REPORT ON ESTIMATING RICE FIELD QUALITY USING ARTIFICIAL INTELLIGENCE

Yahjeb Bouha Khatraty\textsuperscript{1,2}, Nedra Melloulı Nauwynck\textsuperscript{1}, Mamadou Tourad Diallo\textsuperscript{2} and Mohamedade Farouk Nanne\textsuperscript{2}

\textsuperscript{1}University of Paris 8 - Vincennes - Saint-Denis Doctoral School Cognition, Language, Interaction, Paris, France
\textsuperscript{2}University of Nouakchott Al-Aasriya, Faculty of Science and Technology, Doctoral School of Computer Science, Nouakchott, Mauritania

ABSTRACT
The use of artificial intelligence has made life easier for farmers, as analysis of agricultural data allows farmers to make informed decisions, supported by large data sets and processed by machine learning algorithms. In this paper, we present the literature on rice quality assessment according to the different criteria and factors used to predict the quality or yield of rice. Thus, we provided a list of machine learning and deep learning models that had high accuracy in predicting rice yield and quality. In addition, due to the lack of literature of a real-time rice field monitoring system, we proposed a real-time rice field condition monitoring system that collects weather and soil condition data and sends them to the web server. The farmer can see the status of its remotely field and can automate certain tasks such as irrigation based on the collected data values.

KEYWORDS
Machine Learning, Internet of Things, Intelligent rice paddies, Grain Yield Prediction

1. INTRODUCTION

Today, water scarcity is a serious problem, with fresh, clean water accounting for only 0.003% of the world’s total available water, and the increase in the world’s population highlights the need to tackle this problem\textsuperscript{1}.

Agriculture is the highest consumer of water\textsuperscript{1}. Therefore, improving water management in agriculture is essential to preserve the source of life on Earth. Since rice consumes more water than any other crops, so to improve water management, it is necessary to reduce water consumption in rice fields while maintaining rice quality (Bouman 2007; Khush 2005).

To maintain good rice quality, it is essential to know the features and factors that affect it. These features change from the perspective of the farmer, the consumer and the trader. In addition, there are many factors that affect it, such as climatic factors, geographical factors and those related to soil and water supply (Dawn 2001; Pheakdey, Xuan, and Khanh 2017; Wang et al. 2019).

This paper provides a review of the literature on rice quality estimation. Section 2 contains related work. Our Proposed approach is introduced in Section 3. Section 4 concludes and gives perspectives in this research area.

\textsuperscript{1}\url{https://www.theworldcounts.com/challenges/planetearth/freshwater/are-we-running-out-of-water/story}
2. RELATED WORK

Table 1. The models and parameters used in the literature to estimate rice quality and yield and their results

<table>
<thead>
<tr>
<th>Objective</th>
<th>Parameter used</th>
<th>Data type</th>
<th>Model used</th>
<th>Result</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection and classification of rice leaf diseases</td>
<td>Color image, Color, texture and shape</td>
<td>Image</td>
<td>Naive Bayes algorithm</td>
<td>Accuracy reached 90%</td>
<td>(Islam et al. 2018)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Support Vector Machine SVM</td>
<td>Accuracy reached 73.33%</td>
<td>(Prajapati, Shah, and Dabhi 2017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>k-Nearest Neighbor KNN</td>
<td>Accuracy reached 87.02%</td>
<td>(Joshi and Jadhav 2016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Decision tree, logistic regression, KNN and Naive Bayes Classifier</td>
<td>Accuracy was 94.9% for training and 97.9% for testing</td>
<td>(Ahmed et al. 2019)</td>
</tr>
<tr>
<td>Estimation of the number of spikelets in a panicle</td>
<td>45 characteristics of the spikelets and grain such as length, width and surface of the spikelets</td>
<td></td>
<td>Area Method, Faster R-CNN, Cascade R-CNN and SSD</td>
<td>- Faster R-CNN $R^2 = 0.99$ and MAPE &lt; 1.42%</td>
<td>(Yu et al. 2020)</td>
</tr>
<tr>
<td>Classification of indica and japonica rice</td>
<td></td>
<td></td>
<td>Resnet-50, Faster R-CNN, Cascade R-CNN, SSD and SVM</td>
<td>Resnet attained the highest accuracy of 0.96</td>
<td>(Guo et al., n.d.)</td>
</tr>
<tr>
<td>Measuring the correlation between rice yield and phenological, climatic and geographical data</td>
<td>The dates of sowing, transplanting, maturity, Temperature, humidity precipitation, longitude, latitude, altitude ...</td>
<td>Phenological, climatic and geographical data</td>
<td>Random Forest RF, MLR (Multiple Linear Regression), BP (Back Propagation) neural network and SVM</td>
<td>$R^2$ was 0.24 for BP, 0.33 for SVM and 0.31 for RF</td>
<td>(Guo et al., n.d.)</td>
</tr>
<tr>
<td>Estimation of rice dry biomass</td>
<td>Polarisation (VH,VV or VHVV)</td>
<td>Satellite data (Sentinel1A)</td>
<td>Random Forest RF, Gradient Boosting Decision Tree (GBDT), SVM and KNN</td>
<td>RF ($R^2 = 0.73$ and RMSE = 462.4 g/m²) and k-NN ($R^2 = 0.70$ and RMSE = 484.1 g/m²)</td>
<td>(Mansaray, Zhang, and Kanu 2020)</td>
</tr>
</tbody>
</table>

Related work can be divided into two categories. The first category highlights the evaluation models of the rice quality fields. The second category focuses on the estimation of rice yield during the paddy growing stages, or after rice production. For the first category, the parameters used for estimating quality are the characteristics of the rice leaves such as color, shape and texture to determine whether the leaf is diseased or healthy, as has been done in these articles (Ahmed et al. 2019; bank 2020; Islam et al. 2018; Joshi and Jadhav 2016; Prajapati, Shah, and Dabhi 2017; Shrivastava et al. 2019; Yao et al. 2009).

But the most common diseases in the major rice-producing countries (we can cite China, India, Indonesia and Bangladesh) are only four, such as Rice Blast, Rice Bacterial Blight, Rice Brown Spot and Rice Sheath Blight (Ahmed et al. 2019; bank 2020; Islam et al. 2018; Joshi and Jadhav 2016; Prajapati, Shah, and Dabhi 2017; Shrivastava et al. 2019; Yao et al. 2009). Each of these diseases has different symptoms from the others. Hence manual early detection of these diseases is very difficult and imprecise. Therefore, it is mandatory to use smart technologies that enable accurate and rapid disease detection and hence to overcome the limitations of human perception and treat plants as soon as possible.
For the second category, the parameters used to predict rice yield differ according to the prediction phase. Several parameters were used to predict rice yield during growth and post-production. Panicle length was used in (Huang et al. 2013) to estimate rice yield. However, (Mansaray, Zhang, and Kanu 2020) was based on climatic parameters, phenological variables, and geographical criteria to predict rice yield. calculated the relationship between dry rice biomass and Sentinel-1A data. Whereas (Kumar et al. 2020) used grain color and the length of time the grain retains heat to predict rice yield. (Yu et al. 2020) extracted the phenotype of rice clusters (length, width, area, circumference, number of spikes, grains, etc.) to calculate the number of spikelets in a panicle and classify the different types of rice.

The use of machine learning and deep learning to improve the quality and accuracy of crops is becoming a management task known as precision farming (Triantafyllou et al. 2019). Measurements and analysis carried out with machine learning models enable better management of agricultural resources while reducing environmental impact and waste. In Table 1, we list some of the models used in the literature estimating rice yield as quality and their impacts.

We noticed that the studies already carried out were based on only one type of rice quality parameters without considering other parameters and external factors such as climate, geographical location, soil condition, demographics, and water supply. As presented in Table 1, while studies have relied on physical parameters such as color, texture, shape of rice leaves to detect and classify rice diseases. Other studies have also used the physical properties of the spikelets and grains to classify different types of rice.

Therefore, a good yield prediction and quality measurement system must take into account all factors that directly and/or indirectly influence rice quality.

3. PROPOSED APPROACH

Based on our literature, we observed that not all parameters describing rice quality are used in the same study, and that there is also a lack of heterogeneity in data acquisition sources. Consequently, there is no system that monitors the rice fields in real time to act if necessary. Our approach is therefore to create a system that monitors the condition of the plants, soil and atmosphere in real time, from the sowing of the grain to harvesting. Farmers will be able to access this data remotely. In addition, the output of our monitoring system is the input to another system that predicts the yield and quality of rice produced without having to hull the rice spikelets. The forecasting system uses a wide range of parameters such as spikelets properties, climate, phenological, and geographical data. In addition, this system should also allow the difference between a large type of rice. Details of the forecasting systems will be discussed in a future paper.

Figure 1 shows our real-time rice field monitoring system that we detail its components in the next paragraph.

Our monitoring system consists of four layers:

- The sensor layer: includes smart sensors for data collection and monitoring.
- The network layer: includes the communication technologies between the sensors and the gateway. The choice of these technologies depends on the location of the agricultural field (rural or urban area). We chose to work at first with Lora technology because of its low energy consumption and its long range.
- The service layer: responsible for processing and analyzing the collected data.
- The dashboard layer: allows the visualization of the information from the sensor network. This layer is the one that has a direct interaction with the user (the farmer).

A simpler example offered by our system to the farmer is the smart watering system. This system allows the farmer to remotely control and adapt the irrigation with a single touch on the mobile phone. Also, it allows to program the irrigation according to the temperature and/or humidity values of the plant environment.
4. CONCLUSION AND PERSPECTIVES

In this paper, we have investigated the measured parameters and machine learning models used in rice quality and yield estimation. In addition, we presented the models that had high accuracy with their results. We then proposed an approach based on the lack of literature, such as the implementation of a real-time rice field monitoring system that will provide farmers with a remote monitoring application for their rice fields. In addition, it will allow them to automate the irrigation of the fields according to the measured values of the field. The data collected from the rice fields will be used as a dataset for prediction systems of rice quality and yield based on machine learning algorithms.

ACKNOWLEDGEMENT

We would like to thank the French Embassy in Mauritania for its financial support for our research.

REFERENCES


