Development of knowledge based model for the diagnosis of sorghum diseases using Rule-Base approach

Catherine Aaron 1, * and Yusuf Musa Malgwi 2

1 Department of Computer Science, School of Science and Technology, Federal Polytechnic Mubi, Adamawa State, Nigeria.
2 Department of Computer Science, Faculty of Physical Sciences, Modibbo Adama University, Yola, Adamawa State, Nigeria.

World Journal of Advanced Research and Reviews, 2022, 16(01), 198–208

Publication history: Received on 30 August 2022; revised on 04 October 2022; accepted on 07 October 2022

Article DOI: https://doi.org/10.30574/wjarr.2022.16.1.0997

Abstract
The nuances of semantics, diction, and intent are too broad for a computer to understand directly. Instead, humans give the computer a set of rules which define the way to process incoming information. Hence, expertise in this particular field is apparently essential to help diagnose crop pest and disease whilst recommending a proper treatment or control procedure to farmers affected by the incident. Damage occurs, both on the field during the cultivation process and warehouse storage. These conditions will significantly affect the income of farmers and the world's food supply. The aim of this research is to develop a knowledge-based model for prediction and prescription of sorghum diseases using the concept of Machine learning techniques (K- nearest neighbor, KNN) algorithm. The specific objectives include: To seek, create and preprocess dataset, To develop the front end of the system using Java programming language and the back end using PHP, To Develop and implement the K-nearest Neighbour algorithm, Train and test the Model, To validate the Model using Confusion matrix in order to determine the prediction accuracy, The software will bring an end to the manual method of application which leads to applying the wrong insecticide, fungicide etc. to the wrong crop and in a wrong quantity, It will be beneficial to the farmers in Adamawa State and their experts. Plant diseases are the most important reason that leads to the destruction of plants and crops. Detecting that disease at early stages will enable us to overcome and treat them appropriately. The process requires an expert to identify the disease, describe the method of treatment and protection. Identifying the treatment accurately depends on the method that is used in diagnosing the diseases. An expert system was developed using two different methods of plant diagnosis: Step by step description and graphical representational methods. The simulation was carried out using JAVA Net beans for windows.

Keywords: Knowledge based; Model; Diagnosis; Expert system; Disease; Rule base

1. Introduction

Rule-based systems (also known as production systems or expert systems) are the simplest form of artificial intelligence. A rule-based system uses rules as the knowledge representation for knowledge coded into the system [1]. The definitions of rule-based system depend almost entirely on expert systems, which are system that mimic the reasoning of human expert in solving a knowledge intensive problem. Instead of representing knowledge in a declarative, static way as a set of things which are true, rule-based system represent knowledge in terms of a set of rules that tells what to do or what to conclude in different situations. Computers cannot speak and understand language in the same way humans do [2]. The nuances of semantics, diction, and intent are too broad for a computer to understand directly. Instead, humans give the computer a set of rules which define the way to process incoming information. While there are specific forms of rule-based systems in which the computer assists in the defining of its rules, however that machine learning approach is often thought to be separate when understanding rule-based systems. In order to
accomplish feats of apparent intelligence, an expert system relies on two components: a knowledge base and an
inference engine [1]. A knowledge base is an organized collection of facts about the system’s domain. An
inference engine interprets and evaluates the facts in the knowledge base in order to provide an answer. Typical tasks for expert
systems involve classification, diagnosis, monitoring, and design, scheduling, and planning for specialized endeavors.
Facts for a knowledge base must be acquired from human experts through interviews and observations. This knowledge
is then usually represented in the form of “if-then” rules (production rules): “If some condition is true, and then the
following inference can be made (or some action taken).” An expert system may display the sequence of rules through
which it arrived at its conclusion; tracing this flow helps the user to appraise the credibility of its recommendation and
is useful as a learning tool for students [2]. Sorghum is truly a versatile crop that can be grown as a grain, forage or
sweet crop. Sorghum is one of the top five cereal crops in the world. The United States is the world’s largest producer
of grain sorghum, having produced 373 million bushels in 2020. Sorghum is among the most efficient crops in
conversion of solar energy and use of water and is known as a high-energy, drought tolerant crop that is
environmentally friendly. Due to sorghum’s wide uses and adaptation, “sorghum is one of the really indispensable
crops” required for the survival of humankind [1].

Incidence of crop pest and disease is one of the major concerns faced by farmers. For example, in late 2004, soybean
rust (SBR) had largely threatened the soybean crops in the United States, thus reducing its soybean yields. Such
aforementioned incidence can present negative impacts on the quantity and quality of any crop production. In dealing
with this sort of incidence, farmers initially necessitate to useful advices for diagnosing the various pest or disease
confronted before being able to implement a suitable treatment or control onto the affected crop. On the other hand,
much of the literatures have addressed that farmers undertake a prolonged period of time to acquire appropriate advice
for conducting the suitable treatment or control, due to issues of human resource and knowledge. Besides, with
inappropriate and perhaps misleading advice gained by farmers, it might escalate to the use of harmful substances such as
pesticides without having regards of the pesticides proper use and protocol, and subsequently leading to matters like
environmental pollution [3]. Hence, expertise in this particular field is apparently essential to help diagnose crop pest
and disease whilst recommending a proper treatment or control procedure to farmers affected by the incident.
However, most common farmers still lack the convenience of receiving useful consultation from experts due to the few
different constraint and with certain cases; an expert may not be readily available to carry out the diagnosis. Therefore,
the use of ICT such as Artificial Intelligence (AI) can be rendered as an effective tool to treat a certain problem of a
specialized area as an expert would do by emulating human intelligence. Ultimately, an expert system has the advantage
of speeding up pest and disease diagnosis [3]. Research works focusing on advisory expert system for supporting
agriculture practitioners is yet limited. Likewise, the development of an expert system and previous research works
alike in relations to diagnosing and managing the pest and disease problems of a certain crop is also scarce in spite of
the expert system being able to aid farmers, agriculture officers and human experts. Moreover, the majority of expert
system at hand concentrated on a single crop (usually tomato) or mainly one aspect of the crops problem e.g. fungal
diseases or nutritional disorders. Plant pathology involves the study of pathogen identification, disease etiology, disease
cycles, economic impact, plant disease epidemiology, plant disease resistance, how plant diseases affect humans and
animals, pathosystem genetics, and management of plant diseases [4]. The major objective of this Paper is to Model and
represent knowledge acquired from domain experts and codified sources. Build prototype knowledge-based system
which advises experts in the domain. Evaluate the performance of the modeled knowledge-based system.

2. Literature Review

2.1. Review of Related Works

Abu-Saquer & Abu-Naser [5] carried out a research to predict the infected area of the leaves by applying k-means
clustering algorithm and the Otsu’s classifier. Both the shape and texture features were extracted in the proposed work.
The shape oriented features that were extracted in the work included area, color axis length, eccentricity, solidity and
perimeter, whereas the texture oriented features were contrast, correlation, energy, homogeneity and mean. And lastly,
classification in there research was done using a neural network based classifier.

Aravind, Raja, Mukesh, Anirudh & Ashiwin [6] studied maize crop diseases for automatizing the plant disease detection
system. Features known as Speeded Up Robust Features (SURF) were extracted from each image. The features were
clustered using k-means algorithm. Two methods were used for feature extraction namely, histogram and GLCM. These
two methods were used for studying various textural features. For classification, multi-class SVM was applied based on
various kernel functions like linear, polynomial and Radial Basis Function, etc. A best average accuracy of 83.7% was
achieved using the SURF features in the proposed work.
Elsharif & Abu-Naser [7] describe a Software system for disease detection based on the infected images of various rice plants. Images of the infected rice plants using digital camera are captured and then processed using techniques like image growing, image segmentation and zooming to detect infected parts of the plants. Then infected part of the leaf has been used for the classification using neural network. The methods employed in this system are both image processing and soft computing technique.

Alqassas & Abu-Naser [8] proposed method to Accurate Detection and Classification of Plant Diseases In the first step it identify green Colors pixels. Then pixels are masked based on particular threshold values that are obtained using Otsu’s method, and then mostly green pixels are masked. The other additional step is that the pixels with zeros red, green and blue values and the pixels of infected clusters from the boundaries were completely removed. SGDM matrix generated for H and S, and then GLCM Function is called to calculate the features. The experimental results demonstrate that this technique is a powerful technique for the detection of plant leaves diseases.

Galala [9] Surveyed on Remote Area Plant Disease Detection Using Image Processing. In the paper, a method is proposed for detection of disease in malus domestica using methods like k-mean clustering, texture and color analysis. Algorithm used for texture segmentation is CCM method. As RGB images of leaves are converted into HSI color space representation .Then this is used to generate co-occurrence matrix (CCM). By comparing texture and color images plant diseases can be detected.

Husnain, Abdur, Tehseen & Muhammad [10] used an algorithm which uses image processing for disease spot segmentation in plant leaf. In the first phase automatic detection and classification of plant diseases is done. Color transform of RGB image is performed for better segmentation of disease spots. For image smoothing median filter is used. Otsu method is used to calculate the threshold. An algorithm which is independent of background noise, plant type and disease spot color was developed and experiments were carried out on different “Monocot” and “Dicot” family plant leaves with both, noise free (white) and noisy background. In the paper a comparison of the effect of CIELAB, HSI and YCbCr color space in the process of disease spot detection is done.

Karol, Gulhane & Chandiwade [11] used the image recognition of two kinds of grape diseases (grape downy mildew and grape powdery mildew) and two kinds of wheat diseases (wheat stripe rust and wheat leaf rust) was conducted by using image processing technologies and BP networks. Based on the data of the extracted color features, shape features and texture features from disease images and their combined features, BP networks constructed by using different function combinations were used as the classifiers to identify grape diseases and wheat diseases, feature combinations respectively. The image recognition of plant diseases using BP networks was also conducted based on the dimension-reduced data that were obtained by using PCA to process the data.

2.2. Population size

The target population of this study comprises of all the farmers of sorghum crop in Mubi North LGA of Adamawa State. The sample size of the study was drawn from the target population using Yamane’s Formula determine the sample size for the population of the farmers that farms sorghum crop.

\[
\text{The Yamane’s Formula is } n = \frac{N}{1 + N(e)^2}
\]

Where;

- \(n\) = sample size
- \(N\) = population size
- \(e\) = level precision or sampling of error, which is 0.05

To this end the research study using systematic sampling. Some farmers were used in Mubi Local Government of Adamawa State.

2.3. Data collection Techniques

The techniques used for data collection in this work are the secondary and the primary data that were collected from journals, reports, newspapers, textbooks, Internet etc. and through Interview with experts.
2.4. Methods and Data analysis

The objective of the new system is to perform the function more effectively. What does Expert system do in essence It scores some function against entries in database and makes decisions using a predefined algorithm based on calculated scores. It can be much more intelligent than individual human expert because of database capacity. And probably AI will be able to replace human experts in most situations. The major difference with humans is that Sane Stable AI cannot and should not improvise and make responsible moral decisions then there is not enough information. The proposed system is to be designed using JAVA programming language. In this research work, different procedures are followed in developing the proposed knowledge based system. These are: knowledge acquisition, knowledge modeling, knowledge Representation, Knowledge based system development for Sorghum disease diagnosis and Evaluation of the system. Four domain experts are selected using purposive sampling techniques and interviewed to extract the tacit knowledge. Similarly, documented sources of knowledge are consulted on the area of crop protection and treatment from different sources such as agricultural books, journals, publications, internet sources, plant disease protection guidelines and training manuals are analyzed. The acquired knowledge is modeled by using decision tree. Decision tree shows the relationships of the problem graphically and can handle complex situations in a compact form. Knowledge diagramming is often more natural to experts than formal representation methods and decision trees can easily be converted to rules. Decision tree is drawn using flow chart symbols as it is easier for many to read and understand. It helps to identify a strategy most likely to reach a goal and allow the addition of new scenarios. After modeling the acquired knowledge by using decision tree, it is represented in a format that is both understandable by humans and executable on computers. Production rules are the most popular form of knowledge representation which is an easy way to understand and reasonably efficient in diagnosing problems. Knowledge is represented in the form of condition-action pairs: IF this condition (or premise or antecedent) occurs, THEN some action (or result or conclusion or consequence) will (or should) occur. MATLAB programming tool was used to develop a rule based knowledge based system for sorghum disease diagnosis. To achieve the established objective of the study, the prototype system is extensively tested and evaluated to ensure that whether the performance of the system is accurate and the system is usable by research centers and development agents. Model Development Sorghum disease diagnosis KBS is designed based on rule based reasoning techniques. As shown in Figure 1, the model shows that, the knowledge is acquired from experts and documented knowledge sources. Potential sources of knowledge include domain experts, books, journal articles, proceedings, electronic sources and information available on the web. Then the acquired knowledge is effectively coded in the knowledge base by knowledge engineer. Knowledge base contains rules base from which the system draws conclusion through inference engine. The inference engine accepts query from the user via user interface and prompt the action in user understandable form if the goal is satisfied.

Fuzzy reasoning mechanism is used to search and extract the rules for specific type of sorghum disease. Proposed System User Interface to interact with the system the user interface is needed. User interface is a bidirectional communication between the system and the user. It is the window through which the system is able to return information to the user. Once the file 'SDDKBS.pl' is opened, the end-users can start Sorghum disease diagnosis as the system requests end-users to choose the symptom of the sorghum disease. If a choice match based on the user’s response, the system provides conclusions for the users request through the user interface. System Testing and Evaluation Once a proposed knowledge based system is developed, it should be tested and evaluated to measure its performance in diagnosing Sorghum diseases and determined whether the system satisfies the requirements of its users and applicable in the domain area. Testing and evaluation of the prototype system is the final step that can assist the knowledge engineer to measure, if the objective of the proposed system is met or not. To check if the system fits its purpose, system testing and user acceptance testing are used.

3. System Design

An expert system is a computer program that helps in solving problems demanding substantial human expertness by using explicitly exhibited domain knowledge and computational decision procedures. These are designed to make available some of the skills of an expert to non-experts, as they attempt to imitate the thinking patterns and logical decisions of an expert. The FIS makes use of the theory of fuzzy reasoning. Fuzzy inference is the process of developing the mapping from a given input to an output using fuzzy logic which then offers a base from which decisions can be made or patterns perceived. The classical logic has only two truth values, true or false, and so the process of inference is simplified as compared to fuzzy logic, where we have to be concerned not only with propositions but also with their truth values. Every FIS has a fuzzy inference system that reasons using fuzzy logic membership functions, which refers to the degree to which the value of a particular attribute belongs to a set. The FIS designed and employed in this research can be generalized by means of a simple structure as shown in Fig.3. Below:
3.1. KNN Model

KNN Model accepts inputs from the closest neighbor and verifies from the data set and after verifying from the dataset the availability and accuracy of the input then it predicts the supposed outcome.

3.2. Front end fuzzy logic designer of the model

The front end shows the various inputs used in the design and the output as how the model is expected to work.
3.2.1. Data collection Preprocessing

Condition for Hardness, Hypertrophied glumes and Honey dew (HHH) having linguistic variables and parameters as follows: Less Infected [0 0.2 0.4] More Infected [0.2 0.4 0.6], Much Infected [0.4 0.6 0.8].

![Figure 4](image1.png)

**Figure 4** The first condition of the model

Condition for Yellowish, Brown and Copper (YBC) having linguistic variables and parameters as follows: Infected [0 0.3 0.4], Less infected [0.3 0.6 0.7], highly Infected [0.6 0.7 1].

![Figure 5](image2.png)

**Figure 5** The second condition of the model

Condition for green, purple and black (GBP) having linguistic variables and parameters as follows: Low infected [0 0.2 0.4], Medium Infected [0.2 0.4 0.6] High Infected [0.5 0.6 0.9].

![Figure 6](image3.png)

**Figure 6** The third condition of the model
3.2.2. The output variable
The output variable diseases types (DT) parameters are defined based on the linguistic variables which are the different types of sorghum diseases that contained the following diseases as membership functions: Wire worms, Smuts, Sorghum charcoal rot, Sorghum downy mildew, Loose smut, Sorghum anthracnose, Rust and Rust.

![Figure 7 The output variables of the model](image)

3.2.3. System Validation
The system can be validated based on its structure which shows the input, the inputs membership functions the rules that it adopted and the output membership functions and lastly displayed the output, that will prove that the system has now been validated as all the needed steps was adopted to achieved the desired result.

![Figure 8 Structure of the model](image)

3.2.4. System surface
System surface viewer shows how the surface of the structure is looking that contain the inputted symptoms and shows the diseases type of that symptom. Based on the surface you can view how the system contained all the necessary inputs and output needed.
4. Results and discussion

A Confusion matrix is an N x N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with those predicted by the model. This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making. For a binary classification problem, we would have a 2 x 2 matrix as shown below with 4 values:

![Confusion Matrix Diagram]

**Figure 10** How the actual and predicted values are represented

The target variable has two values: Positive or Negative. The columns represent the actual values of the target variable, the rows represent the predicted values of the target variable. True Positive (TP) The predicted value matches the actual value, the actual value was positive and the model predicted a positive value. True Negative (TN) The predicted value matches the actual value, the actual value was negative and the model predicted a negative value. False Positive (FP) – Type 1 error The predicted value was falsely predicted. The actual value was negative but the model predicted a positive value, Also known as the Type 1 error. False Negative (FN) – Type 2 error The predicted value was falsely predicted, The actual value was positive but the model predicted a negative value, Also known as the Type 2 error. We had a classification dataset with 1000 data points. We fit a classifier on it and get the below confusion matrix:
The different values of the Confusion matrix would be as follows: True Positive (TP) = 560; meaning 560 positive class data points were correctly classified by the model. True Negative (TN) = 330; meaning 330 negative class data points were correctly classified by the model. False Positive (FP) = 60; meaning 60 negative class data points were incorrectly classified as belonging to the positive class by the model. False Negative (FN) = 50; meaning 50 positive class data points were incorrectly classified as belonging to the negative class by the model. Now, let’s predict the level of infection on the sorghum crop and isolate them before they destroy the crop completely. The two values for our target variable would be: Symptoms and Disease type. There are 947 data points for the negative class and 3 data points for the positive class. This is how we’ll calculate the accuracy: Accuracy = TP + TN/TP + FP + TN + FN See how our model performed:

### Table 1 How the model performed

<table>
<thead>
<tr>
<th>ID</th>
<th>Actual symptom</th>
<th>Predict. Disease</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>TP</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>TN</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>TN</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>TP</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>TN</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>TN</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>FP</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>FN</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>TN</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>FP</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>FN</td>
</tr>
</tbody>
</table>

The total outcome values are: TP = 30, TN = 930, FP = 30, FN = 10

So, the accuracy for our model turns out to be:

\[
\text{Accuracy} = \frac{30 + 930}{30 + 30 + 930 + 10} = 0.96
\]

\[
\text{Accuracy} = 96\%
\]

Precision tells us how many of the correctly predicted cases actually turned out to be positive. Precision = TP/ TP + FPReward tells us how many of the actual positive cases we were able to predict correctly with our model. And here’s show we can calculate Recall: Recall = TP/TP + FN
We can easily calculate Precision and Recall for our model by plugging in the values into the above questions:

\[
\text{Precision} = \frac{30}{30 + 30} = 0.5 \\
\text{Recall} = \frac{30}{30 + 10} = 0.75
\]

50% percent of the correctly predicted cases turned out to be positive rate. Whereas 75% of the positives were successfully predicted by our model.

The result shows the various sorghum diseases mentioned above imputed into the model. And as any of the symptoms is clicked that will indicate by a ticked sign which is the data creation. it also shows the result of how those symptoms and disease and the supposed treatment are being imputed or added into the model as will be contained in the knowledge base of the system by the experts. Here the machine learning model is being built to accept the inputted knowledge and stores it in its knowledge base data base. Shows the result of how the model accepts those inputted knowledge and add it successfully to its data base showing that is now a built machine. Shows the result of how the inputted symptoms were displaced on the interface of the model indicating that they are contained in the data base as the knowledge of the symptoms as they are now being tested. Shows the result of the symptoms that is ticked immediately showing the desired output which is the disease type and its prescription, now the model is trained and it diagnosed and prescribe the sorghum diseases. From this discussion it can now be said that the following objectives have now been successfully achieved. The Data creation and preprocessing (categorizing the diseases and its symptoms), The Building of the machine learning model using the supervised classification technique K- nearest neighbor (KNN), Train and test the model using the Database, Diagnose and prescribe the solution to the diseases.

5. Conclusion

This paper focuses on the use of computer system with reference to diagnosis and prescription of sorghum crop disease. The work covers the manual system of operations as regards the problems identified, stating the aims of the new system, stating the various specifications and then implementing the programs. The work was successfully developed using Java programing language, K-nearest neighbour algorithm validated by confusion matrix. The model was tested and improved upon which yields a dynamic result. The research work cannot be said to be perfect, but however, its benefits cannot be overemphasized. It has led to the improvement in the speed of processing operation, efficiency, accuracy and improved storage of data.

Realizing a research of this nature is very exciting. However, the researcher encounter a lot of problem which I believe if looked into, will go a long way toward reducing the tension associated with the design implementation and construction of the model. In spite of the constraints encountered during the implementation of this research work, the aim of the research is well accomplished.
Compliance with ethical standards

Acknowledgments
The authors acknowledged the effort of the Dean School of Agricultural Science and the staff of Computer Science Department Modibbo Adama University as well as the world Journal of Advance and Reviews for this opportunity.

Disclosure of conflict of interest
There is no conflict of interest as regards this paper.

References