

# **Prediction of Wheat Rust Diseases Using Data Mining Application**

# Wesenu Bekele Mulatu<sup>1</sup>, Mebratu Fana Bedasa<sup>2</sup>, Gemechu Kena Terefa<sup>3</sup>

<sup>1</sup>Department of Information System, College of Computing, Madda Walabu University, Bale Robe, Ethiopia <sup>2</sup>Department of Computer Science, College of Computing, Madda Walabu University, Bale Robe, Ethiopia <sup>3</sup>Department of Information Science, College of Computing, Madda Walabu University, Bale Robe, Ethiopia Email: wesenubekele1@gmail.com, mebratu.fana@yahoo.com, gemekef@gmail.com

How to cite this paper: Mulatu, W.B., Bedasa, M.F. and Terefa, G.K. (2020) Prediction of Wheat Rust Diseases Using Data Mining Application. *Open Access Library Journal*, **7**: e6717. https://doi.org/10.4236/oalib.1106717

Received: August 14, 2020 Accepted: September 15, 2020 Published: September 18, 2020

Copyright © 2020 by author(s) and Open Access Library Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/

CC Open Access

# Abstract

Stem rust caused by Puccinia graminis f.sp. tritici and stripe rust caused by Puccinia striiformis are the most destructive wheat rust diseases when environment conditions are favorable in regions where wheat crops are grown. An early prediction mechanism can play a great role in forecasting the occurrence of the rust disease. It assists proactive control and early decision making. However, in the absence of prediction mechanism of wheat rust disease hurried wheat production yield loss more. Hence, to overcome these issues, this study was conducted to develop wheat stripe rust and stem rust diseases prediction model using data mining application. The meteorological and disease data for the year 2010 to 2019 of national meteorological agency southern Oromia region Bale Robe service center and Oromia Seed enterprise Sinana farm unit II were obtained and used for the study. Two bread wheat varieties, namely Kakaba and Danda'a were involved in the study. Fist, daily meteorological data mean values of three-day consecutive interval were computed. Then, mean value of meteorological data has been integrated with disease incidence and severity date observed during the most critical infection period (mid-august to November-30) was used to develop the model. WEKA software machine learning tool with J48 decision tree algorithm was used for data preprocess and experiments. It is open source software containing various machine learning algorithms for data mining tasks. The study results showed that, stripe rust predictive model trained and tested with using training set test option achieved accuracy 75.70% and stem rust predictive model achieved accuracy 90.045% with cross-validation fold 10 test option. The results found were much promised to forecast occurrence of wheat stripe and stem rust disease.

# **Subject Areas**

Big Data Search and Mining, Network Modeling and Simulation

## **Keywords**

Prediction, Wheat Rust Disease, Data Mining, Decision Tree, Stem Rust, Stripe Rust

# **1. Introduction**

Wheat is grown on more than 200 million hectares worldwide and is a source of food and livelihoods for hundreds of millions in developing countries [1]. Ethiopia is one of the developing country and the largest wheat producer in sub-Saharan Africa, next to South Africa. Wheat production is the second most important source in total production next to maize and the third in area after maize and sorghum that plays a significant role in assuring food sufficiency [2]. It is largely grown in the highlands of the country and constitutes 24.6% of the annual cereal crop production and plays an important role in supplying the population with various nutrition advantage [3]. It is consumed in various forms such as bread, cakes, biscuits, injera (bread wheat), pasty, and macaroni (durum wheat) [4].

Wheat production is severely affected by abiotic and biotic stresses [5]. Specially, drought, floods and diseases severely affect wheat production. Exacerbated by climatic stress, particularly in rain fed areas, the impact of wheat rust diseases is expected to increase. Wheat production in Northern and Eastern Africa, the Near East and West, Central and South Asia is vulnerable to rust diseases. These regions account for around 37% of worldwide wheat production. The cost of a 10% loss in areas at risk is estimated to exceed USD 5.8 billion. The impact on food as well as on nutrition security is estimable. These regions are vulnerable to epidemics of wheat rust diseases caused by stripe rust (*Puccinia graminis f.sp. striiformis*), stem rust (*Puccinia graminis f.sp. tritici*) and leaf rust (*Puccinia triticina*) also known as yellow, black and brown rust respectively. In late 2013, an epidemic of stem rust affecting over 18000 hectares of wheat crop in Ethiopia was caused by a new strain of stem rust [6].

The potential yield loss caused by wheat rust diseases, depends on host susceptibility, weather conditions, timing and severity of disease outbreaks related to crop growth stage. The greatest yield losses take place when one or more of these diseases occur before the heading stage of development [7]. Stripe rust caused by *Puccinia striiformis* is a very important disease of wheat, particularly in central and west Asia and North Africa. It is reckoned to have caused recurrent, severe damage in crops since the dawn of agriculture. It is mainly a disease of wheat grown in cooler climate  $(2^{\circ}C - 15^{\circ}C)$ . Stem rust is caused by *Puccinia graminis*. It is also referred to as summer rust due to the abundant production of shiny black spores, which form at the end of the crop growing season. It is favored by humid conditions and warm temperatures of  $15^{\circ}C$  to  $35^{\circ}C$ . An apparently healthy crop, three or four weeks before harvest, can be reduced to a black tangle of broken stems and shriveled grain. Harvest losses of 100 percent can

occur in susceptible crop varieties. Another wheat rust disease is leaf rust, caused by *Puccinia triticina* and occurs to some extent wherever wheat is grown. The disease develops rapidly at temperatures between 10°C and 30°C. Leaf rust losses in grain yield are primarily attributed to reduce flower set and to grain shriveling. In highly susceptible wheat varieties, the crop can be killed by early epidemics. Crop losses due to leaf rust are usually small (less than 10 percent), but have been known to cause up to 30% crop losses [8].

Oromia Seed Enterprise Bale branch Sinana Farm, this wheat rust diseases were the causes of wheat production yield loss. In this farm, to identify the rust diseases and to take action, the domain expert visually observed color of wheat rust diseases, symptoms and assess disease incidence and severity. This mechanism is very difficult to recur wheat crop from damage even the chemical is sprayed.

In agriculture, to increase production quantity as well as quality it is needed to early control factors that reduce production. Nowadays, Information Technology (IT) constitutes an important part of our life today [9]. Currently, Data mining Technology is growing rapidly in all domain specifically in agriculture, healthcare, market analysis, financial, banking and telecom, which represents knowledge implicitly stored in huge data sets. Current technologies provide us lot of information on agriculture related activities and also used to retrieve useful hidden information from them [10]. Data mining is a technology that uses various techniques to discover hidden knowledge from heterogeneous and distributed historical data stored in large databases, data warehouses and other massive information repositories so to find patterns in data. According to Namita et al. [11], data mining is the process of extracting useful and important information from large sets of historical data. It is computer assisted method of analyzing huge sets of data to extract useful information which assist to make a decision. Data processing tools forecast behaviors and future trends, permitting business to form proactive, knowledge driven selection. Data mining techniques used in the field of agriculture is useful in predictive of problems, disease detection and optimizing the pesticide. The main objective of this study was to discover predictive model for wheat stripe rust and stem rust disease which forecast existence/absence of the disease.

# 2. Statement of the Problem

Rust diseases are economically important diseases in the world, because of its wide distribution, capacity to form new races that can attack previously resistant cultivars, ability to move long distance, and potential to develop rapidly under optimal environmental condition [7] [12]. Ethiopia is one of African country affected by these epidemic wheat rust disease highly, which causes reduction of wheat harvest and increase the cost of food and pose a real treat to rural livelihoods and regional food security. The main stripe rust epidemics in Ethiopia occurred in 1970's, 1988, and 2010. Disease occurs regularly in highland areas over 2000 m above sea level. In 2010 at least 400,000 ha affected. This was a serious

problem difficult to measure. The epidemic covered almost all wheat growing regions in the country and most of the commercial bread wheat cultivars were susceptible. Emergency fungicide applications are used in 2010 30% of the wheat area was sprayed and over \$3 million worth of fungicides were distributed. This is a big problem for the country which leads to deficiency in wheat production yield [8].

In Oromia Seed Enterprise Sinana Farm, in 2010, there was severe outbreak of stripe rust devastating a popular mega variety named Degalu. The disease spread over large areas where wheat was grown. It damages a lots of wheat production. In addition, in 2013, 2014 and 2015 there was severe outbreak of stem rust and stripe rust devastating on mega varieties. This causes reduction of wheat production and leads insufficiency of food security. To identify and control wheat rust diseases epidemics, the domain expert uses symptoms of each rust diseases, color and calculate disease distribution and severity after the disease occur. This mechanism is very difficult, because after the disease occurs, the probability to recur is very less and wheat production can reduce.

Scientists and agricultural specialists around the world are working to more effectively monitor, track and combat the spread of wheat rust diseases. Early detection and well-organized reporting are the keys to better managing and reducing wheat rust in regions at risk. Diseases predictive model/forecasting system should be needed to provide advice and support for the action [8] [13].That is the main reason many of the research still being conducted in rust disease prediction to early warning and detection for better decision making.

# 3. Overview of Data Mining

Data mining is the process of extracting useful and important information from large sets of data. Data mining in agriculture field is a relatively novel research field [11] [14]. Data mining, through better management and data analysis, can assist agricultural organizations to achieve greater profit. Therefore, it is essential that managers of agricultural organizations get to learn about the idea and techniques, because the amount of available information is sure to grow in the future, and it will not become clearer and easier to understand and make decisions. Understanding of the processes which are carried out and decisions being made in agricultural organizations is enabled through data mining [11]. There are various data mining model. Some of them are described as following section.

# 3.1. Cross Industry Standard Process for Data Mining (CRISP-DM)

Cross-Industry Standard Process for Data Mining (CRISP-DM) majority of sharing in data mining. This Proposes of CRISP-DM methodology for mining: a multi-step iterative process. As shown in **Figure 1**, it consists of six phases such as business understanding special the way of the business perspective is critical because it identifies the business objectives and, thus, the success criteria of data mining projects.



Figure 1. Phases of the CRISP-DM [15].

1) Business understanding: This initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

2) Data understanding: This phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.

**3) Data preparation:** The data preparation phase covers all the activities required to construct the final dataset from the initial raw data. Data preparation tasks are likely to be performed repeatedly and not in any prescribed order.

**4) Modeling:** In this phase, numerous modeling techniques are selected and applied and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type.

**5) Evaluation:** Before proceeding to final model deployment, it is important to evaluate the model more thoroughly and review the steps taken to build it to be certain that it properly achieves the business objectives.

**6) Deployment:** Model construction is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it [15].

# 3.2. Knowledge Discovery in a Database (KDD)

The data-mining field currently relies heavily on known techniques from machine learning, pattern recognition, and statistics to find patterns from data in the data-mining step of the Knowledge Discovery in Databases (KDD) process [16]. As shown in **Figure 2**, the KDD process of data preparation, data selection, data cleaning, incorporation of prior knowledge, and proper interpretation of the results of mining ensure that useful knowledge is derived from the data. An important notion of "interestingness" is usually taken as an overall measure of pattern value, combining validity, novelty, usefulness, simplicity and understandability. As a matter of fact, knowledge in this definition is purely user oriented and domain specific and it is determined by whatever function and threshold the user chooses. The roles of interestingness are to threshold the huge number of discovered patterns and report only those which may be of some use [17].

In view of machine learning process or KDD refers to the overall process of discovering useful knowledge from data. The distinction between the KDD process and the data-mining step (within the process) is a central point of this article. The additional steps in the KDD process, such as data selection, preparation and cleaning and incorporation of appropriate prior knowledge, and proper interpretation of the results of mining, is essential to ensure that useful knowledge is derived from the data [16].

Knowledge discovery concerns the entire knowledge extraction process including how data are stored and accessed, how to use efficient and scalable algorithms to analyze massive datasets, how to interpret and visualize the results, and how to model and support the interaction between human and machine. It also concerns support for learning and analyzing the application domain area. Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. The data mining component of KDD currently relies heavily on known techniques from machine learning, pattern recognition, and statistics to find patterns from data in the data-mining step of the KDD process [18].



Figure 2. Steps of KDD process source (MIT press, Menlo Park, CA, 1996).

Lastly, when as writers explains encounter patterns within a database the researchers state the findings (patterns or rules) as data mining, information retrieval or knowledge extraction and so on. The term data mining is used mostly by statisticians, data analysts and the management information systems (MIS) professionals. The difference between Data Mining (DM) and Knowledge Discovery (KD) is that the latter is the application of different intelligent algorithms to extract patterns from the data whereas KD is the overall process that is involved in discovering knowledge from data. There are other steps such as data preprocessing, data selection, data cleaning, and data visualization, which are also a part of the KDD process. Many people treat DM as a synonym for another popularly used term, KD from Data, or KDD. Alternatively, others view DM as simply an essential step in the process of KD. Hence, in the definition as the writer adopt, DM is just a step in the overall KDD process [17].

## 3.3. Hybrid Models

The development of academic and industrial models has led to the produce of hybrid models that combine aspects of both. One such model is a six-step Knowledge Discover Process (KDP) model developed by Cios *et al.* It was developed based on the CRISP-DM model by adopting it to academic research. The main differences and extensions include providing more general, research-oriented description of the steps and introducing a data mining step instead of the modeling step, more description of the six steps see **Figure 3** of the model follows [19].

1) Understanding problem of the domain: This initial step involves working closely with domain experts to define the problem and determine the project goals, identifying key people, and learning about current solutions to the problem. A description of the problem, including its restrictions, is prepared. Finally, project goals are translated into data mining goals, and the initial selection of data mining tools to be used later in the process is performed.

2) Understanding of the data: This step includes collecting sample data and deciding which data, including format and size, will be needed. Background knowledge can be used to guide these efforts. Data are checked for completeness, redundancy, missing values, plausibility of attribute values, etc. Finally, the step includes verification of the usefulness of the data with respect to the data mining goals.

**3) Preparation of the data:** This step concerns deciding which data will be used as input for data mining methods in the subsequent step. It involves sampling, running correlation and significance tests, and data cleaning, which includes checking the completeness of data records, removing or correcting for noise and missing values, etc. The cleaned data may be more processed by feature selection and extraction algorithms (to reduce dimensionality), by derivation of new attributes (say, by discretization), and by summarization of data (data generalization). The end results are data that meet the specific input requirements for the data mining tools selected in phase 1.



Figure 3. Hybrid process model.

**4) Data mining:** Here the data miner uses various data mining method such as classification, clustering and association rule discovery to derive knowledge from preprocessed data as per the objective of the study.

5) Evaluation of the discovered knowledge: Evaluation includes understanding the results, checking whether the discovered knowledge is novel and interesting, interpretation of the results by domain experts, and checking the impact of the discovered knowledge. Only approved models are retained, and the entire process is revisited to identify which alternative actions could have been taken to improve the results. A list of errors made in the process is prepared.

6) Use of the discovered knowledge: This final step consists of planning where and how to use the discovered knowledge. The application area in the current domain may be extended to other domains. A plan to monitor the implementation of the discovered knowledge is created and the entire project documented. Finally, the discovered knowledge is deployed.

## 3.4. Related Works

Tamene et al. [13], has conducted study in 2013 to assess analysis of climate va-

riability effects on wheat stem rust epidemics in Bale and Arsi zones of Oromia regional state, Ethiopia. The meteorological and disease data for the year 2004 to 2013 of Sinana and Kulumsa were obtained from the respective agricultural research centers and were analyzed. Four bread wheat cultivars, namely Kubsa, Mada walabu, Sofumer and Tusie were included in the study. The researcher used analysis data SPSS (IBM SPSS) version 20 for windows, 2011. The researchers found that total seasonal rainfall had positive effect on the development of wheat stem rust, while the seasonal mean minimum temperature and seasonal average relative humidity significantly affected the development of wheat stem rust in the field and the researchers has develop a regression model.

The researcher [20], has conducted the study on linkage of climate variability, rust diseases and wheat product in central parts of Ethiopia. The data used for the study was monthly rainfall total, maximum and minimum temperatures, monthly Relative Humidity (RH) and wind speed recorded at six stations for the periods 1971 to 2008. Based on meteorological station data and global indices, long-term mean and standard deviation were computed by using INSTAT and SYSTAT software. The results revealed that the severities of stripe rust have severely increased during El Nino years than in non-El Nino years. The stripe rust development over the study site initially depended on Belg season's rainfall while its infection became apparent in Kiremt (summer) season's maximum temperatures, whereas the development and spreading of stem rusts were highly influenced by Kiremt season rain. The researchers propose that skillful and reliable weather-related early warning can be well practiced by acquiring appropriate lead-time climate-based forecasting on the possible occurrence of both climates and diseases on varieties of wheat crops across the study area. The researchers [21], study weather-based prediction models for leaf rust using disease severity and weather data recoded at four locations. The researchers were identifying crop growing season critical periods for relating weather variables to disease. Disease data for cultivar C-306 were collected at all four research centers (locations) in the study .Highly significant correlation coefficients were found between disease severity and a greater number of weather variables in these critical 3-week periods than at other times. The correlation coefficients were greatest for the humid thermal ratio, maximum temperature and special humid thermal ratio and these three weather variables were selected as predictor variables. Linear regressions algorithm were used with these predictor variables (individually) during the critical periods, and a multiple regression with maximum temperature and relative humidity, serve as four disease prediction models, with sufficient lead time to take control measures. According to [21], results of validation of these prediction models with disease severity data showed that the regression equation with maximum temperature (Model 1) was the best among the prediction models, with four out of six simulations matching observed disease severity classes and also having lowest residual sum of squares value of 2727. Models 4 (multiple regression), 2 humid thermal ration and 3 special humid thermal ratio with residual sum of squares values of 2881, 3092 and 3732, respectively are in order of decreasing accuracy of prediction. The researcher concluded that, the model using maximum temperature can be used to predict the disease severity.

In general, most of the researches did on prediction of rust diseases on different topics. But till there is no nonlinear prediction model research done using data mining and with decision tree algorithm. Therefore, this research was done a nonlinear model to predict wheat rust disease by using data mining with decision tree algorithm.

# 4. Material and Methodology

# 4.1. Description of Study Area

The study was conducted in Ethiopia, Oromia region, Oromia Seed Enterprise Sinana Farm (OSESF) site II (**Figure 4**). Sinana Farm is one of Oromia Seed Enterprise Bale branch Farm center. Oromia seed enterprise Sinana Farm 42 km is far away from Robe city, zonal city of Bale zone. Sinana is found in Bale zone of Oromia region at 463 km from Addis Ababa capital city of Ethiopia. Its geographical location is 07°07'N latitude and 40°10'E longitude. The elevation of the center is 2400 m.a.s.l with topography of gentle slope plain, which has beauty scene for vision and is quite conducive for agricultural production system under rain-fed in the present climatic conditions. Sinana Farm is one of the government Farm which more wheat crop is produced in large.



## Figure 4. Map of study area.

# 4.2. Research Design/Procedure

There are numerous data mining techniques presented today with their appropriateness to be applied in different agricultural areas; Such as agricultural decision support system, improve yield production and agricultural research and development. The data mining techniques used to predict the occurrence of different agriculture problems like epidemic disease and production quantity. While introducing this new approach for wheat rust disease prediction can based on wheat rust disease data and meteorological data were used with Weka software decision tree algorithm which is easy and simple to implement, evaluate the trends of the disease. The Cross Industry Standard Process for Data Mining (CRISP-DM) technique was used to discover patterns to predict wheat rust disease. This model was chosen since it exhibits all the advantages of well-known and used methodology called CRISP-DM and provides a more general, researchoriented description. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data. There are six stapes in CRISP-DM as showed above under Section 3.1 in Figure 1 [22]. The researchers were use these CRISP-DM techniques and were every activity following this techniques procedure.

# 4.3. Decision Tree Classifier

The use of decision trees is perhaps the easiest to understand and the most widely used method that falls into the category of supervised learning. Decision tree is powerful in its functions and a very popular tool for classification and making Prediction. The graphical representation of a simple decision tree using two attributes (**Figure 5**). A typical decision tree system adopts a top-down strategy in searching for a solution. It consists of nodes where predictor attributes are tested at each node, the algorithm examines all attributes and all values of each attribute although to determining the attribute and a value of the attribute that will "best" separate the data into more homogeneous sub-groups with respect to the target variable or class variable [23].

Decision trees are an approach of representing a sequence of rules that lead to a set or value. As a result, they are used for directed from mining, mainly classification. One of the main important of decision trees is that the model is quite reasonable since it precedes the form of generate rules that bases of the intervention. Another important classifier is easy and simple to implement. It doesn't have domain knowledge or additional parameter setting. It handle huge amount of dimensional data. It is more suitable for exploratory knowledge discovery. The results attained from Decision tree are easier to interpret and used for the prediction of trends and unknown patterns from the database. A decision tree is a classification tree when the outcomes are predicted interims of a class and a real number then it is known as regression analysis leaves in the decision tree represent the class labels and combination of these class labels are represented by the branches [24].



Figure 5. Graphical representation of decision tree machine learning algorithm.

The decision tree algorithm used in this research is J48 algorithm, which implementation produces decision tree models. The algorithm uses the greedy technique to induce decision trees for classification. A decision-tree model is built by analyzing training data and the model is used to classify unseen data. J48 generates decision tree; the nodes of tree evaluate the existence or significance of individual features. Furthermore, a set of classification rules can be extracted from the decision tree by tracing the path from the root to each leaf (corresponding class). This set of rules can be consequently plugged into settle knowledge-based system. So, the researcher using J48 method in order to get the best fitted model that can appropriate to predict the pattern of existence/present or absent of wheat rust disease [24].

A decision tree is a classifier expressed as a recursive partition of the instance space. It consists of nodes that form a rooted tree, meaning it is a directed tree with a node called "root" that has no incoming edges. All other nodes have exactly one incoming edge. A node with outgoing edges is called an internal or test/decision node. All other nodes are called leaves also known as terminal or decision nodes. In a decision tree, each internal node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values. In the simplest and most frequent case, each test considers a single attribute, such that the instance space is partitioned according to the attribute's value. In the case of numeric attributes, the condition refers to a range. Each leaf is assigned to one class representing the most appropriate target value. Alternatively, the leaf may hold a probability vector indicating the probability of the target attribute having a certain value. Instances are classified by navigating them from the root of the tree down to a leaf, according to the outcome of the tests along the path [25].

## 4.4. Induction Rule

Rule induction is the process of extracting useful "if then" rules from data based

on statistical significance. A Rule based system constructs a set of if-then-rules. Knowledge represents IF-THEN rules for classification. An IF-THEN rule is an expression of the form even though the pruned trees are more compact than the originals; they can still be very complex. Hence, generate rules to make a decision tree model more readable, it can be transformed into an IF-THEN decision rule. Decision rules can be generated from a decision tree by traversing any given path from the root node to any leaf. The complete set of decision rules generated by a decision tree is equivalent to the decision tree itself. Rule induction or decision rule classifiers are set of IF-THEN classification. An IF-THEN rule induction is an expression of the form IF condition THEN conclusion. If the condition in a rule antecedent holds true for a given tuple, we say that the rule antecedent is satisfied and that the rule covers the tuples [15].

# 4.5. Performance Evaluation

Once a predictive model was developed using the meteorological and disease data, the model was checked as to how it was performed for the future data that it has not seen during the model building process. The researcher had used J48 classifiers to build the predictive model and in order to evaluate the performance of the model, for evaluation confusion matrix was used. Confusion matrix is a useful tool for analyzing how well a classifier can recognize tuples of classes. A confusion matrix is a table of size two by two. An entry, in the first rows and columns indicates the number of tuples of class one that were labeled by the classifier as class two. For a classifier to have good accuracy, ideally most of the tuples would be represented along the diagonal of the confusion matrix, this explore in the figure [26]. According to [27], a confusion matrixis a simple performance analysis tool typically used in supervised learning. It is used to represent the test result of a prediction model. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. One benefit of a confusion matrix is that it is easy to see if the system is confusing two classes.

As seen in **Figure 6**, a confusion matrix table of size two by two, the following measures can be calculated to measure the accuracy of the model, true positive rate, false positive rate, accuracy, Precision, recall, F-measure and ROC curve.

1) The Accuracy (AC) is the proportion of the total number of predictions

		Actual class		
		Positive	Negative	
ss	Positive	True positive (Tp)	False positive (Fp)	
Predic clas	Negative	False negative (Fn)	Ture negative (Tn)	

Figure 6. Confusion matrix.

that were correct. It is determined using the equation:

Accuracy = 
$$\frac{\text{True Positive}(Tp) + \text{True Negative}(Tn)}{Tp + Fp + Tn}$$

2) The **True positive rate** of a classifier is estimated by dividing the correctly classified positives by the total positive count.

True Positive Rate = 
$$\frac{Tp}{Tp + Fn}$$

3) The **False positive** rate of the classifier is estimated by dividing the incorrectly classified negatives by the total negatives.

False positive = 
$$\frac{Tn}{Tn + Fp}$$

4) **Precision** is calculated by dividing correctly classified instances by the total number of correctly and incorrectly classified samples.

Precision = 
$$\frac{Tp}{Tp + Fp}$$

5) F-Measure is calculated as the harmonic mean of recall and precision.

F-measure = (2 \* Precision \* Recall)/(Precision + Recall)

## ROC (Receiver Operating Characteristics Analysis) curve

ROC is another model performance evaluation method. ROC analysis is performed by drawing curves in two dimensional spaces, with axes defined by the True Positive rate and False Positive rate. The True Positive rate and False Positive rate values of different classifiers on the same test set are often represented diagrammatically by a ROC Graph. On a ROC Graph, the value of False Positive rate is plotted on the horizontal axis, with True Positive rate plotted on the vertical axis. Each point on the graph can be written as a pair of values (X, Y) indicating that the False Positive rate has value X and the True Positive rate has value Y. The performance of different types of classifier with different parameters can be compared by inspecting their ROC curves.

# 5. Datasets

#### **5.1. Sources of Data**

Sources of data used for this research was historical based rust diseases data and meteorological data for ten successive wheat production year (2010 to 2019) in study area Oromia Seed Enterprise Bale branch Sinana Farm Unit/site II, Bale zone, Oromia region, Ethiopia. The historical data of rust disease of stripe rust and stem rust and meteorological data was based on main wheat production season from mid-August to November-30 of each ten year. This season is a critical period of wheat rust disease occurrence in the study area.

Meteorological data for 10 successive years (from year 2010 to 2019) was col-

lected from two meteorological locations center. National Meteorological Agency Southern Oromia region Bale Robe service center and National Agro Meteorological service station which is close to the study area are the source of meteorological data for this research study. Weather variables data like daily Rainfall (mm), daily Minimum Temperature (°C) and daily Maximum Temperature (°C) were obtained from National Agro Meteorological Service station. Relative Humidity weather variable data was collected from National Meteorological Southern Oromia region Bale Robe service center.

**Disease data:** Historical rust disease data of two Commercial Bread Wheat varieties (*Kakaba and Danda'a*) was obtained from the study area during 10 years (from year 2010 to 2019) wheat production in summer season. The two varieties were chosen because, they are currently under production in the study area. The researcher was recording the history of rust disease first arrival of disease data into excel format into computer from the files of rust disease (stripe and stem). The wheat rust disease assessment was undertook started from Tillering to hard dough growth stage of the wheat crop per blocks as Zadoks guide-lines in study area by domain experts.

Data collection methods used in this research was secondary data by analysis historical data of meteorological data and wheat rust disease data. The meteorological and wheat rust disease historical data was used from 2010 to 2019 during wheat production main season. Wheat rust diseases assessed during year 2010 to 2019 wheat production cycle from mid-august to November thirty of each year data was collected from file of disease assessed by domain experts. The previously disease assessed by domain experts was translated in excel format into computer. The primary data collection methods also applied to discover the knowledge of diseases assessment methods, assessment period, season of rust exist. Missing value, outliers, noisy data were regulated with consultation of domain experts and Weka algorithm and taken into consideration for model development.

The sampling method used in this research study was purposive sampling techniques. The reason chosen this purposive sampling technique was, the goal of this research was to intentionally select subjects to collect information. Researchers are working with a specific goal in mind through the lens of quantitative research. The focus remains on individuals with specific characteristics in a targeted population group of interest. Furthermore, the researcher chosen purposive sampling techniques that help to choose wheat variety that long time used under production wheat in study areas. Two bread wheat varieties were chosen purposively from thirteen varieties, which was under production at this time and takes above ten years underproduction

## 5.2. Data Integration/Combination

The very first step of the analysis correlated, historical series meteorological data with disease severity assessment dates were observed. The combination/integration

of meteorological data and disease severity observed dates was matched with each other and combined. In study area, the domain experts were assessed rust disease in intervals of three days differences. The three days interval data mean value of each weather variables were calculated and matched/combined with dates of disease severity observed. So, the consecutive three days interval average data was used for the development of predictive models. In the study, all weather variables considered as independent variables that was used to predict the occurrence of diseases. The disease severity was considered as dependent variable/predicted target class. The presence/existence of pathogen was assigned as (YES) target class and the absence/non-existence of pathogen to (NO) target class.

After data preprocessing was done, the researcher save file into csv file format prepare to data mining WEKA software. Weather Variables daily rainfall (mm), daily minimum temperature (°C), daily maximum temperature (°C) and relative humidity (%) were selected as final predictor variables and disease severity were indicated as response. The classified outputs were useful to distinguish whether the meteorological variables analyzed can predict an epidemic disease (YES) or the absence (NO) of the disease. **Table 1** illustrates the selected final variables/attributes ready for experiments with the use of algorithm.

# 6. Experimentation and Analysis

The experiments were run in WEKA software 3.8.2 with prepared datasets. This addressed the objectives of the research study with weather variables as predicators and disease severity as a response/dependent variable, the target class output whether the disease was existing or not. The experiments of stripe rust and stem rust were performed as in next section.

#### 6.1. Stripe Rust Experiments

For this research, stripe rust experiments were done with J48 un-pruned decision tree algorithm with all 4 attributes/variables and target class the results of the experiment were presented in experiments below with cross-validation10 folds and use training set test machine learning test and train methods.

#### Table 1. Details of attributes.

No.	Attribute	Data Type	Description
1	Rainfall (mm)	Numeric	Daily registered rainfall
2	Minimum Temperature	Numeric	Daily registered minimum temperature
3	Maximum Temperature	Numeric	Daily registered maximum temperature
4	Relative Humidity	Numeric	Daily registered relative humidity
5	Class/target output	Nominal (YES/NO)	Target predict output as disease occur or not

#### 6.1.1. Result Analysis

This experiment of un-pruned J48 decision tree experimental train and test with test option cross-validation folds 10 and use training set test mode were used. The total four factor variables/independent attributes were used in this experiment for built stripe rust predictive model. For this experiment 177 mean values of consecutive three days intervals of weather instances were finalized and used.

In the case of test mode cross-validation folds were 127 instances correctly classified which is equal to 71.7514% and as well 50 instances were incorrectly classified which is equal to 28.2486% as illustrated in **Figure 7**. In the case of use training set test mode machine learning mode 134 instances/cases were correctly classified which is equal to 75.7062%. Incorrectly classified instances 43 which is equal to 24.2938% as shown in **Figure 8**. As the experiment results shows that, the model built by use training set machine learning train and validation mode achieve the accuracy results of 75.7062%. The number of correctly classified more than the result gained by cross-validation folds 10 machine learning training and validation method. The precision, recall and True positive and Accuracy results obtain by un-pruned J48 decision tree algorithm with both cross-validation 10 folds and use training set machine learning training and validation methods were summarized in **Table 2**.

The tested model has accuracy of un-pruned J48 decision tree algorithm 71.75% using 10-fold cross-validation and 75.70% accuracy using use training set test option. The model tested has true positive rate 71.97% using 10-fold cross-validation and 75.00% using use training set test option. The results of precision were 94.95% using 10 cross-validation and 95.79% using use training set machine learning model test method were achieved. As the experiment results shows in **Table 2**, the results achieved by use training set model test/evaluation

Correctly Classified Instances	127	71.7514 %
Incorrectly Classified Instances	50	28.2486 %
Kappa statistic	0.2295	
Mean absolute error	0.3742	
Root mean squared error	0.4609	
Relative absolute error	84.7679 %	
Root relative squared error	98.1725 %	
Total Number of Instances	177	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.95	0.759	0.72	0.95	0.819	0.625	no
	0.241	0.05	0.7	0.241	0.359	0.625	yes
Weighted Avg.	0.718	0.527	0.713	0.718	0.668	0.625	

=== Confusion Matrix ===

```
a b <-- classified as
113 6 | a = no
44 14 | b = yes
```

Figure 7. Stripe rust experiment result with cross-validation 10 folds test mode.

Correctly Classified Instances	134	75.7062 %
Incorrectly Classified Instances	43	24.2938 %
Kappa statistic	0.3545	
Mean absolute error	0.3373	
Root mean squared error	0.4107	
Relative absolute error	76.4383 %	
Root relative squared error	87.4941 %	
Total Number of Instances	177	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.958	0.655	0.75	0.958	0.841	0.749	no
	0.345	0.042	0.8	0.345	0.482	0.749	yes
Weighted Avg.	0.757	0.454	0.766	0.757	0.724	0.749	

=== Confusion Matrix ===

a b <-- classified as 114 5 | a = no 38 20 | b = yes

Figure 8. Stripe rust experiment result with use training set test mode.

Table 2. Summary of two model test options experiment results.

Algorithm -	True Positive Rate		Pre	cision	Accuracy		
	10-fold	Training set	10-fold	Training set	10 folds	Training set	
J48un-pruned	pruned 71.97% 75.00%		94.95% 95.79%		71.75%	75.70%	

option were achieved higher results than experiment results achieved by crossvalidation fold 10 test option. The experiment results of true positive rate, precision and accuracy using use training set model test option were higher than the experiment results achieved using cross-validation 10-fold model test option. This indicates that the best decision tree model produced by J48 unpruned using use training set model test/validation option shows better performance evaluation than the model tested by cross-validation 10-fold test model option.

## 6.1.2. Confusion Matrix

The confusion matrix is a useful tool for analyzing how well the classifier can recognize records of different classes. The confusion matrix for decision tree illustrated in **Table 3** shows that out of the total 177 records 114 records are correctly classified as class "NO" and 20 records are correctly classified as class "YES". From the total 177 records 38 records are incorrectly classified as class "NO" and 5 records are incorrectly classified as class "YES". As whole, there were 43 instance/records incorrectly classified under both classes Yes and No.

The better accuracy was predicted using use training set test model method as class Yes and no was 75.70% were achieved as shown in **Table 3**.

Table	3.	Confusion	matrix.
Table	э.	Confusion	matrix.

	Confusion Matrix	
А	В	Predicted as
114	5	a = no
38	20	b = yes

#### 6.1.3. Model Building

In this portion, the researchers attempted to develop a model that enables to predict wheat stripe rust disease epidemic to occur. For the model development total 177 instance and four variables/attribute with target class were used for training machine and test/validate the developed model. The model validations were use training set test were used for learning parameters of the model in order to produce hypothesis and evaluates the accuracy the model in predicting. The effectiveness of the J48 un-pruned decision tree algorithm was checked using cross-validation 10-fold model validation method. The overall accuracy of the stripe rust prediction model was inspected using the confusion matrix computing, true positive rate and precision were used to evaluate the model performance.

The researcher has extracted rules that are believed to be unambiguous, relevant and novel to the domain experts to forecast the occurrence/absence of wheat stripe/yellow rust disease. The prediction rules are used for knowledge expression in the form of IF-THEN rules. The following the rule extracted from J48 un-pruned decision tree to predict wheat stripe rust disease existence that assists the stakes holder to take better decision to proactive control of the disease.

## Rule 1:

IF Maximum Temperature <= 23.07 AND Minimum Temperature <= 9.67: THEN YES (22.0/5.0)

## Rule 2:

IF Maximum Temperature <= 23.07 AND Minimum Temperature >= 9.67: THEN NO (110.0/36.0)

#### Rule 3:

IF Maximum Temperature > 23.07 AND Rainfall > 6.7 AND Minimum Temperature > 9.73: THEN NO (4.0)

## Rule 4:

IF Maximum Temperature > 23.07 AND Rainfall > 6.7 AND Minimum Temperature =< 9.73 AND Relative Humidity > 69: THEN YES (3.0)

#### Rule 5:

IF Maximum Temperature > 23.07 AND Rainfall > 6.7 AND Minimum Temperature =< 9.73 AND Relative Humidity =< 69: THEN NO (5.0/1.0)

#### Rule 6:

IF Maximum Temperature > 23.07 AND Rainfall =< 6.7: THEN NO (33.0/1.0)

These rules were generated from J48 decision tree un-pruned algorithm and learned and validated using use training set WEKA machining learning data analysis software.

#### 6.1.4. Model Performance Evaluation

The model selection criteria were done based on the statistical summary obtained from the WEKA open source software machine learning. Based on the evaluation parameters were tested learning and validation model method to compare each option method was done using the two test options. The accuracy of the model, true positive rate, precision and false positive rate were computed and compared between cross-validation 10-fold and use training set machine WEKA machine learning and validation method. The performance of each model learning and validation method was done with four predictor variables and the response class variable. The detail of this was discussed under topic 6.1.1 results analysis above.

## 6.2. Stem Rust Experiment

Stem rust experiments were done with J48 pruned decision tree algorithm with the same variable and 221 instances. Figure 9 shows the results of stem rust experiment with cross-validation 10-fold and percentage split 66% machine learning and validation method. Cross-validation fold 10 training and test method is divided the data sets into ten folds and turn by turn one-fold for testing and nine for training the model. The percentage split 66% is 66% percent of data sets for training the model and the rest percent 34% datasets for validation/testing the developed model. Stem rust model were experimented with both this machine learning and test option/methods.

#### 6.2.1. Result Analysis

This experiment of pruned J48 decision tree experimental train and test with test option cross-validation folds 10 and percentage split 66% test mode were used. The total four variables/independent attributes were used in stem rust experiments for model stem rust predictive model. For stem rust experiments 221 mean values of consecutive three days intervals of weather instances were used during wheat production.

In the case of test mode cross-validation folds were 199 instances correctly classified which is equal to 90.0452% and as well 22 instances were incorrectly classified which is equal to 9.9548% as illustrated in Figure 9. In the case of percentage split 66% mode machine learning mode 65 instances/cases were correctly classified which is equal to 86.66%. Incorrectly classified instances 10 which is equal to 13.33% as shown in Figure 10. As the experiment results show that, the model built by cross-validation machine learning train and validation mode achieved the accuracy results of 90.0452%. The number of correctly is classified more than the result achieved by percentage split 66% machine learning training and validation method.

Correctly Classified Instances	199	90.0452 %
Incorrectly Classified Instances	22	9.9548 %
Kappa statistic	0.5367	
Mean absolute error	0.1409	
Root mean squared error	0.3073	
Relative absolute error	59.3999 %	
Root relative squared error	89.7097 %	
Total Number of Instances	221	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.533	0.042	0.667	0.533	0.593	0.714	yes
	0.958	0.467	0.929	0.958	0.943	0.714	no
Weighted Avg.	0.9	0.409	0.893	0.9	0.896	0.714	

=== Confusion Matrix ===

W

a b <-- classified as 16 14 | a = yes 8 183 | b = no

Figure 9. Stem rust experiment results with cross-validation 10-fold.

Correctly Classified Instances			65		86.6667	8	
Incorrectly C	lassified	Instances	10		13.3333	8	
Kappa statistic			0.42	66			
Mean absolute	error		0.15	44			
Root mean squa	ared error		0.34	65			
Relative absol	lute error		62.44	83 %			
Root relative	squared e	rror	94.16	35 %			
Total Number of	of Instanc	23	75				
=== Detailed A	Accuracy B	y Class ===	=				
	TD Date	FD Date	Precision	Recall	F-Meagure	POC Area	Class
	0 417	0 048	0 625	0 417	0.5	0 764	VAS
	0.417	0.583	0.896	0.952	0.923	0.764	no
Weighted Avg.	0.867	0.498	0.852	0.867	0.855	0.764	
=== Confusion	Matrix ==	=					
a b < (	classified	as					
57 a=	yes						
3 60   b =	no						

Figure 10. Stem rust experiment result with percentage split 66% test option.

In the case of stripe rust experiment result, the accuracy achieved by cross-validation fold 10 was 71.75% but, the experiment result of stem rust with crossvalidation fold 10 were achieve 90.45% which is high than stripe rust. This indicates that, number of instances and J48 pruned/unpruned has a factor. With high number of instances/records, the cross-validation achieved high accuracy than a smaller number of instances/records. This supports machine learning needs huge/large number of datasets for accuracy prediction. Also the experiments done with J48 pruned algorithm achieved high accuracy.

As shown from the results of experiment, true positive rate, precision and Accuracy results obtain by pruned J48 decision tree algorithm with both cross-validation 10-fold and percentage split 66% machine learning training and validation methods were summarized in Table 4.

Algorithm	True Positive Rate		Precision		Accuracy	
	10-fold	Percentage spilt 66%	10-fold	Percentage spilt 66%	10 folds	Percentage spilt 66%
J48 pruned	66.66%	62.50%	53.33%	41.66%	90.045%	86.66%

Table 4. Summary of model test options of stem rust experiment result.

As shown in **Table 4**, the tested model has accuracy of pruned J48 decision tree algorithm 90.045% using 10-fold cross-validation and 86.66% accuracy using percentage spilt 66%. As well as the model tested has true positive rate 66.66% using 10-fold cross-validation and 62.5% using percentage spilt 66% test option. The results of precision were 53.33% using 10 cross-validation and 41.66% using percentage spilt 66% machine learning model test method were achieved. As the experiment results shows in **Table 4**, the results achieved by cross-validation fold 10 evaluation option were achieved higher results than experiment results achieved by percentage spilt 66% test option. The experiment results of true positive rate, precision and accuracy using cross-validation model test option were higher than the experiment results achieved using percentage spilt model test option. This indicates that the best decision tree model produced by J48 pruned using cross-validation model test/validation option shows better performance evaluation than the model tested by percentage spilt 66% test model option.

The model test option shows that accuracy is high in the case of more training data sets used. As the number of training data set increase the accuracy of prediction model also increase.

#### 6.2.2. Confusion Matrix for Stem Rust Experiment Result

The confusion matrix also was analyzed for the experiment result of stem rust model developed. The confusion matrix for J48 pruned decision tree showed in **Table 5** shows that out of total 221 records 183 records are correctly classified as class "NO" and 16 records are correctly classified as class "YES". From the total 221 records 14 records are incorrectly classified as class "NO" and 8 records are incorrectly classified as class "NO" and 8 records are incorrectly classified as class "NO" and 8 records are incorrectly classified as class "NO" and 8 records are incorrectly classified as class "YES". As whole, there were 22 instance/records incorrectly classified under both classes Yes and No.

The better accuracy was predicted using cross-validation fold 10 model test method as class Yes and No was 90.045% were achieved as shown in **Table 5**. In the case of stem rust experiment data sets more than the dataset used for stripe rust experiment. Cross-validation model fold 10 test option were achieved high accuracy with more training dataset used for model development.

## 6.2.3. Model Building

For stem rust experiment test total 221 instance and four variables/attribute with target class were used for training machine and test/validate the developed model. The model validations cross-validation fold 10 test were used for learning

	Confusion Matrix		
А	В	Predicted class	
16	14	a = yes b = no	
8	183		

Table 5. Confusion matrix for stem rust experiment result.

parameters of the model in order to produce hypothesis and evaluates the accuracy of the model in predicting. The effectiveness of the J48 pruned decision tree algorithm were checked percentage spilt 66% model test option. The overall accuracy of the stem rust prediction model was inspected using the confusion matrix computing, true positive rate and precision were used to evaluate the model performance. The researcher has extracted rules that relevant and novel to the domain experts to forecast the occurrence/absence of wheat stem rust disease. The prediction rules are used for knowledge expression in the form of IF-THEN rules. The following the rule extracted from J48 pruned decision tree to predict wheat stem rust disease existence prediction.

# Rule 1:

IF Minimum Temperature <= 10.7: THEN NO (143.0/9.0)

## Rule 2:

IF Minimum Temperature > 10.7 AND <= 12.17 AND Maximum Temperature <= 19.83: THEN NO (18.0)

## Rule 3:

IF Minimum Temperature > 10.7 and <= 12.17 and Maximum Temperature > 19.83 AND <= 22.53 AND relative humidity <= 77.33: THEN YES (27.0/6.0)

## Rule 4:

IF Minimum Temperature > 10.7 AND <= 12.17 AND Maximum Temperature > 9.83 AND <= 22.53 AND Relative Humidity > 77.33: THEN NO (4.0)

In this case Minimum Temperature used as splitting attribute. This implies that minimum temperature has high information gain that other, for that matter chosen as root tree.

## 6.2.4. Stem Rust Model Performance Evaluation

The same procedure was done for stem rust model performance evaluation like stripe rust. For developing and validation model both test option cross-validation fold 10 and percentage spilt were used. This test options were compared to each other on accuracy of the model, true positive rate, precision and false positive rate. The performance of each model learning and validation method was done with four predictor variables and the response class variable. Based on these parameters, the best performed model was selected. This statically information are indicators for the developed model performance evaluation.

## 6.2.5. ROC Analysis for J48 Decision Tree Model

ROC area analysis power full tools to select the best models. ROC analysis is re-

lated in a directly measure the cost/benefit analysis of control decision making. **Figure 11** shows the area under ROC for prediction stem rust instances. Class value yes gives the ROC accuracy of 71.38% of algorithms selected from all 4 attribute with un-pruned experiment.

# 7. Conclusions and Recommendation

Wheat is one of the most important crops in the world in production and food. It is a source of food and essential for smallholder farmers in developing country; however, is several abiotic and biotic danger for wheat production which was cause to reduce the quality as well quantity of wheat production. Wheat stripe rust, and stem rust are among the most destructive rust disease which reduces/damages wheat yield. In Ethiopia especial stripe rust and stem rust are more destructive rust diseases. This epidemic can travel a long distance in a little time and damage large amount of wheat production. To control this wheat rust diseases, the domain experts visually observe and assess the disease incidence and its severity after the diseases happen to action. This way is very dangerous because once the epidemic exists, it is very difficult to recur the infected wheat crop. Biotic factors such as rainfall, temperature and relative humidity are among the most environment factors that cause the wheat rust to exist. The favorable environment, susceptibility of the host and virulent of the pathogen were the main factors for rust disease to occur. Therefore, it is very important to early warn the existence rust disease to reduce the risk due to late response.

This research study was attempted to develop nonlinear model to predict stripe rust and stem rust disease. The prediction model was developed based on historical weather data and disease rust data from during year 2010 to 2019



Figure 11. ROC curve of the decision tree model for stem rust.

wheat production main season. Rust diseases of selected two bread wheat varieties of *Kakaba and Danda'a* were used for prediction model development. The weather and disease data from mid-august to November during year 2010 to 2019 wheat production main season were used for the research purpose. The weather variable daily rainfall, daily minimum temperature, daily maximum temperature and daily relative humidity were used for the study. The fist arrival rust disease incidence and severity assessed in three days interval were combined with mean values of weather datasets for experiments. The dates of disease incidence and severity observation were matched with weather variables and correlation analyzed. The weather variable was used predictor and disease incidence and severity as response class, predicted class whether the disease occur or not.

In order to achieve the objective of the study, the authors used J48 decision tree algorithm for data analysis. J48 un-pruned decision tree algorithm was used to develop model of stripe rust and J48 pruned decision tree algorithm was used to develop stem rust prediction model. During experimentation, the developed model was trained and validated with cross-validation folds 10. This test methods were compared with each based on static's information output like precision, accuracy, and true positive rate and ROC areas and the best performed model were selected. The study was found that, stripe rust prediction model developed with use training set training and test option accuracy 75.70% was achieved. The second results finding was stem rust prediction model. The study was found that, the stem rust developed prediction accuracy 90.045% were achieved using cross-validation fold 10 train and test WEKA machine learning method. This found is very promise to early warning of the existence/absence of the pathogen.

In recommendation, rust disease is destructive; incidence and severity assessment are difficult, and time-taking to make decision. Hence, early preventive mechanisms like simulation and image detection are recommended.

# **Conflicts of Interest**

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

# References

- John Dodds, Director of Research, International Maize and Wheat Improvement Center (CIMMYT), Safeguarding the World's Wheat Harvests from Stem Rust: A Global Initiative, Apdo. <u>http://www.cimmyt.org</u>
- [2] Ethiopian Commodity Exchange Authority (2008) Understanding Wheat: A Review of Supply and Marketing Issues, Addis Ababa, Ethiopia.
- [3] CSA (Central Statistical Agency) (2010) Large and Medium Scale Commercial Farms Sample Survey. Statistical Report on Area and Production of Crops, and Farm Management Practices. Statistical Bulletin 505. Addis Ababa.
- [4] Marasas, C.N., Smale, M. and Singh, R.P. (2004) The Economic Impact in Developing Countries of Leaf Rust Resistance Breeding in CIMMYT-Related Spring Bread

Wheat. Economics Program.

- [5] Singh, R.P., Hodson, D.P., Huerta-Espino, J., Jin, Y., Njau, P., Wanyera, R., Herrera oessel, S.A. and Ward, R.W. (2008) Will Stem Rust Destroy the World's Wheat Crop? *Advances in Agronomy*, **98**, 271-309. https://doi.org/10.1016/S0065-2113(08)00205-8
- [6] FAO's Wheat Rust Diseases Global Programme (2014-2017).
- Plant and Pest Diagnostic Clinic University of Nebraska-Lincoln 448 Plant Sciences Hall Lincoln, NE 68583-0722 (402) 472-2559 Publication Number: EC 1899.
- [8] International Center for Agricultural Research in the Dry Areas (ICARDA)—Research to Action—Strategies to Reduce the Emerging Wheat Stripe Rust Disease International Wheat Stripe Rust Symposium, Aleppo, Syria. 2011.
- [9] Fetanat, H., Mortazavifar, L. and Zarshenas, N. (2015) The Application of Data Mining Techniques in Agricultural Science.
- [10] Sathiamoorthy, S., Ponnusamy, R. and Natarajan, M. (2018) Sugarcane Disease Detection Using Data Mining Techniques. *International Journal of Research in Advent Technology*, Special Issue, 296-301. <u>http://www.ijrat.org</u>
- [11] Namita, M. and Smitha, H. (2016) Application of Data Mining in Agriculture Field. *International Journal of Computer Engineering and Applications*, Special Issue, 94-105.
- [12] University of Nebraska-Lincoln Extension Educational Programs Abide with the Nondiscrimination Policies of the University of Nebraska-Lincoln and the United States Department of Agriculture, 2012.
- [13] Tamene, M., Chemeda, F. and Bekele, H. (2018) Analysis of Climate Variability Effects on Wheat Stem Rust (*Puccinia graminis* f. sp tritici) Epidemics in Bale and Arsi Zones of Oromia Regional State, Ethiopia. *American Journal of Biological and Environmental Statistics*, 4, 49-65. <u>https://doi.org/10.11648/j.ajbes.20180402.12</u>
- [14] Patel, A.A. and Kathiriya, D.R. (2017) Data Mining Trends in Agriculture: A Review. *AGRES—An International Journal*, **6**, 637-645.
- [15] Han, J.W., Kamber, M. and Pei, J. (2006) Data Mining Concepts and Techniques. Elsevier, San Francisco.
- [16] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1997) From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17, 37-54.
- [17] Harding, A.C.J. (2009) Data Mining in Manufacturing: A Review Based on the Kind of Knowledge. *Journal of Intelligent Manufacturing*, 20, 501-521. https://doi.org/10.1007/s10845-008-0145-x
- [18] Rao, D.R., Pellakuri, V., Tallam, S. and Harika, T. (2016) Performance Analysis of Classification Algorithms Using Healthcare Dataset. *International Journal of Computer Science and Information Technologies*, 6, 1103-1106.
- [19] Lakshmi, K.R. and Prem Kumar, S. (2013) Utilization of Data Mining Techniques for Prediction of Diabetes Disease Survivability. *International Journal of Scientific* & Engineering Research, 4, 933.
- [20] Debele, S.E. and Koricha, D. (2010) Linkage of Climate Variability, Rust Diseases and Wheat Production in Parts of Central Ethiopia.
- [21] Vijaya Kumar, P. (2014) Development of Weather-Based Prediction Models for Leaf Rust in Wheat in the Indo-Gangetic Plains of India. Central Research Institute for Dryland Agriculture, Santoshnagar. <u>https://doi.org/10.1007/s10658-014-0478-6</u>
- [22] Siri, K. and Ha, K. (2006) Empirical Study on Applications of Data Mining Tech-

niques in Healthcare. *Journal of Computer Science*, **2**, 194-200. https://doi.org/10.3844/jcssp.2006.194.200

- [23] Chhieng JMaC. (2007) Data Mining and Clinical Decision Support Systems.
- [24] Vijiyarani, S. and Sudha, S. (2013) Disease Prediction in Data Mining Technique— A Survey. *International Journal of Computer Applications & Information Technol*ogy, 2, 2278-7720.
- [25] Rokach, L. and Maimon, O. (2015) Decision Trees, Data Mining and Knowledge Discovery Handbook. Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv.
- [26] Society IEMaS (2010) International Congress on Environmental Modelling and Software Modelling for Environments. Fifth Biennial Meeting, Ottawa.
- [27] Performance Evaluation of the Data Mining Models, Chapter 5.