Developing and Evaluating a Neuro-Fuzzy Expert System for Improved Food and Nutrition in Nigeria

Oluwatoyin Mary Yerokun[1], Moses Okechukwu Onyesolu[2]

[1]Department of Computer Education, Federal College of Education (Technical), Asaba, Nigeria
[2]Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria
Email: yerokunoluwatoyin@gmail.com, mo.onyesolu@unizik.edu.ng

Abstract

Protein is a class of food needed daily by humans and even animals. Protein is rated as 20% to 30% of daily food requirements making it a very significant part of daily needs in compliance with the International Labour Organization (ILO) order on food. This class of food from animals has been threatened and carries a lot of health risks unlike protein from plant sources. The need for an alternative to plant protein led to this work. A neuro-fuzzy expert system for detection of leghemoglobin in legumes was developed and evaluated. Knowledge acquisition was done by oral interview of prominent biochemists and botanists that provided key technical facts on leghemoglobin and visits to botanical gardens of Society for Underutilized Legumes (SUL) in Nigeria. Production rule-base technique and forward-chaining mechanisms with linguistic antecedent conditions were used. MATLAB platform was employed for the development of the system. Confusion matrix was employed for the performance evaluation of the developed system. The result is a neuro-fuzzy expert system with gaussian membership functions with accuracy of 100% as against 99.56% for triangular, trapezoidal and gaussian combination functions, precision of 100% for all the membership functions evaluated and recall of 100% for gaussian membership functions and 99.53% for triangular, trapezoidal and Gaussian combination functions.

Subject Areas

Network Modeling and Simulation,
Simulation/Analytical Evaluation of Communication Systems

Keywords

Neuro-Fuzzy, Leghemoglobin, Expert System, Confusion Matrix, Dataset

1. Introduction

In 2015, the United Nations rolled out a list of seventeen (17) sustainable development goals (SDGs) to cover a fifteen-year period of 2016 to 2030. Ranking second in these SDGs is "end hunger" (end hunger, achieve food security and improved nutrition and promote sustainable agriculture). Hunger according to World Health Organization (WHO) is a craving or urgent need for food or a specific nutrient [1]. Social scientists declare hunger situation when a person is unable to meet basic nutritional needs for a sustained period [2]. International Labour Organization (ILO) declared in a worldwide publication in 1976 that food—balanced diet and improved nutrition is the first and topmost priority of man [3]. A diet is balanced when it gives the body the nutrients it needs to function correctly and when it consists of a variety of different types of food and provides adequate amounts of nutrients necessary for good health [3].

Food is any substance consisting essentially of protein, carbohydrate, fat, and other nutrients used in the body of an organism to sustain growth and vital processes and to furnish energy [4]. Food has three physiological functions in the body, to provide energy, body building, regulation of body processes and protection against diseases [5]. Body building is the most important physiological function of food because the human body is made up of thousands of cells and new cells are required to be added constantly to these to help the body grow, or to repair and replace damaged and/or dead cells. In order to achieve this, food is needed for the formation of the needed new cells. Regulatory functions refer to the controlling of body processes such as maintenance of body temperature at 98.6°F or 37°C and the heart beats at 72 beats/minute. Excretion of waste products from the body is also categorized as regular body function controlled by food, the body may suffer a condition referred to as constipation if body waste is not well regulated, which is capable of causing further complications. The fight against germs/diseases and maintenance of immunity defense system is the crown of all processes regulated by balanced diet and nutrition. Food is grouped into six (6) classes: proteins, vitamin, carbohydrates, fats and oils, mineral salts and water. Protein is rated as 20% to 30% of daily food requirements [6] because it contains 22 essential amino acids, out of which nine (9) cannot be manufactured by the body and so must be supplied directly from human diet [7], thus making protein a very significant part of daily needs in compliance with the ILO order on food.

It is therefore important that protein in all its forms should be available, affordable and reachable by all. The provision of affordable protein foods for zero hunger and balanced nutrition is therefore the imperative of every government and leaders, meaning that the food sector is the most attractive for business creators and administrators worldwide, scientists and technologists inclusive. Computer scientists and investors have pressed into food and nutrition sector with the same zest of applying computer into every human endeavor to make life easier for the populace by development of intelligent systems for professional

decision supports.

Nutrition experts such as [3] and [6] agree protein can be obtained from animals or plants. Beef, milk, cheese, yoghurt, pork, tenderloin, chicken, eggs, fish, etc. are products of animals while plants such as nuts and legumes contain purer and higher volume of essential nutrients and amino acids, which are supplied directly to the body when consumed [8]. Animal protein has the edge of tasting sweeter, looking richer, smelling appetizingly and even cooking better; this is simply because of its major component, heme or hemoglobin (an iron containing compound of the pophyrin class). Meat on the other hand is risky because it is very high in cholesterol, a fat/oil component which has been medically confirmed as dangerous for the heart. But leguminous plants also contain heme/ hemoglobin and are healthier, easier to prepare and cheaper than meat with no risk to the body.

Leguminous crops are plants in the family Fabacea (formerly known as Leguminosae). They are grown primarily for their grain seed, livestock forage and their soil enhancing properties [8]. This is as a result of the nitrogen-fixing nodules in their roots causing a rhizobium infection in the roots. Some legumes are eaten while some are used as fallow crops (tephrosia and mucuna). Some are shrubs (tephrosia) while some are small plants not more than ten centimeters in height (peanut). The two major groups of legumes are those that produce mainly fat and oil such as groundnuts and soya bean and those that mainly produce protein, cowpeas and beans [9]. Leguminous plants contain leghemoglobin in high volume. Leghemoglobin is simply the hemoglobin in legumes and has almost same chemical and structural similarities to hemoglobin in meat from animals. It is found in the nitrogen-fixing root nodules of leguminous plants and is produced by legumes in response to the roots being colonized by nitrogen-fixing bacteria known as rhizobia. Leghemoglobin is a part of the symbiotic interaction between plant and bacterium and can meet the high demand for meat. Simply put, it is meat from plants, having the same nutritional value, content, taste, feel and look [10]. However, not all legumes have this substance in them. Common leguminous crops such as beans (asparagus, green, kidney, navy, soy, pinto, garbanzo, adzuki, Anasazi, wax, mung, dwarf), southern peas, English peas, snow peas, sugar snap peas, alfalfa, clover, lespedeza, lentils, licorice and peanuts contain large volumes of protein and are adequate to supply the daily requirements [11].

In recent years, Nigeria has witnessed a high increase in cases of communal clashes between herdsmen and farmers for pastures for grazing lands, this has risen to a national crises. The cost of raising poultry birds and fishes has also become so high as a result of inflation and economic challenges, such that an average family can barely afford meat to supply their daily dietary requirements. Eating meat is considered a health risk because of its cholesterol content. All these pose great threats to the stable health of the Nigerian populace. It is imperative therefore for scientists, agriculturists, nutritionists and all involved in food

production to find other alternatives within the boundaries of safety, affordability and availability, to meet the protein requirements of Nigerians. Since leguminous plants stand as a healthy source of protein, there is need to take advantage of this natural providence in solving food problems. However, not all legumes contain leghemoglobin that can supply the body with the same unique taste and all essential amino acids [9]. There is therefore a need to detect which legumes have rhizobium infection or colonization and therefore contain leghemoglobin. Intelligent management of information and knowledge becomes vital for individual awareness and reduction in national food and health issues. The main focus of this work is to develop and evaluate a neuro-fuzzy expert system for improved food and nutrition in Nigeria by ascertaining alternative to animal protein which is safer for human consumption.

2. Scientific Innovations in Food and Nutrition Technology

Many automated systems have been built to assist and improve food nutrition, especially in fruitful attempts to end hunger. Many organizations have shown interest in using intelligent systems to reduce the complexity and iterative processes of educating, training and sharing expert knowledge on food production [12]. Various techniques have been employed by food industries to implement production of foods based on extraction and incorporation of nutrients (supplements) to ensure required daily diet rations into food items for supply [13]. One of such innovation is the discovery of leghemoglobin in leguminous plants and the search for intelligent application of technology to detect its presence, extract and produce it in better consumable forms for the populace. Reference [14] deployed AI to develop an expert system that can detect diseases in coffee plants. By applying forward chaining technique, decision trees of hierarchical classification were used to achieve the fuzzy logic based system that can help researchers or observers working in coffee plantation to diagnose coffee plants diseases. A system for meat drying processing was developed by Zhang [15]. Neuro-fuzzy approach was used to control relative humidity and temperature for the meat drying room, this method decoupled the influence of temperature and relative humidity in real-time drying room control process. It reduced energy consumption with results based on real time measurements. Subsequently, the cost of food production and preservation was reduced with increase in quality and efficiency of production. Reference [16] used data mining concepts in Support Vector Machine (SVM) algorithm and segmentation algorithm from captured images of plant leaves. Java JSP was used to extract basic features of a leaf, identify the pattern of points, find edges of the leaf and then formulate a forecasting model of plant disease.

Onions made the list of important foods when Sitanggang and others [17] deployed AI to develop an application that can diagnose diseases at the earliest stage. IF-THEH-ELSE rule-based reasoning method was used to build the fuzzy knowledge base and inference system. The user is prompted to enter or select

various symptoms observed on the plant till a match is found in the knowledge base for certain characteristics of onion diseases, then a conclusion or decision is inferred and displayed for the user.

Reference [18] used java package to develop an application system for nutritional guidance and a user-friendly Nutriflow app, which used a food composition database from the government of Finland. It linked with several research and academic institutions databases, food industries and Finland National Health institute to comprise the knowledge base, which converts the information in the food diaries to average daily nutrient intake levels. Its success was in its ability to interpret nutrients level results from the first fuzzifier level to guide and advice a balanced diet without altering the original diet intake of user too much.

Reference [19] designed a fuzzy system for the assessment of food security risk level using type 2 fuzzy system. A type-2 fuzzy model was constructed for the determination of the food security risk level using input parameters cereal yield, cereal production, and economic growth for Turkey. Reference [20] used back propagation technique in artificial neural network (ANN) to develop a system based on image processing techniques to monitor growth of red spinach plants. A plant growth monitoring program was created with Borland Delphi 7 software and DS Pack version 2.3.4 component and ran the analysis/prediction on MATLAB. The system was aimed at reducing the complexity of image extraction from plants and simultaneously increasing the accuracy of results obtained from growth monitoring exercises.

3. Materials and Methods

A neuro-fuzzy expert system that detects the presence of leghemoglobin in plants was developed and evaluated using the architecture presented in Figure 1.

3.1. Data Collection and Processing

3.1.1. Knowledge Acquisition

Knowledge acquisition in general simply refers to obtaining raw facts and figures from experts, which can be carried out manually or automatically. In the manual mode, a knowledge engineer process knowledge from one or more domain experts or published papers, whereas in the case of automatic mode, various machine learning algorithms are applied for knowledge elicitation from a vast base of data (e.g. the internet) and refinement of such knowledge. In this work, the knowledge acquisition component was done by orally interviewing prominent biochemists and botanists who provided key technical facts about the subject domain leghemoglobin, followed by visits to botanical gardens of Society for Underutilized Legumes (SUL), Ibadan, Oyo state; Olusegun Agagu University of Science and Technology (OAUSTECH), Okitipupa, Ondo State, the Federal University of Technology, Akure (FUTA), various tacit (unconscious knowledge), procedural (knowledge on how to do something) and declarative (knowledge of something being true or false) were captured via interview. Other relia-

ble sources are certified published journals from various servers on the internet and books. The most important aspect of the subject domain (leghemoglobin) explained by the various experts and implemented in the knowledge base of the expert system developed includes: 1) Identification of legumes from other plants; 2) Identification of rhizobia growth in nodules; 3) Identification of Symbiotic leghemoglobin in root nodules of legumes.

The morphological properties of each identification properties were obtained from the various experts, each structure was broken down to the bare minimum and structured into an executable form (production rules) and serve up to inference engines. In cases where the expert knowledge was not readily understood, further questions were asked to verify that the knowledge elicited was correct, and to get rid of semantic errors. In most cases where there seems to be a similarity between morphological structures for the identification of legumes from other plants, experts were asked how to further differentiate between such plant. Interviews were conducted repeatedly, until both the knowledge elicitor and the domain experts (botanist and biochemist) were on the same page for each identification process and extraction. Knowledge elicited were also reviewed by the experts to ascertain the correctness of collated knowledge and make correction where needed. Data collected were all summed in Table 1.
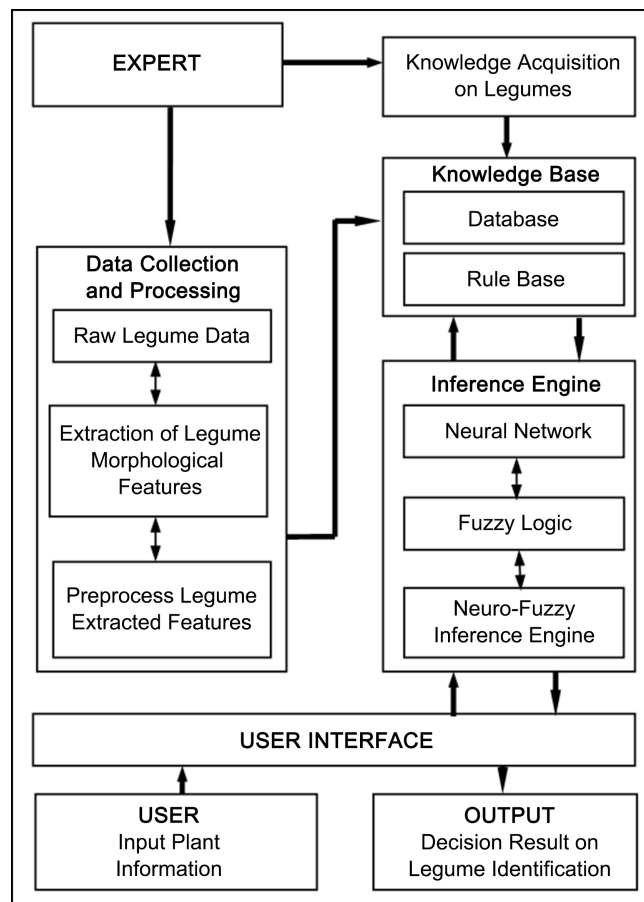

Figure 1. Neuro-fuzzy system architecture.

3.1.2. Data Extraction, Reduction and Feature Extraction

The next stage is to extract features that can be used as predictors. It is necessary to find an approximate method or technique for reducing the dimension and extracting features that can be used to characterize data which will be suitable as input for classification and learning algorithm. Botanical data sets are usually stored in notes and procedures outlined in laboratory manuals.

The lists of attributes (Table 1) were further reduced to manageable groups and formats. Data reduction is very important to the effective management of data. In this work, the plants attributes elicited from the experts are twenty three (Table 1), which was later compressed into Table 2 with various values (number of attributes). Table 1 highlighted (in bold) the lists of attributes selected for the reduction. In Table 1, fruit and seed were collapsed to pods, roots to nodules (Table 2). These are features that were used as predictors that serve as input into the ANFIS.

To further reduce clumsiness and ambiguity in the identification process, metrics/weights were obtained and attached to each morphological feature for precision as shown in Table 3.

Table 1. Essential attributes elicited from the domain experts for identification process.

S/No	Parts of Plant	Attributes	Number of Attributes
1	Inflorescence or Flower	**colour**, symmetry, length, orientation.	4
2	Leaf	type, shape, arrangement, blade edges, **stipule or petiole**	5
3	**Fruit**	type, length, surface, colour, number (cluster, single)	6
4	**Seed**	number, shell (hardness/softness)	2
5	Root	number, hair (primary/secondary), colour (primary/secondary)	3
6	Cross section of root	**colour**, content (juicy or hard)	3

Table 2. Data reduction.

S/No	Attributes	List of Attributes	Value
1	Colour of Flower	white, yellow, red, blue, purple	5
2	Stipules	present, not present	2
3	Pods	size, colour, shape, dehisces	4
4	Nodules	hard, soft, cluster, single	4
5	Cross Section	red, pink, white, cream	4

Table 3. Input variables, ranges, and fuzzy sets.

Input Variables	Ranges	Fuzzy Set
	<128	Dull
Color of Flower	110 - 180	Bright
	170 - 255	Very Bright
	<0.3 cm	Very Small
Stipules	0.25 - 0.6 cm	Small
	0.5 - 1 cm	Large
	<0.03 mm	Very Small
Pods	0.025 - 0.06 mm	Small
	0.05 - 0.1 mm	Large
	<0.3 cm	Very Small
Nodules	0.25 - 0.6 cm	Small
	0.5 - 1 cm	Large
	<128	Dull
Color of Cross Section of Root (CCSR)	110 - 180	Bright
	170 - 255	Very Bright

Colour of flower and colour of cross section of root used the RGB (Red, Green and Blue) colour system. The RGB refers to the three hues of visible light, for representing colours on any computer display (not printing). It can be combined to get any colour in the visible spectrum. It is additive in nature because it generates all colours when mixed in varying percentages, when up to 100% of each of the three are nixed can generate white, while 0% of each can generate black. CMYK (Cyan, Magenta, Yellow and Black) is another alternative but not appropriate in this work because of its subtractive nature, the colours get darker when mixed to higher percentages. RGB measures from zero (0) to two hundred and twenty five (225) as follows: 1) 0 to 128 = Dull; 2) 110 to 180 = Bright; 3) 170 to 255 = Very Bright.

Stipules, Pods and Root Nodules: the simple measuring apparatus, meter rule was used to determine the thickness in centimeters (cm) of each of them. Calibrating the variables' measurements into three levels as follows: 1) 0 to 0.3 cm = Very Small; 2) 0.25 cm to 0.6 cm = Small; 3) 0.5 cm to 1 cm = Large.

Triangular membership function (Equation (1)), Trapezoidal membership function (Equation (2)), Gaussian membership function (Equation (3)), and combination of two Gaussian membership functions (Equation (4)) were employed in the fuzzification of the ANFIS model of Figure 1.

$$\mu_{(X)} = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \leq x \leq b \\ \dfrac{b-x}{c-b} & \text{if } b \leq x \leq c \\ 0 & \text{if } x > c \end{cases} \tag{1}$$

$$\mu_{(X)} = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } b \leq x \leq c \\ \dfrac{d-a}{b-a} & \text{if } c \leq x \leq d \\ 0 & \text{if } x > d \end{cases} \qquad (2)$$

where a is the start point, b represents the peak point, and c, d end point of the linguistic term of the linguistic variable x of input parameters (color of flower, stipule, pods, nodules, and the color of roots) respectively.

$$f_2 = \mu_{A_i}(x) = e^{-\frac{(x-c_i)^2}{2\sigma_i}} \qquad (3)$$

c_i, σ_i are the parameters of the membership function governing the Gaussian function. c_i, is the parameter for locating center and σ_i is the parameter for determining the width of the curve.

$$y = \text{gauss2mf}\left(x, \left[sig_1 c_1 sig_2 c_2\right]\right) \qquad (4)$$

The first function specified by $sig_1 c_1$ determines the shape of the left-most curve, the second function specified by $sig_2 c_2$ determines the shape of the right-most curve. When $c_1 < c_2$, gauss2mf function reaches a maximum value of 1. Otherwise, the maximum value is less than 1.

3.1.3. Evaluation of Data Reduction Accuracy

In the development of expert systems, evaluation is important and must be carried out after data reduction to assess and validate the data obtained before further processing [1]. From the list of attributes provided in Table 1, accuracy of the feature extraction and data reduction can be ascertained from the ratio of total attributes used in each identification process, to the attributes used by human experts. Reference [21] listed twenty five attributes for identifying any plant, while twenty three is identified and used in this work.

The ratio is calculated as follows:

Attributes given by experts = 25;

Attributes used = 23;

Ratio = (23/25)*100 = 92%.

Therefore, accuracy for the data used is 92%.

3.2. Knowledge Engineering

On completion of data reduction process, the proceeding step is designing the knowledge base; this involves picking the best method of representing the knowledge and the rules governing the inference making. Python (Anaconda/Miniconda version) programming platform was used for the creation, packaging and wrapping of the system. Aesthetics and the flow of various segments were linked with Anaconda/miniconda. The control center presented the two

sections of the work: Identification Decision Support Panel and Information Support System. MATLAB platform was employed for the development of the analysis and detection section of the system. MATLAB is an Integrated Development Environment which allows both procedural and object-oriented approach for the representation of knowledge in artificial intelligence mode. Class and objects were used to model the information elicited. A class named Rule was created and houses various objects such as: Rule(), AddAntecedent(), setConsequent(), AddRule(), AddFact(), and etc which serves as subsidiary methods. Each rule is saved into the system using the specific object to the class Rule, some rules added contains media contents which are saved on the database, pointers are used to point to contents in the database as supported by MATLAB and access control to the database. In order to model the identification of legume, rhizobia, leghaemoglobin processing a rule instance is created for each identification and a rule name is created and antencedents are added to the rule created using the AddAntecedent() object to ascertain the conditions to be met to lead to the setconsequent (conclusion) added using setConsequent() object for each identification. To query the inference engine the AddFact() object is used to provide facts entered into the system by users to the inference engine for antecedent cross-matching.

3.3. The System Database

A relational database was created with Microsoft Excel and linked with the MATLAB files to serve as the local database for mapping of media antecedents contained in the knowledge base and inference engine. These contents provide direct interaction platform with the user in more practical manner and easy use of the system, thereby increasing the accuracy of inferred reasoning presented to a user by the inference engine.

3.4. The Inference Engine

This is the brain of the expert system which does the reasoning and conclusion derivation based on production rules in the knowledge base. It is also the subsystem that is directed through knowledge base rules to manipulate the knowledge in the working memory so as to draw an inference from it. The type of inference engine developed in this work is a rule-based inference engine. MATLAB provided the platform as an interactive IDE. The inference engine works by using backward chaining technique, it compares facts supplied by users to the system with the production rules already available in knowledge base. This means that if facts supplied by user matches one or sets of antecedents of rule then the consequent of such rule is fired by the inference engine which activates various actions by the system. Antecedents are expressions involving attributes connector "AND" in the inference engine. The system only executes a consequent when a rule containing it is fired and a rule is executed only when all of its antecedents are satisfied by facts supplied to the user to the inference engine.

3.5. The Working Memory

This is where all data that are received from user during interaction is kept; these values are used to evaluate antecedents in the knowledge base. Consequents from rules may create new values in learning or training set, update old values or adjust some existing values, as necessary.

3.6. The User Interface

The user interface consist of graphical clickable controls which allow ease of input of queries by users and a natural language response controls to enable ease of use and understanding by expert and non-expert users. At the physical level, the user interacts with the system with the following controls: 1) Navigation controls (for ease of navigation between forms); 2) Button controls (for executing of queries); 3) Explanation controls (for explaining consequents of the inference engine). At the system level the user interface also consists of two basic components: 1) The interviewer component: this interface is used for asking various questions from the user to gather facts for the inference engine to compare against the rules in the knowledge base of the system; 2) The explanation component: Conclusion/consequents are displayed in this component to users and antecedents leading up to the consequent are shown to the user from this component, reassuring users of the reasoning process. For examples it might output the rules used to come to the consequent.

3.7. Neuro-Fuzzy Inference Component

The decision support system for identification of legume plants is based on the integration of the concepts of fuzzy logic and artificial neural networks. The fuzzy logic component introduces the notion of a membership function, artificial neural networks, with their different architectures built on the concept of artificial neurons; provide opportunities for learning, adaptation and generalization. However this ANFIS inference system supports fuzzy system of **Sugeno** type due to its ability to work well with adaptive techniques like the neural network. The membership functions, triangular (trimf), gaussian (gaussmf), trapezoidal (trapmf), and gaussian combination (gauss2mf) functions were used.

3.8. Performance Evaluation

Performance of the system was evaluated using confusion matrix from where the accuracy, precision and recall of the system were calculated using Equation (5) to Equation (7). The confusion matrix is a square matrix table that is used in supervised learning to describe the performance of any classification models on test dataset by representing the actual (column) and predicted (row) dimensions. It makes it easy for programmers to clearly see the performance of the model designed. It gives an idea of what classification models predict correctly and what they predict incorrectly. Hence, the number of correct and incorrect predictions are summarized with count values and broken down by each class (leghemoglo-

bin and non-leghemoglobin). In this work, the test dataset (without the class labels) were fed to the trained ensemble model and prediction for each row in the test dataset was performed. Thereafter a comparison between the expected outcomes and the predicted outcomes were performed by counting the number of correct and incorrect prediction for each class. These numbers are then organized into a table where each row corresponds to the predicted class and each column corresponds to an actual class. The counts of correct and incorrect classification are then filled into the table as shown in Table 4, where all correct classifications lie along the principal diagonal (TP and TN) and the incorrect classifications correspond to numbers off the diagonal (FP and FN).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100 \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{7}$$

4. Results and Discussion

4.1. The Neuro-Fuzzy Expert System

The system comprises of two sections: the decision support and information support sections. Screenshots shown in the following figures clearly illustrate the windows and options on both sections of the system. Figure 2 shows the decision support with the three options for users to choose or navigate. Clicking any of the icons will lead to opening the menu and further interactions between user and system. The options are identification of legumes, identification of rhizobia and identification of leghemoglobin in the root nodules and the user can interact with the system in two modes on the three sections: the interactive terminal or selection panel.

Selection or clicking on identification of legumes opens to the window in Figure 3 with the dialogue box. The user supplies the data as prompted to eventually lead to results in sample in Figure 4.

Table 4. Confusion matrix for a binary classification.

		Predicted Class	
		Non-leghemoglobin	leghemoglobin
Actual Class	Non-Leghemoglobin	TN	FP
	Leghemoglobin	FN	TP

where: 1) True Positive (TP)—This is the proportion of actual positive cases which are correctly detected as leghemoglobin present; 2) False Positive (FP)—Known as Type I Error, represents represent the proportion of data samples detected as leghemoglobin present (positive) but not present (negative); 3) True Negative (TN)—The proportion of actual negative cases correctly detected as leghemoglobin not present. 4) False Negative (FN)—This is also known as Type 2 Error, represents the proportion of data samples detected as leghemoglobin not present (negative) but false, *i.e.* leghemoglobin is present (positive).

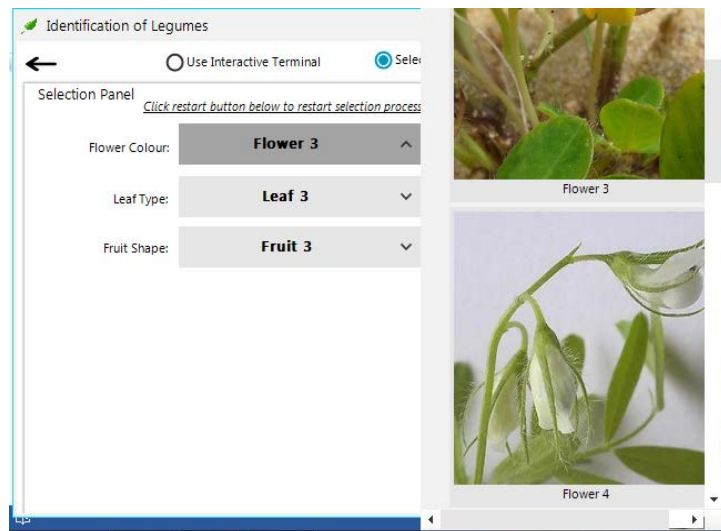Figure 2. Decision support section of the system.
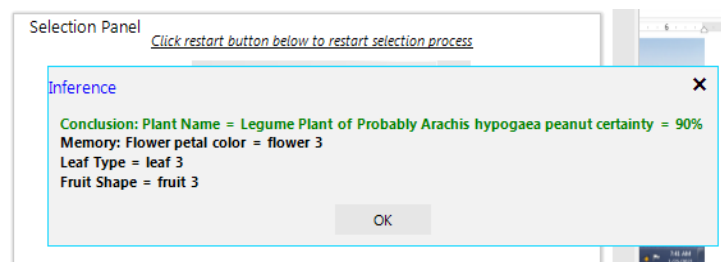

Figure 3. Identification of legume window.

Figure 4. Result of identification of legume on selection panel.

The second option is the identification of rhizobium presence shown in **Figure 5** and the sample result in **Figure 6**.

Confirmation of symbiotic leghemoglobin is the third window/option. When users click this option, the following window in **Figure 7** is displayed and the consequent result sample is shown in **Figure 8**.
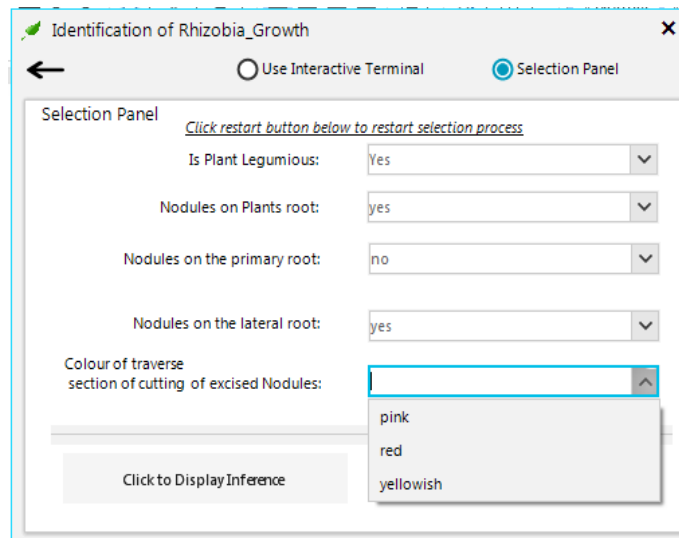
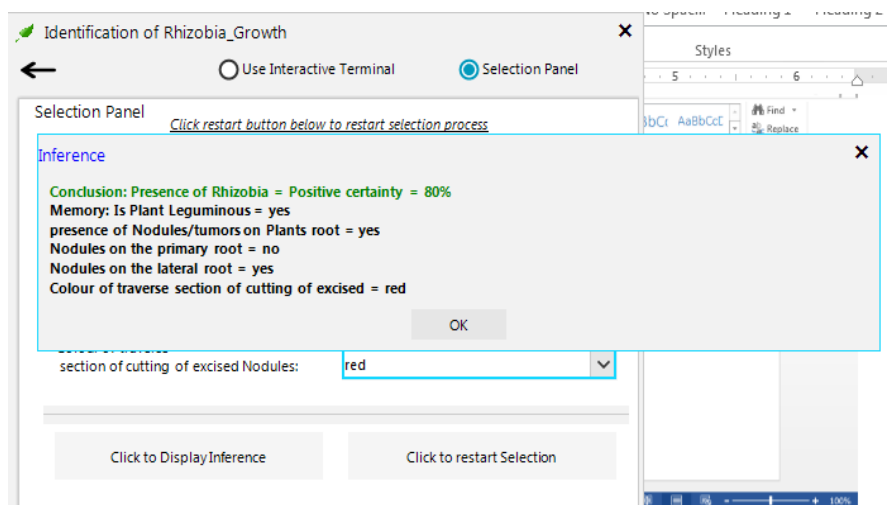**Figure 5.** Identification of rhizobia growth.

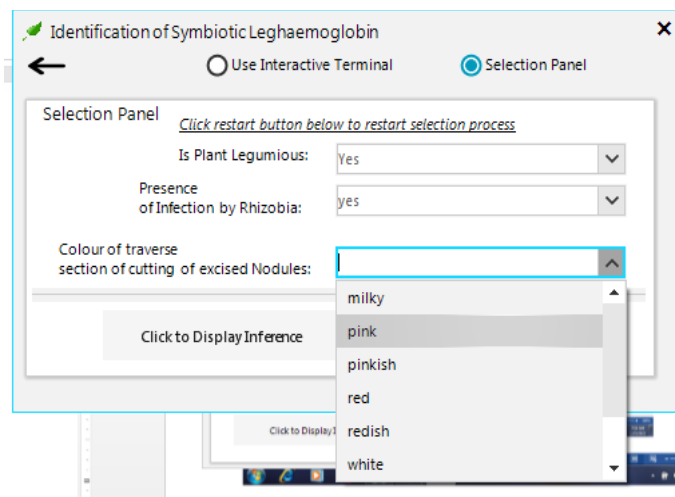**Figure 6.** Result of identification rhizobia growth.

Figure 7. Identification of leghemoglobin.

The second section of this work is the information system containing educative information on leghemoglobin providing menu to select from (**Figure 9**). This section supplies brief explanation of leghemoglobin, its extraction procedures, more illustrations on symbiotic leghemoglobin and some video clips on this nutrient (**Figure 10**).
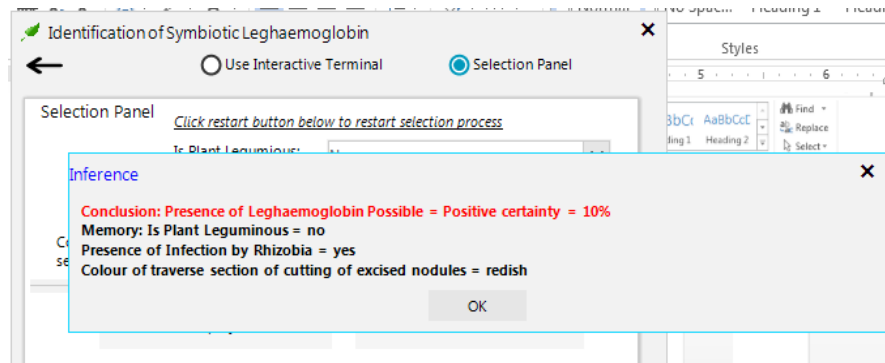

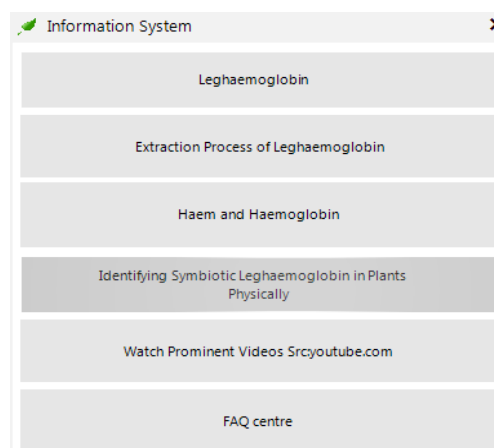**Figure 8.** Sample result of identification of leghmoglobin.

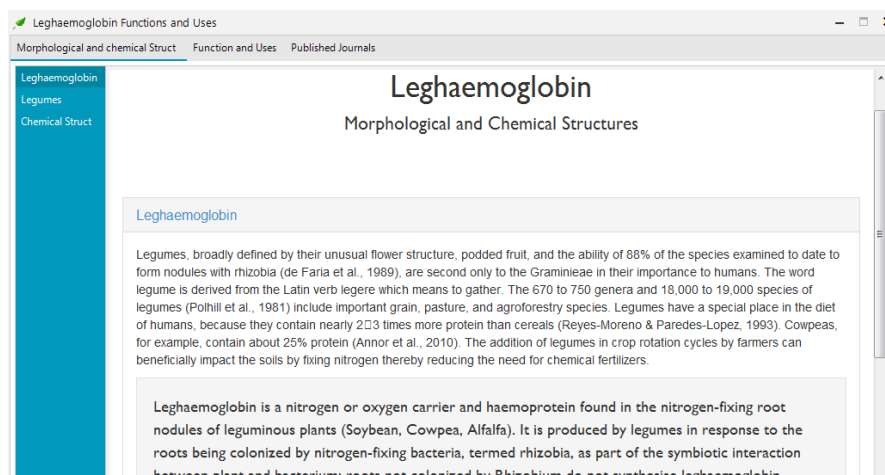**Figure 9.** Menu of information system on leghemoglobin.

Figure 10. Information on leghemoglobin displayed when selected from the menu.

The total number of dataset for this work is Seven Hundred and Fifty Three (753) instances of plants. The dataset is partitioned into training and test set using 70:30 splitting ratio as the training set is comprised of 527 plant instances while the test set is comprised of 226 plant instances as shown in Table 5. The training set is used to train the developed model, while the test set is used to validate the performance of the trained model.

The final result obtained from the system is one of two options: 1) Leghemoglobin Present; 2) Leghemoglobin not Present. After confirmation of the systems accuracy and precision by the evaluations, the detection function is activated via the DETECT button. The five morphological features are open for input of values through the drop-down buttons. The expected input values are as follows: 1) Color of Flower: Dull, Bright, Very Bright; 2) Stipules: Very Small, Small, Large; 3) Pods: Very Small, Small, Large; 4) Nodules: Very Small, Small, Large; 5) Colour of Cross Section of Root: Dull, Bright, Very Bright. Prompt each of the variables by clicking on its button and select values as necessary, then click on DETECT button. Instantly the result is displayed as Leghemoglobin Present or Leghemoglobin not Present.

4.2. System Evaluation

Performance evaluation provides the result and evaluation of the developed ANFIS-based expert system for leghemoglobin detection with four basic membership functions such as triangular (trimf), gaussian (gaussmf), trapezoidal (trapmf), and gaussian combination (gauss2mf) functions. The developed model was evaluated to determine the performance of its detection as applied to the different membership earlier mentioned. The performance was measured based on computational time, accuracy, precision, and recall rate obtained from confusion matrix plots. The confusion matrix plot shown in Table 6 displays the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), when the model is validated with the test data.

Accuracy describes the general performance of the developed model on datasets. It determines the closeness of actual values to the predicted values. However, this work examined the accuracy of the developed model by sampling the enlisted memberships in Table 7 using Equation (5). It was observed that the developed ANFIS-based expert system performed better using Gaussian membership function with an accuracy of 100% accuracy, while the developed model

Table 5. Plant distribution.

Class Labels	Plant Dataset			
	Training	Training %	Testing	Testing %
Heghemoglobin (1)	430	81.59	213	94.24
Non-Heghemoglobin (0)	97	18.41	13	5.76
Total	527	100	226	100

Table 6. Confusion matrix for test data.

Membership Functions	TP	FP	TN	FN
Triangular (trimf)	213	0	12	1
Gaussian (gaussmf)	213	0	13	0
Trapezoidal (trapmf)	213	0	12	1
Gaussian combination (gauss2mf)	213	0	12	1

Table 7. Accuracy of model on different membership functions.

Membership function	Accuracy	Accuracy (%)
Triangular (trimf)	0.9956	99.56
Gaussian (gaussmf)	1.0000	100.00
Trapezoidal (trapmf)	0.9956	99.56
Gaussian combination (gauss2mf)	0.9956	99.56

performance was the same for the other three membership functions with 99.56%. Table 6 shows the accuracy of developed model on each membership function.

There is need to determine the model's performance in detecting leghemoglobin in roots of leguminous plants from the predictions made. This can be determined using precision as it determines the percentage of leguminous plants that is predicted as leghemoglobin plant from the model's prediction. Applying Equation (6), Table 8 shows how different membership functions performed using precision metrics. The study revealed that the model recorded a perfect 100% precision when it was applied on the memberships used in this research.

Recall measures the ratio of correctly predicted leguminous plants in the actual class to all the plants in the dataset. That is, it computes the number of leguminous plants that was predicted correctly as found in the actual class. This work shows the recall rate of the developed ANFIS-based expert system when applied on the triangular, trapezoidal, guassian, and gaussian combination membership functions (Table 9). From this study, it is observed that the model's recall rate is highest using Gaussian membership function with 100%, while the three other membership function recorded an equal 99.53% using Equation (7).

Table 10 shows the time taken per seconds for complete execution of task by the developed model based on the four membership functions. The table shows that the developed model was fastest when applied on trapezoidal membership function with least computational time of 11.91 seconds.

Figure 11 show the bar plots accuracy, precision, and recall rate in a single chart for clarity and understandability. The horizontal axis of the plot holds the performance metrics, while the vertical axis represents the percentage of their performance.

**Table 8.** Precision of model on different membership functions.

Membership Function	Precision	Precision (%)
Triangular (trimf)	1.000	100.00
Gaussian (gaussmf)	1.000	100.00
Trapezoidal (trapmf)	1.000	100.00
Gaussian combination (gauss2mf)	1.000	100.00

Table 9. Recall rate of model on different membership functions.

Membership function	Recall	Recall (%)
Triangular (trimf)	0.9953	99.53
Gaussian (gaussmf)	1.0000	100.00
Trapezoidal (trapmf)	0.9953	99.53
Gaussian combination (gauss2mf)	0.9953	99.53

Table 10. Computational time of model on different membership functions.

Membership function	Time
Triangular (trimf)	12.200
Gaussian (gaussmf)	12.094
Trapezoidal (trapmf)	11.910
Gaussian combination (gauss2mf)	12.063
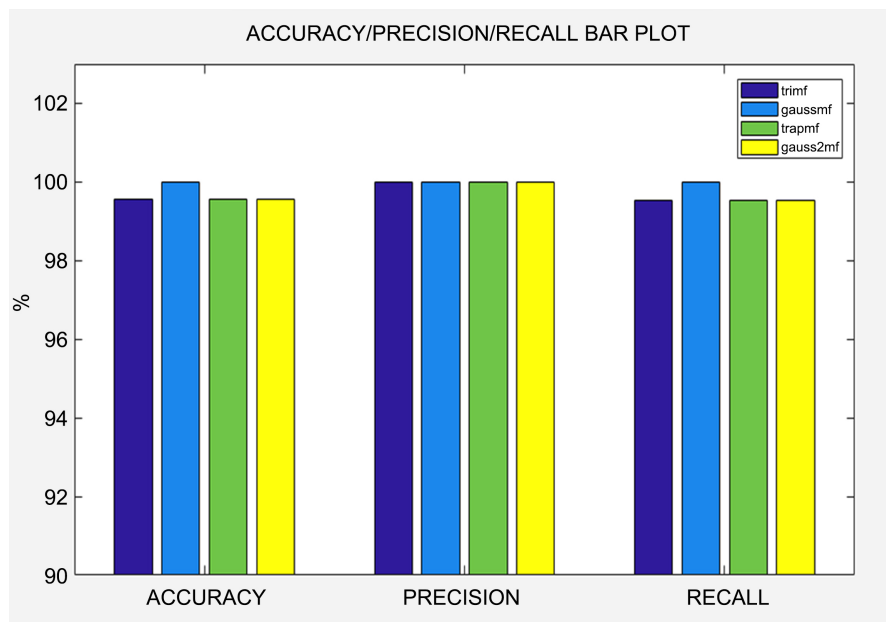
Figure 11. Multiple bar plot showing summary of accuracy, precision, and recall.

5. Conclusion

This work developed and evaluated a neuro-fuzzy expert system for the detection of leghemoglobin in plants. The accuracy of the system is up to the level of an expert in the subject domain of leghemoglobin and it is interactive and easy to use for non-experts. This system is illustrative and educative, supplying relevant information about leghemoglobin and its safety as food. It is therefore recommended for all and sundry especially food and nutrition experts seeking to accomplish the millennium development goals of zero hunger. The work will be of immense benefit to the entire Nigerian populace. Food availability and affordability are both lifelines of human survival; the research will provide a new direction on how to use computer technology with plant morphological data to meet nutritional requirements with the knowledge of healthy choice of plants as replacement for meat. The information collected, and the development of the system will expose the user of the system to the knowledge of plants containing leghemoglobin and how to identify these plants among others. This will be a springboard for further studies on how computer technology can further advance botanical research studies; it will use intelligent means to solve the problem of food supply and location and is expected to produce information for ministries of food and industry. If the food industries are boosted, basic needs of the people are met; food and agriculture will become easier and more attractive to investors, cleaner job opportunities and in the long run to boost the economy of the country.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Akomolafe, D.T., Yerokun, O.M. and Olufowobi, S.B. (2019) Design and Development of a Rule Based Expert System for the Identification of Common Terrestrial and Aquatic Plants. *The First Faculty of Science Annual Conference* (*FOS*2019) *of the Ondo State University of Science and Technology*, Okitipupa, 16-18 July 2019, 54-62.

[2] Garcia, J.S., Perez, I.B. and Diez-Sanz, J.M. (2019) Hunger and Sustainability. *Journal of Economic Research*, **32**, 850-875.
https://doi.org/10.1080/1331677X.2019.1583588

[3] Krans, B. (2016) Balanced Diet.
http://www.healthline.com/health/balanced-diet#Overview1

[4] Ratini, M. (2018) Good Protein Sources. Healthy Protein Food Society.
http://www.webood.com

[5] Arai, S. (1994) Physiological Functions of Foods. In: Yano, T., Matsuno, R. and Nakamura, K., Eds., *Developments in Food Engineering*, Springer, Boston, 48-53.
https://doi.org/10.1007/978-1-4615-2674-2_9

[6] Collins, C. (2017) What Is a Balanced Diet Anyway? The Conversation: Eating

Science. https://theconversation.com/what-is-a-balanced-diet-anyway-72432

[7]   Ware, M. (2017) How to Get Enough Protein.
      https://www.medicalnewstoday.com/articles/310118.php#what_does_protein_do

[8]   Towler, B.F. (2014) How Is Energy Measured: The Future of Energy. Elsevier Inc.,
      Academy Press, Cambridge, 17-26.
      https://doi.org/10.1016/B978-0-12-801027-3.00002-6

[9]   Yingling, G.L. (2017) Gras Notification for Soy Leghemoglobin Protein Preparation
      Derived from Pichiapastoris. Impossible Foods Inc., Redwood City, GRAS Notice
      (GRN) No. 737.
      https://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/NoticeInventory/d
      efault.htm

[10] Fraser, R., Shitut, M., Agrawal, P., Mendes, O. and Klapholz, S. (2018) Safety Evalu-
      ation of Soy Leghemoglobin Protein Preparation Derived from Pichiapastoris, In-
      tended for Use as a Flavor Catalyst in Plant-Based Meat. *International Journal of
      Toxicology*, **37**, 241-262. https://www.ncbi.nlm.nih.gov
      https://doi.org/10.1101/196766

[11] Food and Agriculture Organisation of the United Nations (2016) Legumes Can
      Help Fight Climate Change, Hunger and Obesity in Latin America and the Carib-
      bean. http://www.fao.org/americas/noticias/ver/en/c/409536

[12] Olanloye, D.O. (2017) Development of an Artificial Intelligence Geoinformatics
      System for Solid Minerals Prospecting. A Dissertation Submitted to the Department
      of Computer Science, Faculty of Physical Sciences, Nnamdi Azikiwe University,
      Awka.

[13] Yerokun, O.M. and Onyesolu, M.O. (2021) On the Development of Neuro-Fuzzy
      Expert System for Detection of Leghemoglobin (NFESDL) in Legumes. *The Journal
      of Digital Innovations & Contemporary Research in Science*, *Engineering & Tech-
      nology*, **9**, 129-140. https://doi.org/10.22624/AIMS/DIGITAL/V9N1P1

[14] Surhatono, D., Aditya, W., Lestari, M. and Yasin, M. (2013) Expert System in De-
      tecting Coffee Plant Diseases. *International Journal of Electrical Energy*, **1**, 156-162.
      https://doi.org/10.12720/ijoee.1.3.156-162

[15] Zhang, W. (2015) Adaptive Decoupling Control and Energy Efficient System for
      Meat Drying Processing Based on Neuro-Fuzzy Approaches. A Thesis Presented to
      the Department of Electrical Engineering, Faculty of Engineering, University of
      Guelph, Ontario. https://doi.org/10.1109/ICCIS.2015.7274551

[16] Shamkuwar, D.O., Thakre, G., More, A.R., Gajakosh, K.S. and Yewale, M.O. (2018)
      An Expert System for Plant Disease Diagnosis by Using Neural Network. *Interna-
      tional Research Journal of Engineering and Technology*, **5**, 369-372.

[17] Sitanggang, D., Siregar, S.D., Situmang, S.M., Indra, E., Sagala, A.R., Sihombig, O.,
      Nababan, M., Pasaribu, H., Damanik, R.R., Turnip, M. and Saragh, R.I. (2018) Ap-
      plication of Forward Chaining Method to Diagnosis of Onion Plant Diseases. *Jour-
      nal of Physics*: *Conference Series*, **1007**, Article ID: 012048.
      https://doi.org/10.1088/1742-6596/1007/1/012048

[18] Heinonen, P., Mannelin, M., Iskala, H., Sorsa, A. and Juuso, E. (2009) Development
      of a Fuzzy Expert System for a Nutritional Guidance Application. *International
      Fuzzy Systems Association World Congress and European Society of Fuzzy Logic
      and Technology Conference*, Lisbon, 20-24 July 2009, 1685-1690.
      https://dblp.org/ifsa/eusflat/conference2009

[19] Abiyev, R., Uyar, K., Ilhan, U. and Imanov, E. (2016) Assessment of Food Security
      Risk Level Using Type-2 Fuzzy System. Elsevier, Amsterdam, 547-554.

https://doi.org/10.1016/j.procs.2016.09.441

[20] Wijaya, R., Hariono, B., Saputra, T.W. and Rukmi, D.L. (2019) Development of Plant Monitoring System Based on Multi Camera Image Processing Techniques on Hydroponic Systems. *IOP Conference Series: Earth and Environmental Science*, **411**, Article ID: 012002. https://doi.org/10.1088/1755-1315/411/1/012002

[21] Jagtap, S.B. and Hamarde, S.M. (2014) Agricultural Plant Leaf Disease Detection and Diagnosis Using Image Processing Based on Morphological Feature Extraction. *IOSR Journal of VLSI and Signal Processing*, **4**, 24-30. https://doi.org/10.9790/4200-04512430