Acquisition and Analysis of Soil Parameters for Vine Monitoring using the IoT Sensor Network

Mihaela Hnatiuc, Mirel Paun, Domnica Alpetri

Electronics and Telecommunications Department  
Constanta Maritime University, Constanta, Romania  
mihaela.hnatiuc@cmu-edu.eu

*Abstract*— In recent years, the concept of the Internet of Things (IoT) has spread in most fields due to the multiple benefits it offers, so it recently started to be used in viticulture, together with artificial intelligence (AI). These two innovative technologies combined with classical methods can lead to a much more efficient and rapid prediction of diseases that may occur in grapevine culture. The present study aims to detect diseases in the period when evolution can still be prevented, by minimizing the favorable environment in which the pathogen can develop. This prevention can be achieved using algorithms for predicting and comparing the data from IoT sensors with the data classically gathered by farmers. The IoT sensors system developed to acquire information about environmental parameters is tested in the Murfatlar vineyard, Romania. Two types of untreated Cabernet and Sauvignon Blanc varieties are used in the experimental study with the aim of reducing the chemical treatment concentration. All the sensors' data are stored in a cloud. This paper presents an analysis of sensor data consisting of soil parameters such as temperature, humidity, conductivity, nitrogen (N), phosphorus (P), and potassium (K). Big data algorithms using machine learning (ML) for clustering and prediction are tested to identify the vine diseases. The results are detailed at the end of the paper.

Keywords— nutrient sensor, pH sensor, silhouette parameter

# Introduction

Disease detection is an intensive area of research in viticulture. Given the significant impact and economic costs of diseases, it is important to automate the early detection of these diseases in vineyards. Disease occurrence is dependent on ever-changing environmental factors, nutrient fluctuations, and prevention or control technologies, as the case may be. The disease occurs when the pathogen encounters the right organism in an environment favorable for its development and multiplication. The main mineral substances that must be considered in the crop plant are nitrogen, phosphorus and potassium. In addition to these, the amount of humus, water and acidity (expressed in pH) must be controlled in the soil. These are the important elements for most plants, and the deficiency or excess of one or more of the substances mentioned above can make the difference between a profitable crop and a harmful one [1, 2]. The concentration of nutrients is variable depending on the atmospheric conditions and the soil depth [3, 4]. The sensor data is compared with data collected using conventional methods in order to check the data accuracy [5].

Among the technologies implemented in Precision Agriculture (PA), along with the advantages it brings, we list: artificial intelligence and machine learning that offer intelligent software applications based on learning algorithms using training models; cloud computing that integrates computing power into cloud servers and storage options for different types of data; Internet of Things (IoT) that automatically provides data exchange between the devices we use. This field is constantly developing due to the advent of devices that have the possibility of being connected to the Internet, making the way to obtaining access to a multitude of very attractive datasets. With the introduction of digitized methods for agriculture, the modernization of rural communication networks has begun [6, 7].

The occurrence of the diseases analyzed in this work is influenced by humid environments, with average precipitation around 510 mm/m2, with average daily temperatures between 14-16℃, the lush vegetation of the stumps, the excessive development of weeds, the high density of the foliage. These aspects will be well observed to help identify diseases in time to combat or prevent them, depending on the intensity of the attack.

WSN (Wireless Sensors Network) is a network of connected sensors consisting of several nodes. The sensors take data, such as temperature, humidity, wind speed, etc., and send the information to the processing unit. Each node consists of an antenna, microcontroller, and an interface circuit that serves as a communication, processing and detection unit. To power the node, solar panels are used in combination with batteries, for more autonomous network management. Another important element proposed is an SD card, which allows to record the data locally. For the monitoring sensor of fluctuating concentrations of nutrients (nitrogen, potassium and phosphorus) we used SNiP-MP4.

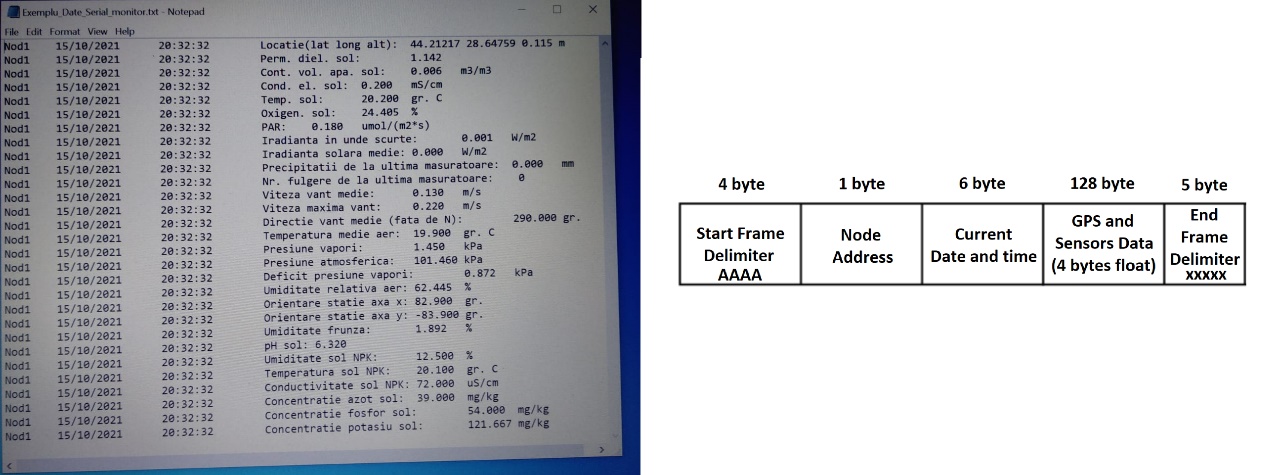
The next sections of the paper present the data communication and the results of data analysis.

# Data Communication

The sensors system using IoT technology, developed in the laboratory for vine monitoring, was deployed in the Murfatlar vineyard in august 2022. The system consists of nodes directly connected to sensors. The data from the nodes are sent to the cloud server by means of a gateway device. The communication protocol between the node and the gateway is Long Range (LoRa) radio technology operating on 868 MHz. Two high-level communication protocols operating over LoRa have been designed and tested specifically for this system.

The first one is intended for broad compatibility with third-party devices and sends the sensors data in human-readable format as ASCII text messages.

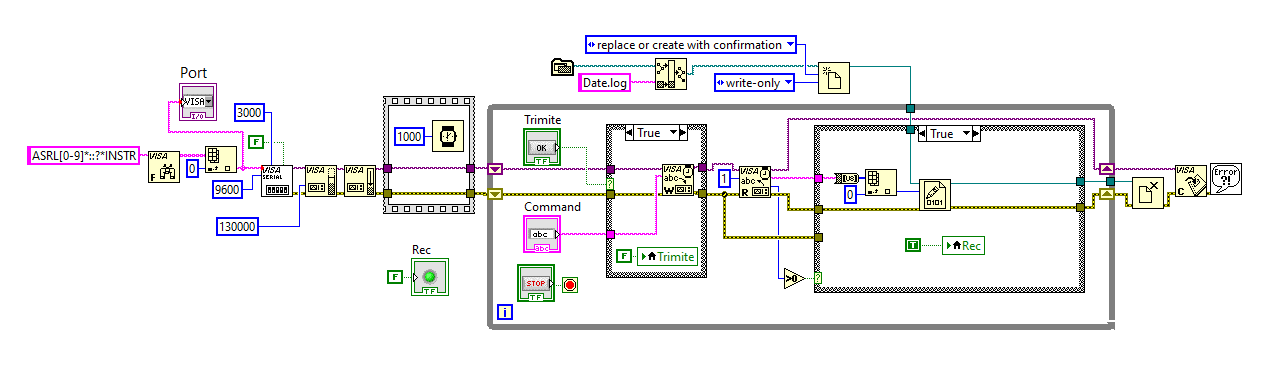
The second protocol sends the sensors data in position-encoded binary format using fixed frames of 144 bytes, as presented in Fig. 2. The advantage of the last one is the reduction of data flow by an order of magnitude compared to the human-readable protocol, thus reducing the energy consumption and the channel dwelling of the radio network.



|  |
| --- |
| 1. Data communication protocol. |

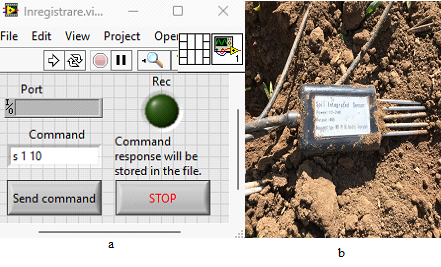
The protocol doesn’t use a redundancy check (CRC), because it relies on the CRC performed by the lower layer of the LoRa implementation existing in the RFM98 modem.

As an anti-theft measure, the nodes have GPS modules for locating them in case of disappearance. GPS localization was preferred over LoRa localization because it offers greater accuracy and works even for a network consisting of only one node and a gateway, contrary to LoRa localization, which needs at least 3 devices [8]. The LoRa gateway is connected to a PC for local storage or internet access. The PC program is implemented in LabVIEW 2013, as shown in Fig. 2 and Fig.3a



1. LabVIEW gateway implementation

The sensors used in the data analysis are NPK sensors which give data about the soil parameters. For the monitoring sensor of fluctuating concentrations of nutrients (nitrogen, potassium and phosphorus) we used CWT-SOIL, shown in Fig. 3b.



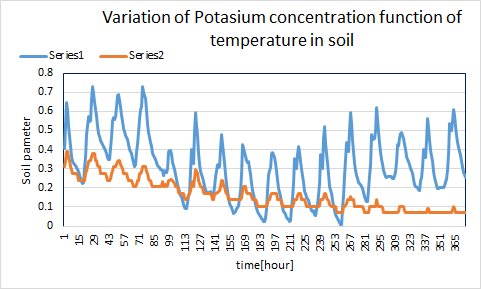
1. LabVIEW gateway interface a) and soil sensor b)

We mention that the sensors are deployed in the field by means of a non-invasive method in order not to disturb the natural development processes of the plant and inadvertently limit the occurrence of monitored diseases.

The next Section has presented the influence of the soil parameters on the occurrence of the disease.

# Results

The nutrient concentrations depend on the soil temperature and humidity Fig.4.

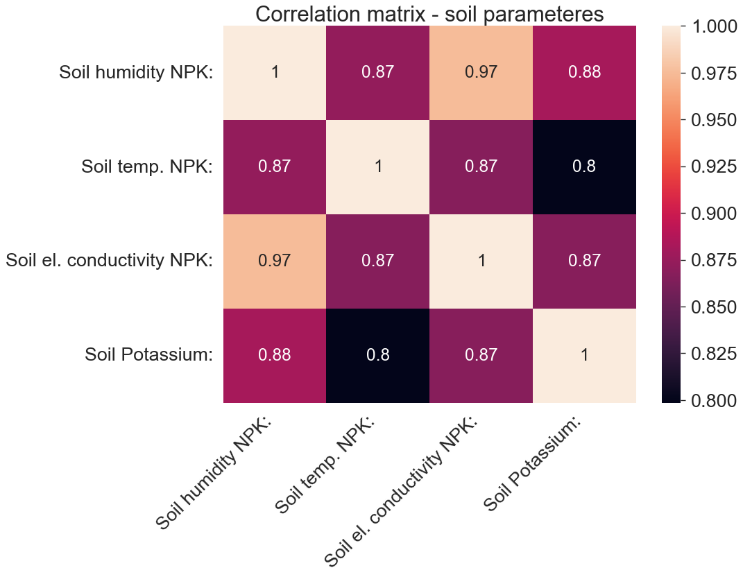


1. Fig.4. The variation of potassium contraction (orange) function of temperature (blue)

Note: Series1- soil temperature [0C]

Series2 – potassium concentration [mg/kg]

The soil data are correlated using the temperature, humidity, potassium concentration and conductivity (Fig.5) in order to identify the data which gives the most relevant information about the vine. The processing aims to identify two diseases using the sensors placed on the Cabernet Sauvignon vine.



1. Data correlation of soil parameters

The correlation between the soil nutrients like Potassium concentration and soil temperature and humidity is observed in Fig.5

The environmental parameters and nutrients can be used in the prediction algorithm to identify the disease.

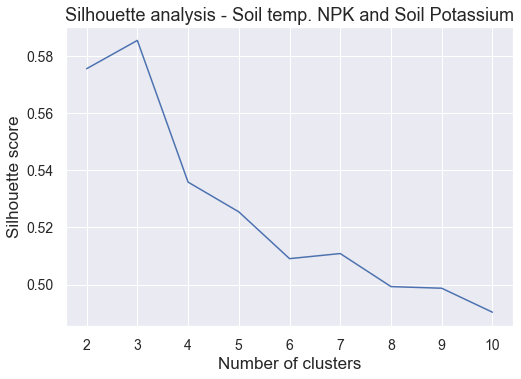
There are a number of internal measures for assessing data clustering quality. However, it should be taken into account that many of these measures are based on the evaluation of consequent clustering, whereas only one potential cluster is known when evaluating.

The Calinski-Harabasz index is a measure of the quality of a dataset's splits in clustering. This index is based on the between-groups and within-groups variances. The distance to the cluster center is presented in [9]. Another algorithm of clustering is the Dunn index and the Davies-Bouldin Index. This compares the cluster spacing to the cluster diameter. The Index determines the average similarity between a cluster and its nearest neighbors. These indices are computed using inter-cluster or distance-to-cluster-center values.

The Silhouette index is a method for interpreting and checking the consistency within clusters. The silhouette value measures an object’s similarity with its own cluster [9].

This study is presented the computing test for data clustering. The processing aims to identify two diseases using the sensors placed on the Cabernet Sauvignon vine. Samples of soil parameters data for predicting the vine disease are used as initial data. These parameters are soil temperature [deg. C] and soil Potassium [mg/kg] measured with NPK (Fig.6). Clustering methods are used to identify the number of data groups of diseases. The silhouette parameter obtained for the three classes is 0.58 (Tabel I).

The soil data are correlated using the temperature, humidity, potassium concentration, and conductivity (Fig.7) in order to identify the data which gives the most relevant information about the vine. After the cluster data, we obtain the same number of classes that represent the normal and disease vine as the classical data collected by the farmers.

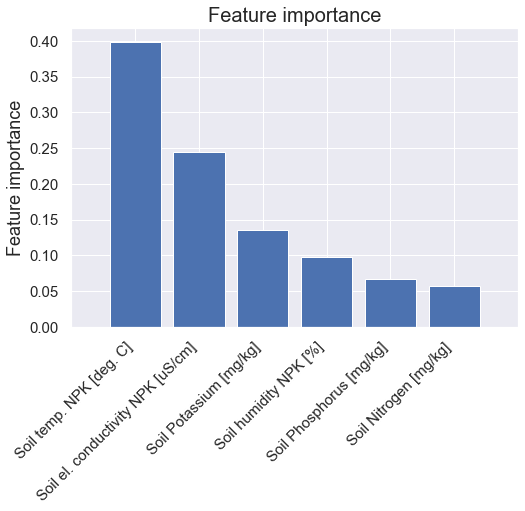


1. Silhouette analysis for soil temperature [deg. C] and soil Potassium [mg/kg]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Inertia | Silhouette Score | Calinski-Harabasz Index | Davies-Bouldin Index |
| K-Means - Soil temp. NPK and Soil Potassium | 0.55 | 0.58 | 1227.92 | 0.49 |

Tabel I. Clusters method used to identify the number of diseases that can be identified with soil parameters. Performance metrics for K-Means models (03.08.2022 - 30.09.2022)

Feature importance (variables) indicates the contribution of each feature to the prediction. Determines the degree of importance of a specific variable on the feature importance ranks of a classifier[10]. The measure of this method is score. The higher the score value, the more important that variable is processed with the Random Forest method say which features of data are most useful for the objective, which can be both classification and regression.



1. Feature importance for soil parameters

The regression and classification algorithms are tested for identifying the attack degree and disease occurrence. The degree of attack is observed using a classic method in the period of May – September 2022. These methods consist of leaves analysis on the field. The number of spots on the leaves is computed and correlated with the degree of attack. The number of samples used for training is 1460 and for testing is 20. The parameters used are Soil temp. NPK [deg. C] and Soil el. conductivity NPK [uS/cm] because in the feature importance, these have a big score. The performance obtained after the regression algorithm is presented in Table II. In regression, we used the leaves attached by Plasmopara Viticola and Botrytis Cinerea disease.

Table II. Performance metrics for regression models (15.05-22.09 in 2022)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | R^2 | MAE | MSE | RMSE |
| XGBoost Regressor | Deg. of Attack - Plasmopara | 0.902 | 0.019 | 0.0017 | 0.042 |
| Deg. of Attack -Botrytis | 0.917 | 1.044 | 3.27 | 1.808 |
| Random Forest Regressor | Deg. of Attack - Plasmopara | 0.926 | 0.015 | 0.0013 | 0.036 |
| Deg. of Attack -Botrytis | 0.921 | 1.023 | 3.103 | 1.761 |
| LSTM | Deg. of Attack - Plasmopara | 0.706 | 0.043 | 0.005 | 0.064 |
| Deg. of Attack -Botrytis | 0.913 | 1.703 | 3.405 | 2.323 |
| Deep Neural Network | Deg. of Attack - Plasmopara | 0.762 | 0.036 | 0.004 | 0.065 |
| Deg. of Attack -Botrytis | 0.886 | 1.373 | 4.501 | 2.121 |
| Support Vector Regression | Deg. of Attack - Plasmopara | 0.633 | 0.073 | 0.006 | 0.081 |
| Deg. of Attack -Botrytis | 0.607 | 2.66 | 15.549 | 3.943 |

The R^2 represents the coeficient of determination. It is the regression sum of squares divided by the total sum of squares.

Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are metrics for evaluateing a Regression Model. These metrics give information about the accury of prediction, what is the amount of deviation from the real values. For the regression, we used different methods such as XGBoost Regressor, Random Forest Regressor, Long Short Term Memory  (LSTM), Deep Neural Network, and Support Vector Regression [11]. The data we used is nonliniar, as such, the tree based algorithms perform best. Neural networks are also capable of capturing nonliniar relationships, but they require more data in order to train effectively. The Support Vector Regression (SVR) is the poorest performing of the three, as it assumes a liniar relationship between the features and the targets.

# Conclusion

The present study has the goal of disease identification using soil parameters. The feature importance data are used to identify the most influential parameter in the disease occurrence. The leaves give information about the disease occurrence and the percentage of them. So using the regression methods we identify two diseases that can occur due to the soil temperature, conductivity, and nutrients. In the next study, we achieve an automatic system for data classification for disease identification. These systems will be based on the soil and air parameters that influence the disease occurrence.

##### Acknowledgment

This work is supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CCCDI - UEFISCDI, project number 203, COFUND-ICT-AGRI-FOOD-MERIAVINO-1, within PNCDI III.

##### References

1. D. B. Bhoyar, S. K. Mohod, S. S. Khade, R. A. Burange, “Detection of NPK and pH components of soil”, International Journal of Scientific and Research Publications, ISSN: 2250-3153.
2. K. Misbah, A. Laamrani, K. Khechba , D. Dhiba, A. Chehbouni, “Multi-Sensors Remote Sensing Applications for Assessing, Monitoring, and Mapping NPK Content in Soil and Crops in African Agricultural Land”, Remote Sens. 2022, 14, 81, 2022, DOI: 10.3390/rs14010081.
3. J. Lin, M. Wang, M. Zhang, Y. Zhang, L. Chen, “Electrochemical Sensors for Soil Nutrient Detection: Opportunity and Challenge”, International Conference on Computer and Computing Technologies in Agriculture CCTA 2007: Computer And Computing Technologies In Agriculture, Volume II, pp. 1349–1353, 2007.
4. S. G. Lemos, E. A. Menezes, F. S. Chaves, A. R. A. Nogueira, A. Torre-Neto, A. Parra, and J. Alonso, “In Situ Soil Phosphorus Monitoring Probe Compared with Conventional Extraction Procedures”, Communications in Soil Science and Plant Analysis, 40, pp. 1282–1294, 2009, ISSN 0010-3624.
5. S. Ghita, M. Hnatiuc, A. Ranca, V. Artem, M.A. Ciocan “Vegetation Dynamics, Changing Ecosystems and Human Responsibility”, Studies on the Short-Term Effects of the Cease of Pesticides Use on Vineyard Microbiome, Intechopen, 2022, DOI: 10.5772/intechopen.105706
6. M. Wu, T. J. Lu, F.Y. Ling, “Research on the architecture of Internet of Things”, Conference: Advanced Computer Theory and Engineering (ICACTE), Vol. 5, 2013.
7. G. Shen; B. Liu, “The visions, technologies, applications and security issues of Internet of Things”, 2011 International Conference on E-Business and E-Government (ICEE), DOI: 10.1109/ICEBEG.2011.5881892, 2011.
8. K. H. Lam, C. C. Cheung, W. C. Lee, “LoRa-based localization systems for noisy outdoor environment”, 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Rome, Italy, 2017, pp. 278-284, doi: 10.1109/WiMOB.2017.8115843.
9. E. Sivogolovko, B. Novikov, “Validating cluster structures in data mining tasks”, In EDBT-ICDT'12; Association for Computing 804 Machinery: New York, NY, USA, 2012, 245–250.
10. J. Jiarpakdee, C. Tantithamthavorn, and A. E. Hassan, “The impact of correlated metrics on the interpretation of defect models,” IEEE Transactions on Software Engineering, 2019

[11] E.Mooi, “Regresion Analysis”, Chapter 7 in Book “A Concise Guide to Market Research. The Process, Data, and Methods Using IBM SPSS Statistics”- 3rd Edition, DOI: 10.1007/978-3-642-53965-7\_7, 2014