Fuzzy Rough Set Theory (FRST) Classifier for Pest Prediction in India

S. Shanmuga Priya[a] and Dr.M. Sengaliappan[b]

[a]
 Research Scholar,Kovai Kalaimagal College of Arts and Science,
Bharathiar University, Coimbatore.
[b]Dean - Computer Science ,Kovai Kalaimagal College of Arts and Science, Coimbatore.

Abstract: Total cotton production is affected by an important factor called cotton pests occurrence. During the growth of cotton, environmental factors having great dependence especially during climate change. In multi-disciplinary agri-technologies domain, for data intensive science, for creating new opportunities, machine learning with big data technologies is used as a high-performance computing technique and good results are produced using this. However, running time and accuracy are highly important task. Affected crop pests are only predicted using available machine learning techniques. But there wont be any accurate level of pest affection by cotton. In this paper, proposed a Fuzzy Rough Set Theory (FRST) for handling this pest prediction level. There are four major steps in this proposed work, namely, results evaluation, prediction, pre-processing and dataset collection. From 2010-2020, from All India Coordinated Research Project (AICRP) collected the dataset in the first stage. Irrelevant features are removed by performing pre-processing in the second stage and for prediction, needed features only considered. However, for prediction, weather parameters like Rainfall (RF), Relative Humidity-Evening(RH-E), Relative Humidity-Morning(RH-M), Minimum Temperature(Min-Temp) and Maximum Temperature(Max-Temp) are used. For pest prediction, only these weather parameters are used and from original dataset removed the remaining parameters. This produces a pre-processed dataset. For pest prediction, proposed a Fuzzy Rough Set Theory (FRST) classifier in third step. Cotton pest pre-processed dataset is used for training FRST. Activation corresponds to high approximation and lower approximation functions. Weather factors belonging to higher cotton pests are predicted in this result. For future pest, FRST network can be used as a better predictor according to pest records. Moreover, better performance is shown by proposed FRST network, when compared with traditional machine learning techniques like Random Forest (RF) and Multi-Layer Perceptron (MLP). At last, these pest prediction techniques effectiveness is measured using metrics like F-measure, Area Under the Curve (AUC), and Accuracy (ACC).

Keywords: Agriculture, Cotton, Prediction, Fuzzy Rough Set Theory (FRST) classifier and Pests.

1. Introduction

An important economic crop is cotton. In national economy, an important position is occupied by this. However, during the growth, various diseases and pests damage the cotton. About 15–20% economic loss is produced by perennial diseases and pests. Sometimes it may leads to around 50% loss. So, pests need to be controlled for growing cotton. More than 900,000 tons of cotton can be recovered annually by controlling pests (Cui et al., 2007).

Production of cotton is affected by various factors during its growth. Abnormal climate change is a most significant factor affecting the growth. Pests continuous evolution is resulted by abnormal climate change and pest should be made adaptive to environment. Quality and yield has serious influence of it and pests control is made highly a difficult one (Wu et al., 2009).

For controlling pest occurrence, various techniques are developed. According to biochemical perspectives one technique is developed for suppressing pests occurrence like biological control and pesticide screening (Luo et al., 2017); (Singh et al., 2018). They are effective insecticidal and in cotton fields, they are direct. However, highly toxic nature is exhibited by pesticides and serious residual pollution may be caused by this very often.

Subsequently for control and prevention, new pesticide types with environment friendly, low degree and high efficiency. Historical data based technique is developed in another technique and pests future occurrence can be predicted using this (Zhang et al., 2017). To forecast future cotton pest level, prediction technique based on historical data is focused in various researches.

At the same time, in pest development, an important role is played by weather. Pest incidences are highly influenced by the weather parameters like rainfall, evening humidity, morning humidity, minimum temperature and maximum temperature.

However, there won't be any prior knowledge about pests outbreak severity and time. So, for minimizing the losses, timely control measures needs to be included if prediction is followed. Hence, for reducing yield in advance, for cotton, an attempt is made for developing pest forewarning models.

In agricultural operational environments, for creating new opportunities to unravel, understands and quantifies data, an intensive process is developed using Machine Learning (ML) with high performance computing and big data technologies (Jordan & Mitchell, 2015); (Liakos et al., 2018); (Chlingaryan et al., 2018). From output and input data, system model can be build using ML algorithms (Wani & Ashtankar, 2017). Future values can be predicted using this system.

When compared with conventional statistical techniques, better results are produced by these techniques. User-specified parameters are don't have any affect on it and available data is used in this. The ML applications are multi-disciplinary, where construction is not possible in traditional algorithms based on rules or they will not produce better results. An emerging technology is ML which can be used for discovering rules and patterns in agriculture data sets with weather parameters.

Upcoming can be forecasted using ML. So, for solving prediction problems and for producing better results, use of ML technique is focused in various researches.

There are four major steps in this proposed work, namely, results evaluation, prediction, pre-processing and dataset collection. From 2010-2020, from All India Coordinated Research Project (AICRP) collected the dataset in the first stage. Irrelevant features are removed by performing pre-processing in the second stage and for prediction, needed features only considered. This produces a pre-processed dataset. For pest prediction, proposed a Fuzzy Rough Set Theory (FRST) classifier in third step.

Weather factors belonging to higher cotton pests are predicted in this result. For future pest, FRST network can be used as a better predictor according to pest records. Moreover, better performance is shown by proposed FRST network, when compared with traditional machine learning techniques like Random Forest (RF) and Multi-Layer Perceptron (MLP). At last, Matrix Laboratory (MATLAB) is used for experimenting the results.

2. Literature Review

Computer science techniques like machine learning algorithms are used for predicting crop in recent research efforts.

On cotton pests incidence, weather parameters influence is focused in the technique proposed by Raghavendra et al., (2014). Cotton pests are collected at Acharya N. G. Ranga Agricultural University, Lam farm, Guntur, from 2006 to 2010. Pooled pest's data are analysed statistically using Generalized Linear Model (GLM) and Multiple Linear Regression (REG) techniques. Statistical Analysis System (SAS) is used for analysing weather parameters.

The multiple linear regression models are used for developing regression expression for cotton pest. The determination coefficients are identified using a comparative study. For all pests types, this is same and best fitting models are GLM procedure and REG procedure. On thrips, it is significant to have all weather parameters influence and but on, whitefly, aphids and jassid, it is non-significant.

For computing association rules among weather factors and cotton pests occurrence, an Apriori algorithm is proposed by Xiao et al., (2019). A time series prediction is formulated from disease and pests occurrence prediction problem and this problem can be solved by using a technique based on Long Short Term Memory Network (LSTM). In winter and autumn, disease and cotton pests occurrence are common due to rain fall, low wind speed, humid air and moderate temperature as revealed in the association analysis.

Disease and pests occurrence is then predicted using this discovery. In cotton fields, on diseases and pest occurrence prediction, better performance is produced by LSTM as shown in experimental results and around 0.97 Area Under Curve (AUC) results are obtained. Moreover, LSTM produces better performance than traditional machine learning techniques like Random Forest (RF) and Support Vector Machine (SVM).

Patil & Mytri, (2013) proposed an intelligent system for effectual prediction of pest On cotton (Gossypium Arboreum) crop, for Thrips Tabaci Linde (Thrips)'s pest population dynamics effective prediction, an intelligent system is proposed Patil & Mytri, (2013). From College of Agriculture, Raichur, India, obtained the raw data of this proposed system. At first, following stages are used for preparing raw-pest surveillance data. They are, data transformation, Normalization and data pre-processing.

In intelligent system design, back-propagation training algorithm with feed forward Multi-Layer Perceptron (MLP) Neural Network is used. With prepared data, tested and trained the neural network. On cotton crop, in Thrips pest population dynamics prediction, proposed system's effectiveness is shown by experimental results.

Also, a relative examination is performed between the proposed framework and two of the current works. The outcomes indicated that the proposed framework dependent on feed forward neural organizations was most appropriate for successful irritation expectation.

Xiao et al., (2018) expected to anticipate the event of bugs and ailments for cotton dependent on long short term memory (LSTM) organization. To begin with, the issue of event of vermin and ailments was detailed as time arrangement forecast. At that point LSTM was embraced to take care of the issue. LSTM is a unique sort of Recurrent Neutral Network (RNN), which acquaints entryway instrument with forestall the disappeared or detonating angle issue.

It has been indicated acceptable execution in taking care of time arrangement issue and can deal with the drawn out reliance issue, as referenced in numerous writings. The trial results demonstrated that LSTM performed great on the forecast of event of nuisances and ailments in cotton fields, and yielded an AUC of 0.97. Further checked that the climate factors surely have solid effect on the event of bugs and illnesses, and the LSTM network has incredible preferred position on taking care of the drawn out reliance issue.

Seeds quality are classified using an machine learning technique according to cotton crop's various stages of growth in the system proposed by Jamuna et al., (2010). This model is trained using Multilayer Perceptron (MLP), Decision Tree (DT) and Naïve Bayes (NB), which are the machine learning methods. For implementation and training facilitating, set of 900 records various classes are used for extracting features.

The 10-fold cross validation is used for evaluating model performance. In seed cotton yield classification, same accuracy is given by MLP and DT classifier as shown in obtained results. When compared with DT classifier, MLP needs more time to build this model.

On field, an idea for deploying Wireless Sensor Network(WSN) and how disease/pest can be predicted using machine learning algorithm like NB kernel Algorithm is provided by Wani & Ashtankar, (2017). A comprehensive dataset is curated by Meisner et al., (2016) with yield information, pest management and pest are obtained from1498 commercial cotton crops in California's San Joaquin Valley during 1997 to 2008.

A key cotton pest's optimum management policy can be identified using a constructed Markov decision process model using this dataset and between pesticide's application cost and yield loss, trade-off can be balanced using this. During the first 2 weeks of June, L. hesperus are only economically optimal as shown in results where pesticide applications are targeted. Increased unprofitable harvest risk is associated with this pesticide applications.

About 46% of the perceptions in dataset included in any event one pesticide application outside of this ideal window, showing the requirement for a data-driven way to deal with crop the board. Affectability examinations on boundary irritations and decreased informational index sizes recommend that Markov Decision Process (MDP) approach gives a hearty policy-making apparatus, even in loud informational collections.

To predict cotton pests occurrence and diseases with climate factors, bi-directional Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) units is used by Chen et al., (2020). A time series problem is formulated from pests prediction occurrence problem. For solving this problem, adopted the Bi-Directional LSTM network (Bi-LSTM). On sequential data's future and past context, long-term dependencies can be captures using this.

In cotton fields, on pests prediction occurrence and diseases, better performance is shown by Bi-LSTM as exhibited in experimental results and 0.95 AUC can be produced using this. On pest occurrence and disease, strong impact is shown by climate as verified in this work, with certain influence of circulation parameters. A best analytical platform using machine learning algorithm is proposed by Durgabai et al., (2019). Analytical model building can be automated using this.

Problems can be classified and better solutions can be provided using a major machine learning technique called classification. On cotton pests data, three classifiers namely DT, NB, K-Nearest Neighbor (K-NN) are implemented in this work. For analysis, better results are provided by decision tree classifier.

3. Proposed Methodology

There are four major steps in this proposed framework. In cotton pest and two areas, records are selected and unwanted features are removed for pre-processing and analysing it in first stage. To build and test pest prediction's model, pre-processed data is split as test and training dataset in the second stage. At last, after error minimization in prediction model, performance are measured. Figure 1 shows the entire work's flowchart.
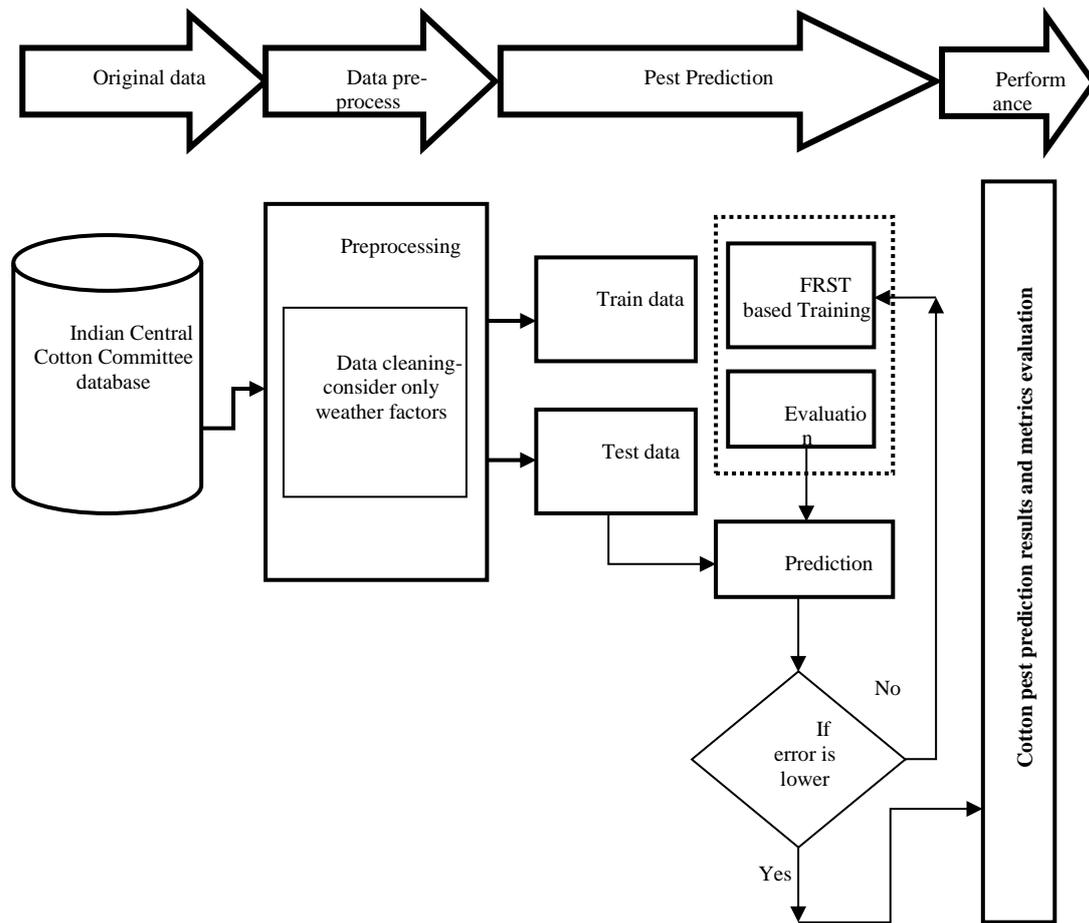
Figure 1. Overall Proposed Framework for Prediction of Cotton Pests

1. 3.1 Dataset and Data Preprocessing

From http://www.aiccip.cicr.org.in/main_aiccip_reports.html, dataset is collected initially. For pest prediction, used a technical program termed as Entomology in this website. For simulation, used the cotton pest samples from 2010-2020. With various Standard Week (SW), in Sriganganagar and Akola, during 2010-2020, monitored and recorded the Thrips population dynamics. Between two regions, varied the SWs every year. For experimentation, used Sriganganagar and Akola with abiotic factors like Rainfall (RF), Relative Humidity-Evening (RH-E), Relative Humidity-Morning (RH-M), Minimum Temperature (Min-Temp) and Maximum Temperature (Max-Temp). In the report, remaining are not used for experimentation. To effectively represent data, it is necessary to pre-process the data and in an effective manner, testing and training of machine learning classifier should be done. From Entomology reports derived from 2010-2020, noisy features are removed in this research. For implementation, considered only the samples from Sriganganagar and Akola with abiotic factors. As prediction is performed according to weather parameters, remaining factors are excluded. At last, with three levels, to predict cotton pest, Thrips are used. The levels are medium, higher and lower. From dataset, removed the attributes which are not utilized in this research work. The pre-processed dataset corresponds to those samples. For cotton pest's Akola and SG, sample during 2014-2015 are provided in table 1 and table 2. In prediction, for facilitating records, information with pest occurrence, Rainfall (RF) (mm), Relative Humidity-Evening(RH-E) (%),Relative Humidity-Morning(RH-M) (%), Min- Temp (°C) and Max-Temp(°C) is represented by every record and are adopted using Thrips.

Table 1. Cotton Pest Data from 2014-2015 in SG Region

SW	Max-Temp	Min-Temp	RH-M	RH-E	RF	Thrips- Cotton-hs6	Thrips- Cotton-rs2013
27	37.4	27.7	69.86	47.14	7.9	5.8	5.0
28	41.4	29.5	52.57	29.43	0.0	6.6	8.6
29	36.7	26.9	77.14	60.14	47.6	9.0	13.8
30	35.6	27.7	78.71	56	31.2	19.4	28.8
31	35.1	27.2	80.29	67.71	7.1	35.6	47.2
32	37.4	27.5	73.57	51	0.0	88.0	138.4
33	36.3	27.2	70	46.14	0.0	60.4	53.0
34	38.1	26.32	58.86	35.14	0.0	42.4	31.6
35	33.6	24.9	81.29	64.29	4.9	28.4	7.40
36	32.1	27.2	92.86	80	226.6	12.6	4.0

Table 2. Cotton Pest Data from 2014-2015 in Akola Region

SW	Max-Temp	Min-Temp	RH-M	RH-E	RF	Thrips-DCH32
36	28.80	22.70	64.90	92.90	109.20	8.1
37	30.30	22.60	65.30	88.40	0.70	23.3
38	32.50	23.10	56.30	90.40	0.50	6.4
39	34.50	20.70	37.00	81.40	2.00	5.1
40	36.50	21.10	29.30	73.10	0.00	10.4
41	36.80	20.90	26.30	66.00	0.00	3.6
42	34.50	21.80	37.10	75.70	0.00	1.4
43	31.90	18.00	36.60	77.40	0.00	1.4
44	33.10	15.90	20.70	68.30	0.00	0.5
45	33.50	16.60	28.00	69.40	0.00	0.2
46	30.00	20.40	46.00	87.10	20.10	1.1
47	31.70	12.90	15.70	72.00	0.00	0.5

2. 3.2 Pest Prediction

Cotton pests occurrences are predicted using this work. Assume a as pest data's weather feature records vector set and b as cotton pests occurrence. With a specified training feature vectors $(a_{t0_i}^i, b_{t0_i}^i), i = 1\ldots N$, relationship between a_i and b_i, are captured by building a model and for future test vectors $a_{t1_j}^j, j = 1, j = 1\ldots M$, non-occurrence $(b_{t1_j}^j, j = 1, j = 1\ldots M)$ or rate $\left(b_{t1_j}^j, j = 1, j = 1\ldots M\right)$ are identified, where time t0 may be earlier than time t1.

Based on pest values (b) and pest data's past weather factors (a), a binary classification problem can be formulated from prediction problem. The Fuzzy Rough Set Theory (FRST) is used in this work for performing prediction. Two different uncertainty in data called incompleteness and vagueness are dealt using this mathematical model called FRST (Zhang & Zhan, 2019); (Guo et al., 2018). Rough set theory (Das et al., 2018)

is integrated with fuzzy set theory (Lin et al., 2018) for developing this. In two ways, class C is approximated by rough sets using pest weather factors.

Elements certainly belonging to C are there in lower approximation, elements possibly belonging to it are there in upper approximation. In cotton pest data, there won't be any uncertainty if these two pest data's weather factors sets are equal. In other conditions, according to pest data's observed weather factors, C cannot be conclusively described and it can be approximated alone.

For obtaining useful results, pest data's real-valued weather factors must be discretized in set theory, which is a major limitation of it. This issue is addressed by fuzzy rough set theory extension. Removed the discretization requirement via measurement of similarity between pest data's weather factors. A fuzzy set is produced by class C's fuzzy rough lower approximation and is given in expression (1).

$$C_{lo}(a) = \min_{b \in Ts}[I(R(a,b), C(b))] \tag{1}$$

Where, training set is represented as Ts, for pest prediction, indiscernibility or similarity between weather factors are represented as $R(\cdot, \cdot)$. A fuzzy logic operator is expressed as $I(R(a,b), C(b)) = I(x,y), x = R(a,b), y = C(b)$ and it is termed as implicator. Training weather factor's membership degree to this class is represented using values $C(\cdot)$. Decision class is represented as C and only two values are taken by this. Element y's not availability is represented using 0 and its availability is represented as 1.

That is $[0,1]^2 \rightarrow [0,1]$, in first argument, there will be decrease in mapping and there will an increase in second argument and boundary conditions $I(0,0) = I(0,1) = I(1,1) = 1$ and $I(1,0) = 0$ are satisfied by this. As only two values 0 and 1 are taken by C(b), expression (2) can be formed by simplifying expression (1) using popular choices of I like Łukasiewicz implicator ($I(x, y) = \min(1 - x + y, 1)$) or Kleene-Dienes implicator ($I(x, y) = \max(1 - x, y)$)

$$C_{lo}(a) = \min_{b \notin C}(1 - (R(a,b)) \tag{2}$$

For *C's* lower approximation, *a*'s membership is given using complement to one of its similarity with highly similar weather factors which are not in C. The *C's* fuzzy rough upper approximation membership is expressed as,

$$C_{up}(a) = \max_{b \in Ts}[\mathcal{T}(R(a,b), C(b))] \tag{3}$$

Where, a triangular norm (t-norm) is taken by fuzzy logic operator $\mathcal{T}$, in both arguments, there will be an increase in associative and commutative $[0,1]^2 \rightarrow [0,1]$ mapping and boundary condition $(\forall x)(\mathcal{T}(x, 1) = x)$ is satisfied by this. Expression (4) can be formed by simplifying expression (3) as C(b) takes only two values 0 and 1.

$$C_{up}(a) = \max_{b \in C} R(a,b) \tag{4}$$

With a training weather factors of *C*, *a's* largest similarity is given by this. The pest's weather factor's membership degree is given in expression (2) and (4). The decision classes fuzzy rough approximations are computed according to the similarity with pest data's single training weather factors. Therefore, these procedures exhibits high susceptibility to noise. Noise is reduced using introduced fuzzy rough sets based on OWA (Cornelis et al., 2010).

In expression (1) and (3), Ordered Weighted Average (OWA) (Jin et al., 2019) aggregations are used for replacing maximum and minimum operations. All elements weight vector $We = \langle w_1, \ldots w_n \rangle$ are used in set of values $V = \{v_1, \ldots, v_n\}$ OWA aggregation and are drawn from 0 to 1 and the sum is 1. In an ordered sequence, according to its position, in V, elements are assigned with these weights. Concretely, performed two steps in this.

1. In descending order, elements in V are sorted. Assume $S = \langle s_1, \ldots s_n \rangle$ as this sorted sequence, where, in V, i[th] largest value is represented as s_i.
2. As discussed in expression (5), V's OWA aggregation is computed.

$$OWA^W(V) = \sum_{i=1}^{n} w_i s_i \tag{5}$$

For upper and lower approximations, weight vectors W_{up} and W_{lo} used appropriately using Fuzzy Rough Set Theory (FRST) model based on OWA (Jin et al., 2019), which softens the maximum and minimum. Expressions (6) and (7) are used for replacing expression (1) and (3).

$$C_{lo}(a) = OWA_{b \in Ts}^{W_{lo}}[I(R(a,b), C(b)] \tag{6}$$

and

$$C_{up}(a) = OWA_{b \in Ts}^{W_{up}}[\mathcal{T}(R(a,b), C(b)] \tag{7}$$

The fuzzy relation $R(\cdot, \cdot)$ is used in this study for measuring similarity between pest data's two weather and is expressed as,

$$R(a,b) = \frac{1}{|X|}\sum_{x \in X} R_x(a,b) \tag{8}$$

Where, feature set is represented as X, which is a pest data's weather factor. Between pest data, weather factor-wise similarity is expressed using function $R_x(\cdot, \cdot)$ and using expression (9), it is computed.

$$R_x(a,b) = 1 - \frac{|x(a) - x(b)|}{range(x)} \tag{9}$$

Where, numeric weather factor is represented as *x* and expression (9) can be modified as,

$$R_x(a,b) = \begin{cases} 1 & if\ x(a) = x(b) \\ 0 \end{cases} \tag{10}$$

Where, nominal value is represented as *x*. Any additional information is not carried by classes upper approximation on top of it as shown with only two possible class labels that are represented using lower approximation. In classification step, computed the two classes lower approximation of membership and pest prediction uses the class to which this value is highest. Used the alternative expression (5) based on OWA.

For OWA aggregation, in a class-dependent weight, incorporated the distinction between majority and minority classes. For two classes, different weight vector W_{lo} definitions can be used. There will be a slight difference in these weights definitions when compared with original proposal, but it can be interpreted clearly. In specific, instead of expression (5), in expression (2), this description is used in OWA integration.

Expressions (11) and (12) indicates negative (non pest) classes *N* and positive (pest) classes *P*'s lower approximations based on weight vectors W_{lo}^+ and W_{lo}^-.

$$P_{lo}(a) = \underset{b \notin P}{OWA}^{W_{lo}^+}\big(1 - R(a,b)\big) = \underset{b \in N}{OWA}^{W_{lo}^+}\big(1 - R(a,b)\big) \tag{11}$$

$$N_{lo}(a) = \underset{b \notin P}{OWA}^{W_{lo}^-}\big(1 - R(a,b)\big) = \underset{b \in P}{OWA}^{W_{lo}^-}\big(1 - R(a,b)\big) \tag{12}$$

In order to ensure the highly balanced contribution, weights increase is opposed to exponential weights. The FRST is also used for identifying occurrence of pests in future. The extent to which prediction can be done using historical observation should be computed.

3. 3.3 Performance Measurement

Techniques performance needs to be measured after predicting pests. Predcition techniques effectiveness is measured using F-measure, Area Under the Curve (AUC) and Accuracy (ACC). Two types of classes called negative and positive are produced as an output of every binary classification and are recorded as N and P. For case predictions, there are four outcomes in bivariate model called False Negative (FN), False Positive (FP), True Negative (TN) and True Positive (TP).

4. Experiment and Results

Detailed layout of this study, dataset specification, evaluation measures and tests utilized for statistical analysis are presented in this section. Proposed Fuzzy Rough Set Theory (FRST) classifier is compared with state-of-the-art classifiers like Random Forest (RF) and Multi-Layer Perceptron (MLP).

To compare with FRST model, for cotton pests occurrence prediction, implemented the classification models. Using Intel (R) Core (TM) i7-4790 CPU operating at 3.60GHz (8CPUs) with 8G RAM with Windows 10 64 bits operating system is used to run the experimentation. MATLAB is used for programming.

4. 4.1 Performance Evaluation Metrics

In pest prediction, popular evaluation measures used for evaluating various technique's performance are F-Measure, Area Under the Curve (AUC) and Accuracy (ACC). Recall and precisions harmonic mean defines F measure. For computing the score, both <u>Recall</u> (REC) and <u>Precision</u> (PRE) values are considered.

Ratio between correct positive results count to total positive results count returned by classifier defines p and ratio between correct positive results count to all relevant samples count (all samples that needs to be identified as positive) defines r. Expression (13) describes these values.

$$F - MEASURE\ = 2.\frac{PRE.REC}{PRE+REC} \tag{13}$$

Precision is also termed as Positive Predictive Value (PPV). Expression (14) describes this precision.

$$PRE = \frac{TP}{TP+FP} \tag{14}$$

In this context, recall is termed as sensitivity or True Positive Rate (TPR). Expression (15) describes this recall.

$$REC = \frac{TP}{TP+FN} \tag{15}$$

Ratio between correct predictions count to total predictions count defines the accuracy. In order to turn it in percentage, it is multiplied with 100. Expression (16) describes this.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{16}$$

Classifier performance is evaluated using introduced Area Under the ROC Curve (AUC), in addition to Receiver Operating Characteristic (ROC) curve. Expression (17) indicates the AUC definition.

$$AUC = \frac{\sum_{i \in positive\ class} rank_i - \frac{M(M+1)}{2}}{M \times N} \tag{17}$$

Where, positive class count is represented as M and negative class count is represented as N. Every samples predicted using this model are sorted based on its probability value from small to large. The ith samples serial number is represented as $rank_i$, i = 1, ..., n, total data is given by n and n = M + N.

5. 4.2 Evaluation Results

Various metrics like Area Under the ROC(AUC), ACCuracy(ACC), F-measure, RECall(REC), PREcision (PRE) are used for analysing three classifiers like RF, MLP and FRST's results and they summarized in table 3.

Table 3. Cotton Pest Methods Vs. Metrics (Two Regions)

	AKOLA REGION	SG REGION

METRICS /METHODS	RF	MLP	FRST	RF	MLP	FRST
PRECISION (PRE) (%)	71.847	82.014	**84.533**	74.837	81.297	**92.867**
RECALL (REC) (%)/SENSITIVITY (%)	72.847	83.014	**86.957**	75.837	82.297	**91.667**
F-MEASURE (%)	72.343	82.511	**85.728**	75.334	81.794	**92.263**
ACCURACY (ACC) (%)	73.171	82.927	**88.710**	75.61	82.927	**91.667**
AREA UNDER THE ROC (AUC) (%)	80.00	82.10	**92.824**	80.14	82.14	**84.828**
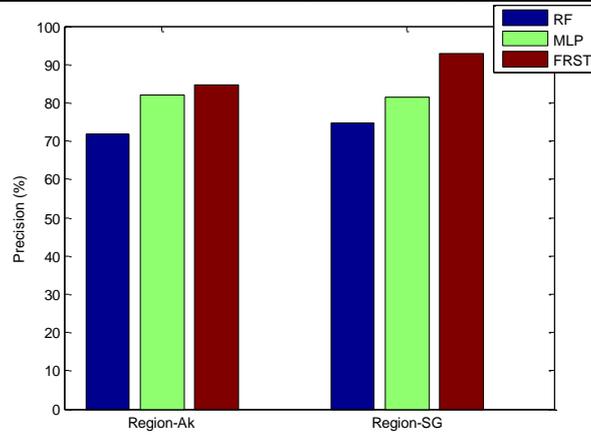
**Figure 2. Precision Results Comparison Vs. Pest Prediction Methods**
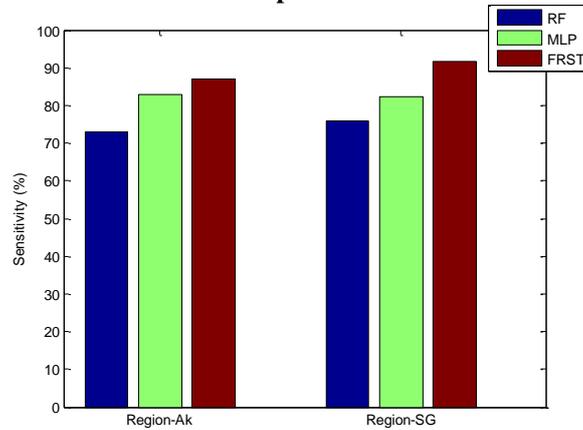
Figure 3. Recall (OR) Sensitivity Results Comparison Vs. Pest Prediction Methods

Figure 2 shows the precision results of three pest prediction techniques for two regions. Around 84.533% precision results are produced by proposed FRST prediction algorithm, which is a greater value when compared with 82.014% precision results of MLP and 71.847% F-measure results of RF for Akola region.

Figure 3 shows the recall results of three pest prediction techniques for two regions. Around 86.957% recall results are produced by proposed FRST prediction algorithm, which is a greater value when compared with 83.014% precision results of MLP and 72.847% F-measure results of RF for Akola region.
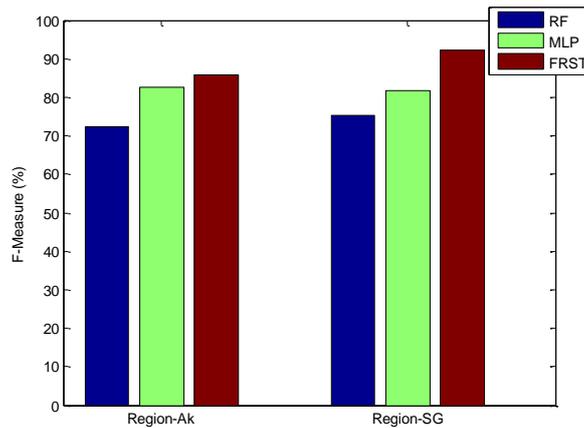
Figure 4. F-Measure Results Comparison Vs. Pest Prediction Methods

Figure 4 shows the F-measure results of three pest prediction techniques for akola. Around 85.728% F-measure results are produced by proposed FRST prediction algorithm, which is a greater value when compared with 82.511% F-measure results of MLP and 72.343% F-measure results of RF for Akola region.
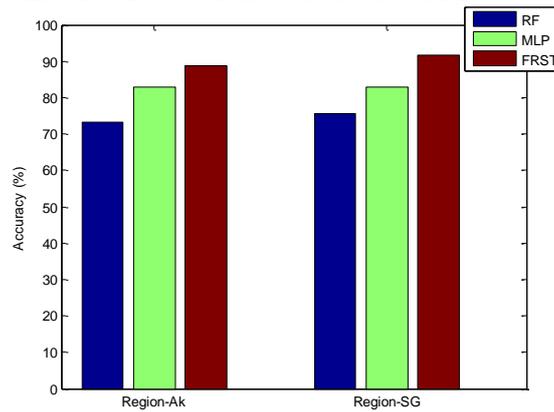

Figure 5. Accuracy Results Comparison Vs. Pest Prediction Methods

Figure 5 shows the accuracy results of three pest prediction techniques for akola. Around 88.710% accuracy results are produced by proposed FRST prediction algorithm, which is a greater value when compared with 82.10% accuracy results of MLP and 80.00% accuracy results of RF for Akola region.
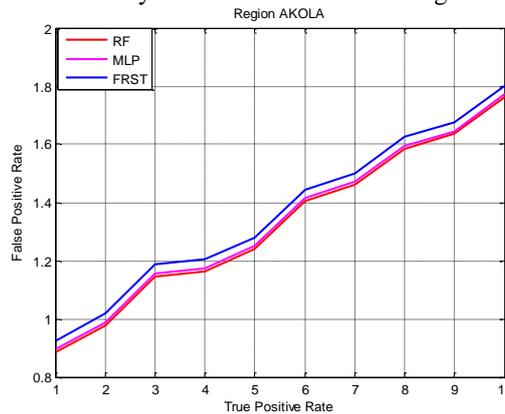
Figure 6. ROC Value vs. Pest Prediction Methods (Akola Region)

Figure 6 shows the ROC curve of three pest prediction techniques for akola. Around 92.824% AUC results are produced by proposed FRST prediction algorithm, which is a greater value when compared with 82.10% AUC of MLP and 80.00% of RF for Akola region.
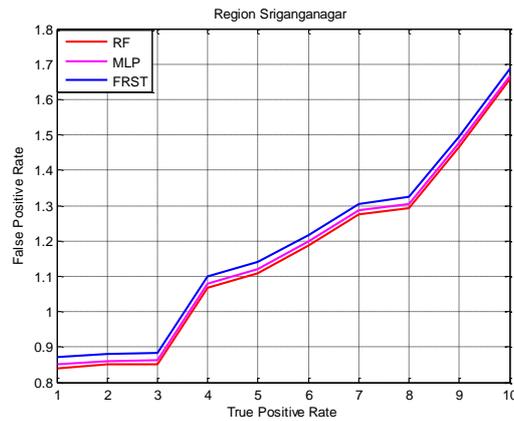
Figure 7. ROC Value Vs. Pest Prediction Methods (SG Region)

Figure 7 shows the ROC curve of three pest prediction techniques for SG. Around 84.828% AUC results are produced by proposed FRST prediction algorithm, which is a greater value when compared with 82.14% AUC of MLP and 80.14% of RF for SG region.

5. Conclusion and Future Work

Proposed a classifier based on Fuzzy Rough Set Theory (FRST) in this paper for predicting future cotton pets occurrence according to historical data like pets data and weather factors. For cotton pets future prevention and control, it is very important and for agriculture development. In two ways, pest weather factors are approximated in FRST namely, higher approximation and lower approximation. This work introduces a FRST model based on OWA for enhancing FRST classifier's accuracy.

For upper and lower approximations, weight vectors W_{up} and W_{lo} are used appropriately for softening maximum and minimum accordingly. Weight vector W_{lo} is used for modelling pest data and weight vector W_{lo} is used for mapping the output as a final pest prediction. For illustrating the prediction comparison with FRST model, implemented some traditional machine learning techniques like Multi-Layer Perceptron (MLP) and Random Forest (RF).

With respect to Area Under the ROC(AUC), ACCuracy(ACC), F-measure, RECall(REC) and PREcision (PRE), certain advantages are shown by FRST as indicated in results. However, weather factors based cotton pest's prediction is only addressed in this work. So, a model can be constructed in future to predict pests hazard level. So that prediction results are made highly responsive to data and detailed pest control strategies can be developed easily.

References

1. Cui, J. J., Chen, H. Y., Zhao, X. H., & Luo, J. Y. (2007). Research course of the cotton IPM and its prospect. *Cotton Sci*, *19*(5), 385-90.
2. Wu, K., Lu, Y., & Wang, Z. (2009). Advance in integrated pest management of crops in China. *Chinese Bulletin of Entomology*, *46*(6), 831-836.
3. Luo, J., Shuai, Z., Ren, X., Limin, L., Zhang, L., Ji, J., ... & Cui, J. (2017). Research progress of cotton insect pests in china in recent ten years. *Cotton Sci*, *9*, 100-12.
4. Singh, S., Gupta, M., Pandher, S., Kaur, G., Rathore, P., & Palli, S. R. (2018). Selection of housekeeping genes and demonstration of RNAi in cotton leafhopper, Amrasca biguttula biguttula (Ishida). *PLoS One*, *13*(1),1-21
5. Zhang, W., Jing, T., & Yan, S. (2017). Studies on prediction models of Dendrolimus superans occurrence area based on machine learning. *Journal of Beijing Forestry University*, *39*(1), 85-93.
6. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.
7. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674.
8. Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and electronics in agriculture*, *151*, 61-69.
9. Wani, H., & Ashtankar, N. (2017). An appropriate model predicting pest/diseases of crops using machine learning algorithms. *International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 1-4.
10. Raghavendra, K. V., Naik, D. B., Venkatramaphanikumar, S., Kumar, S. D., & Krishna, S. R. (2014). Weather Based Prediction of Pests in Cotton. In *2014 International Conference on Computational Intelligence and Communication Networks*, pp. 570-574.

11. Xiao, Q., Li, W., Kai, Y., Chen, P., Zhang, J., & Wang, B. (2019). Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network. *BMC bioinformatics*, *20*(25), 1-15.
12. Patil, J., & Mytri, V. D. (2013). A prediction model for population dynamics of cotton pest (Thrips tabaci Linde) using multilayer-perceptron neural network. *International Journal of Computer Applications*, *67*(4), 19-26.
13. Xiao, Q., Li, W., Chen, P., & Wang, B. (2018). Prediction of Crop Pests and Diseases in Cotton by Long Short Term Memory Network. In *International Conference on Intelligent Computing*, pp. 11-16. Springer,
14. Jamuna, K. S., Karpagavalli, S., Vijaya, M. S., Revathi, P., Gokilavani, S., & Madhiya, E. (2010). Classification of seed cotton yield based on the growth stages of cotton crop using machine learning techniques. In *2010 International Conference on Advances in Computer Engineering*, pp. 312-315.
15. Wani, H., & Ashtankar, N. (2017). An appropriate model predicting pest/diseases of crops using machine learning algorithms. In *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 1-4.
16. Meisner, M. H., Rosenheim, J. A., & Tagkopoulos, I. (2016). A data-driven, machine learning framework for optimal pest management in cotton. *Ecosphere*, *7*(3), 1-13.
17. Chen, P., Xiao, Q., Zhang, J., Xie, C., & Wang, B. (2020). Occurrence prediction of cotton pests and diseases by bidirectional long short-term memory networks with climate and atmosphere circulation. *Computers and Electronics in Agriculture*, *pp. 1-9.*
18. Durgabai, R. P. L., Bhargavi, P., & Jyothi, S. (2019). Classification of Cotton Crop Pests Using Big Data Analytics. In *International Conference On Computational And Bio Engineering* (pp. 37-45). Springer, Cham.
19. Zhang, L., & Zhan, J. (2019). Fuzzy soft $$\beta$$ β-covering based fuzzy rough sets and corresponding decision-making applications. *International Journal of Machine Learning and Cybernetics*, *10*(6), 1487-1502.
20. Guo, Y., Tsang, E. C., Xu, W., & Chen, D. (2018). Logical disjunction double-quantitative fuzzy rough sets. In *2018 International Conference on Machine Learning and Cybernetics (ICMLC)* (Vol. 2, pp. 415-421). IEEE.
21. Lin, Y., Li, Y., Wang, C., & Chen, J. (2018). Attribute reduction for multi-label learning with fuzzy rough set. *Knowledge-based systems*, *152*, 51-61.
22. Das, A. K., Sengupta, S., & Bhattacharyya, S. (2018). A group incremental feature selection for classification using rough set theory based genetic algorithm. *Applied Soft Computing*, *65*, 400-411.
23. Cornelis, C., Verbiest, N., & Jensen, R. (2010). Ordered weighted average based fuzzy rough sets. In *International Conference on Rough Sets and Knowledge Technology* (pp. 78-85). Springer, Berlin, Heidelberg.
24. Jin, L., Mesiar, R., & Yager, R. (2019). Ordered weighted averaging aggregation on convex poset. *IEEE Transactions on Fuzzy Systems*, *27*(3), 612-617.