rain registered every 30 minutes**. Some meteorological information was missing in the original database due to failures in the station; **we removed entries that contained missing data** (as did (Meira et al., 2008), (Luaces et al., 2010) and (Luaces et al., 2011)). **The rust observations were taken the ﬁrst day of every month by picking 100 leaves from each plant of each plot and computing the average number of infected leaves.**

We used the following data**:**

**(1)** monthly information of the incidence of the coﬀee rust, and

**(2)** a daily summary of all the weather observations taken that day. After pre-processing, the ﬁnal database comprises 1716 records of 34 variables. **We ignore the time series aspect of the data and treat each record as independent and identically distributed** (although we incorporate aspects of time through the month and year variables).

The incidence of the rust over successive months (plotted in Fig. 3) is highly correlated (correlation coeﬃcient ρId0, Idt = 0.557157).

**We use a threshold of 4.5% incidence to conﬁrm an infection, the threshold typically used by farmers when deciding to chemically treat the plants,**

In this paper we will analyse the behaviour of Bayesian networks for the coﬀee rust prediction problem. The advantages of BNs are

(1) they are **able to capture (in)dependencies existing among the variables involved, and the use of conditional probabilities make them capable of inherently dealing with uncertainty;**

(2) the **graphical representation of relationships between variables facilitates the interpretation and formulation of conclusions about the domain of study**; and

(3) BNs **can combine causal relationships** with **probabilistic logic**, which helps to **incorporate expert knowledge into the model.**

**3 The CaMML BN Learner**

CaMML attempts to **learn the best causal structure to account for the data**, using a **minimum message length (MML) metric** (Wallace, 2005) with a two-phase search, simulated annealing followed by **Markov Chain Monte Carlo search, over the model space.**

**Paper: Local dynamics of the coffee rust disease** and the potential effect of shade

**Paper:** Modelling coffee **leaf rust risk in Colombia with climate reanalysis data**

**Paper:** The **coffee rust crises in Colombia and Central America** (2008–2013**): impacts, plausible causes and proposed solutions**

Abstract:

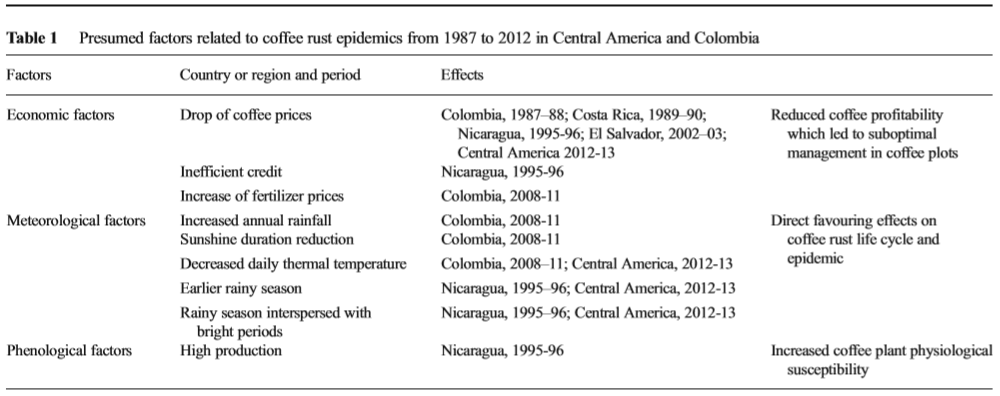
Coffee rust is a leaf disease caused by the fungus, Hemileia vastatrix. Coffee rust epidemics, **with intensities higher than previously observed**, have **affected a number of countries including: Colombia, from 2008 to 2011; Central America and Mexico, in 2012–13; and Peru and Ecuador in 2013**.There are many contributing factors to the onset of these epidemics e.g. the state of the economy, crop management decisions and the prevailing weather, and many resulting impacts e.g. on production, on farmers’ and labourers’ income and livelihood, and on food security.

**Production has been considerably reduced** in Colombia (**by 31 % on average** during the epidemic years compared with 2007) and Central America (**by 16 % in 2013** compared with 2011–12 and **by 10 % in 2013–14** compared with 2012–13). These reductions have had direct impacts on the livelihoods of thousands of small holders and harvesters. As a result, the coffee rust epidemic has had indirect impacts on food security. The **main drivers of these epidemics are economic and meteorological.**

**These epidemics should be considered as a warning for the future, as they were enhanced by weather conditions consistent with climate change.**

**Introduction**

**Coffee rust is caused by the fungus Hemileia vastatrix, an obligate parasite that affects the living leaves of the genus Coffee. Among the cultivated species, C. arabica is the most severely attacked.**

****

Despite coffee rust not being feared by the farmers, the disease had caused significant losses even before the period 2008–2013. Most of these **losses went unnoticed due to an interaction between the disease and the coffee tree phenology**. **For some (not well understood) reason, coffee rust attacks are more severe on high yielding coffee trees**.

In addition, coffee trees exhibit a biennial production rhythm, particularly at full sun exposure high yielding trees normally produce low yields the following year and vice versa. After a high yield year, farmers therefore expect a low yield, with or without coffee rust. However, the reduction in production may not be attributed to the biennial production rhythm alone. Coffee rust may also contribute to this decrease by killing severely infected branches, such that they cannot bear fruits in the following year

**These secondary losses, i.e. losses caused by the current epidemic on productivity in the following years, have probably always existed. Primary losses, i.e. losses caused by the current epidemic during the current year’s production, are less important. Usually, the fastest growth and the peak of the epidemic are during and at the end of the harvest, respectively.**

**Paper: The effect of temperature on the development of epidemics of coffee leaf rust in Papua New Guinea**

J. S. Brown,\* M. K. Kenny,+ J. H. Whan\* and P. R. Merriman\* \*Department of Agriculture, Victorian Institute for Dryland Agriculture, Private Bag 260, Horsham 340 1, Australia, ‘Coffee Industry Corporation, Coffee Research lnstitu te, P. 0. Box 105, Kainantu, Eastern Highlands Province, Papua New Guinea, and \*Depa/tment of Agriculture, Institute for Horticultural Development, Burwood High way, Knox field, Victoria 3 180, Australia

**Abstract:**

Epidemics of coffee leaf rust (CLR) were monitored **to establish the seasonal pattern of epidemic development in various coffee growing regions of Papua New Guinea (PNG) and to determine the effect of temperature on those epidemics.** There is a seasonal pattern to the epidemics of CLR in PNG. Rust incidence (% infected leaves) **is lowest during the October-November to January-February** period. Thereafter rust incidence increases and **reaches a peak in May-June-July,** after which the disease incidence declines. The average maximum disease incidence (MDI) was less than 20% leaves infected. MD1 was positively correlated with the mean minimum monthly temperature five months before the MDI.

This paper is a study.

**Introduction:**

Coffee leaf rust (CLR) (caused by the fungus Hemileia vast&ix Berk & Br.) **was first detected in the main coffee growing regions of Papua New Guinea (PNG) in 1986** (Waller and Turner, 1986). The **major aspects of the epidemiology of CLR** are known from work undertaken in parts of Africa and south and central America, where rust is now endemic. These include *the infection process, incubation period and spore dispersal, and the influence of climate, location and host* (Waller and Turner, 1986).

Although this knowledge **can be used to predict, by extrapolation,** how coffee rust might develop in PNG, *any predictions must remain tentative and await confirmation through direct monitoring of the disease (*Waller and Turner, 1986).

Factors such as **altitude, temperature and rainfall** all play **significant roles** in the development of CLR epidemics by affecting one or several processes of the infection cycle. For instance, in parts of Kenya, Bock (1962) showed that the **disease severity decreased with increasing altitude**. This was *attributed to generally lower night temperatures at higher altitudes because temperatures below 15°C adversely affect the germination and infection process* (Nutman and Roberts, 1963; Kushalappa, Akutsv and Ludwig, 1983; Jong, Eskes, Hoogstraten and Zadock, 1987). Temperature is import- ant because it affects the length.

The **objective of the study** described in this report was **to obtain information on the epidemiology of CLR in PNG**. This is a necessary prerequisite to the development of control strategies. Epidemics of CLR were monitored *to establish the seasonal pattern of epidemic development in various coffee growing regions of PNG and to determine the effect of temperature on those epidemics*.

**Seasonal pattern of epidemic development:**

**At all sites there was only one peak in the CLR epidemic** (the maximum disease incidence, MDI) each year, **and at most sites this occurred in the May-July period**, although the precise timing varied with the locality and year.

**Effect of temperature on CLR epidemics:**

For this analysis MD1 was correlated with various temperature variables. These **included the average monthly minimum (M.MinT) and average monthly maximum (M.MaxT) temperatures**, the *number of days per month on which the maximum temperature was above 30°C (ndMax>30) and the number of days per month on which the minimum temperature was below 15°C (ndMin<15*). The latter two variables were selected because **temperatures below 15°C and above 30°C are detrimental to rust development** (Nutman and Roberts, 1963; Kushalappa and Chaves, 1980).

**Discussion:**

This study has shown that there is a **definite seasonal pattern to the epidemics of CLR in PNG**. The rust incidence is **lowest during the October-November to January-February period**. Thereafter rust incidence **increases and reaches a peak in May-June-July**, after which the disease incidence declines. *The generation of a single epidemic peak in any one year is similar to CLR epidemics in India (Mayne, 1930; cited by Waller, 1982) and Brazil (Kushalappa and Chaves, 1980)* **but contrasts to the situation in Kenya**, especially east of the Rift valley, **where there are usually two peaks in the rust epidemic every year** (Bock, 1962).