International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

Sørensen-Dice Cuckoo Feature Selection based Gaussian Neuro Fuzzy Classification for Improved Agriculture Seed Growth

Ms. M. Indira
Assistant Professor,
Department of Computer Science,
P.K.R. Arts College for Women (Autonomous),
Gobichettipalayam, India.

Dr. S. Jayasankari
Associate Professor,
Department of Computer Science,
P.K.R. Arts College for Women (Autonomous),
Gobichettipalayam, India.

***Abstract* - Seed classification is a significant issue to be solved in data mining to increase the agriculture growth. Few research works are designed for classifying seed as normal or abnormal using different techniques. However, accuracy of seed classification was not sufficient. Further, time taken to classify the seed data was more. In order to overcome such limitations, a Sørensen–Dice Cuckoo Feature Selection Based Gaussian Neuro Fuzzy Classifier (SCFS-GNFC) Technique is proposed. SCFS-GNFC Technique is designed to classify the seed data as normal or abnormal with higher classification accuracy and minimal feature selection time and thereby enhancing agriculture growth. The SCFS-GNFC Technique initially takes number of seed data and their features as input. After that, Sørensen–Dice Similarity Based Cuckoo Feature Selection (SS-CFA) algorithm is introduced to choose relevant features from the large volume of data with lower time. By using the selected significant features, the seed classification is performed in SCFS-GNFC Technique with application of Gaussian Neuro-fuzzy classifier. During the classification process, SCFS-GNFC Technique employs Gaussian membership function and fuzzy if-then rules to precisely classify the each input data as normal or abnormal class with a minimal time. This helps for SCFS-GNFC Technique to improve the classification performance of seed data categorization with a lower error rate. The experimental evaluation of SCFS-GNFC Technique is carried out using soybean dataset on factors such as feature selection time, classification accuracy and error rate with respect to number of features and seed data. The experimental result shows that the SCFS-GNFC Technique is able to increase the classification accuracy and also reduces the feature selection time of seed classification in agriculture field as compared to state-of-the-art works.**

Keywords-Agriculture Dataset, Fitness Function, Gaussian Neuro-fuzzy classifier,Relevant Features, Seed Data, Sørensen–Dice Similarity Based Cuckoo Feature Selection

I. INTRODUCTION

In agriculture field, seed plays a vital role. The classification of seeds as the normal seed or defected is challenging issue in data mining. In this advanced era, seed classification plays major role to improve the agricultural growth. Most of the recent research focused on classification of seeds using different data mining techniques. But, the classification accuracy was not improved. In order to improve the seed classification performance with a minimal time, SCFS-GNFC Technique is introduced in this research work by using Sørensen–Dice Similarity Based Cuckoo Feature Selection (SS-CFA) algorithm and Gaussian Neuro-Fuzzy (GNF) Classifier.

Multilayer perceptron network classifier was introduced in [1] for classifying high-quality seeds from low-quality seeds. However, accuracy of Multilayer perceptron network classifier was not enough. A Hybrid Kernel based Support Vector Machine (H-SVM) was designed in [2] for categorizing the multi-class agricultural data with attributes. But, the feature selection was not carried out that resulted in maximal error rate.

The C-band, dual polarimetric and temporal satellite of RISAT-1 was discussed in [3]. But, the error rate was not reduced by using the divergence method. Decision-making tool was introduced in [4] for choosing the best crop in a given agricultural land with enhanced accuracy. However, time complexity was very higher.

A comparative result analysis were carried out in [5] for post-harvest growth discovery using geometric features. However, the classification time was not reduced due to the comparative analysis of classification process. The survey of diverse data mining techniques designed for analysis of agriculture data and thereby enhancing the crop production was presented in [6].

Hybrid rough fuzzy soft classifier was presented in [7] for agriculture crop selection with a lower time. However, error rate during classification process was higher. A hybrid Multi-Criteria decision making technique was developed in [8] for improving the diseases diagnosis performance in plants. But, feature selection was not performed in this technique.

Machine Learning Approach was designed in [9] to find proper crops according to climatic conditions and thereby maximize yield rate. But, classification accuracy was lower. A novel diagnostic method was introduced in [10] for predicting fungal pathogens on vegetable seeds. But, computational time taken for diagnosis was more.

In order to addresses the above mentioned existing issues, SCFS-GNFC Technique is introduced in this research work. The key contributions of SCFS-GNFC Technique is described in below,

- To increase the feature selection performance as compared to state-of-the-art works, Sørensen–Dice Similarity Based Cuckoo Feature Selection (SS-CFA) algorithm is proposed in SCFS-GNFC Technique as it provides robust solution for optimization problems during feature selection process. Because, proposed SS-CFA contains many advantages for example easy implementation, stable convergence characteristic and good computational efficiency on the contrary to conventional optimization algorithms. The designed SS-CFA algorithm identifies optimal features from an input dataset for efficient seeds classification with enhanced accuracy.
- To improve the seed classification performance in agricultural field as compared to conventional works, Gaussian Neuro-Fuzzy (GNF) Classifier is proposed in SCFS-GNFC Technique. The advantages of neural networks and fuzzy systems are combined as a neuro-fuzzy approach in GNF Classifier. GNF Classifier is a fuzzy network that contains a fuzzy inference system to solve some drawbacks of neural networks and fuzzy systems. Because, GNF Classifier in proposed SCFS-GNFC Technique can learns and represents knowledge in an interpretable way and learning ability. This supports for SCFS-GNFC Technique to exactly classify the seed data as normal or abnormal with a lower time.

The rest of paper is constructed as follows. Section 2 portrays the related works. Section 3 presents the exhaustive process of proposed SCFS-GNFC Technique. In section 4, an experimental setting of proposed technique is demonstrated. Section 5 provides the results and discussion of certain parameters. Finally, the conclusion of the research work is depicted in Section 6.

II. RELATED WORKS

A fuzzy-based multi-criteria decision-making was developed in [11] to discover the crop pattern. A novel method was designed in [12] to offer an effective classification of the soybean seed vigor.

A real-time, non-invasive, micro-optrode technique was introduced in [13] for determining the seed viability. A seed yield evaluation modeling using classification and regression trees (CART) was presented in [14].

A hybrid ensemble approach was designed in [15] for solving multiclass classification issues. But, the classification time was not reduced. In [16], the differentiations in seed germination and seed growth were evaluated at intra- and inter-provenance levels.

A novel predictive model was presented in [17] to find seed classes with application of machine learning algorithms to attain high crop production. A back-propagation (BP) neural network was applied in [18] to find out the seed distribution.

An analysis and impact factors on agriculture field using different data mining techniques were presented in [19]. Rough set and decision tree ensemble was employed in [20] to enhance the detection performance of agricultural data.

Based on the above existing techniques, a novel seed classification technique called SCFS-GNFC Technique is developed which detailed described in below section.

III. SØRENSEN–DICE CUCKOO FEATURE SELECTION BASED GAUSSIAN NEURO FUZZY CLASSIFIER TECHNIQUE

Sørensen–Dice Cuckoo Feature Selection Based Gaussian Neuro Fuzzy Classifier (SCFS-GNFC) Technique is proposed in order to improve the performance of seed classification for predicting the agriculture growth. The SCFS-GNFC Technique is designed by combining the Sørensen–Dice Similarity Based Cuckoo Feature Selection (SS-CFA) algorithm and Gaussian Neuro-Fuzzy (GNF) Classifier on the contrary to conventional works. Therefore, proposed SCFS-GNFC Technique gives best classification result for identifying the seed growth in the agriculture field as compared to existing algorithm. The architecture diagram of SCFS-GNFC Technique is shown in below Figure 1.

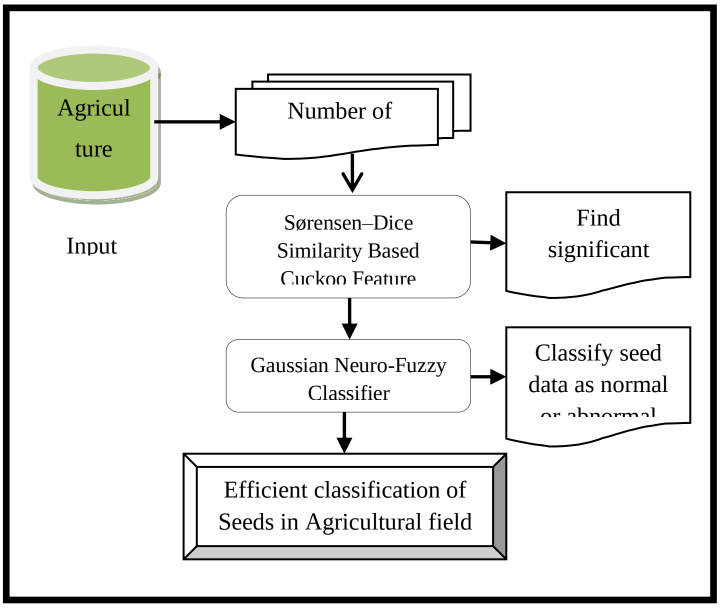

Figure 1 Architecture Diagram of SCFS-GNFC Technique for Predicting Agriculture Seed Growth

Figure 1 shows the overall process of SCFS-GNFC Technique to achieve enhanced classification accuracy for finding seed quality. As presented in above figure, SCFS-GNFC Technique initially takes agriculture dataset i.e. Soybean Dataset as input. After getting input, SCFS-GNFC Technique carried out Sørensen–Dice Similarity Based Cuckoo Feature Selection process where features that are more related for seed classification are selected with minimal time. After feature selection, SCFS-GNFC Technique applies Gaussian Neuro-Fuzzy Classifier that efficiently classifies each input seed data as normal or abnormal by using selected features with a lower amount of time. Thus, SCFS-GNFC Technique improves the seed classification performance as compared to state-of-the-art works. The detailed processes of SCFS-GNFC Technique are described in below sub-section.

A. Sørensen–Dice Similarity Based Cuckoo Feature Selection

The Sørensen–Dice Similarity Based Cuckoo Feature Selection (SS-CFA) algorithm is designed to choose the features that are more imperative for classifying seeds in agriculture field. On the contrary to conventional works, SS-CFA algorithm is proposed with application of Sørensen–Dice similarity measurement in cuckoo search optimization. The SS-CFA is an optimization algorithm which depends on the obligate brood parasitism of cuckoo species by putting their eggs in the nests of other host birds. Each egg in a nest denotes a solution (i.e. best features for seed categorization), and a cuckoo egg indicates a new solution. The aim of SS-CFA algorithm is to find optimal solutions (i.e. relevant features) for seed classification by replacing a not-so-good solution (i.e. irrelevant features) in the nests (i.e. input agriculture dataset). The SS-CFA algorithm depends on three idealized rules explained in below,

- ✓ Each cuckoo puts one egg at a time, and leaves its egg in an arbitrarily selected nest.
- ✓ The best nests with high quality eggs are considered for the next generation.
- ✓ The number of host's nests is constant, and the egg laid by a cuckoo is identified by the host bird with a probability.

Based on the above rules, SS-CFA algorithm selects the more relevant features for improving the seed classification accuracy. The process of SS-CFA algorithm is depicted in below Figure 2.

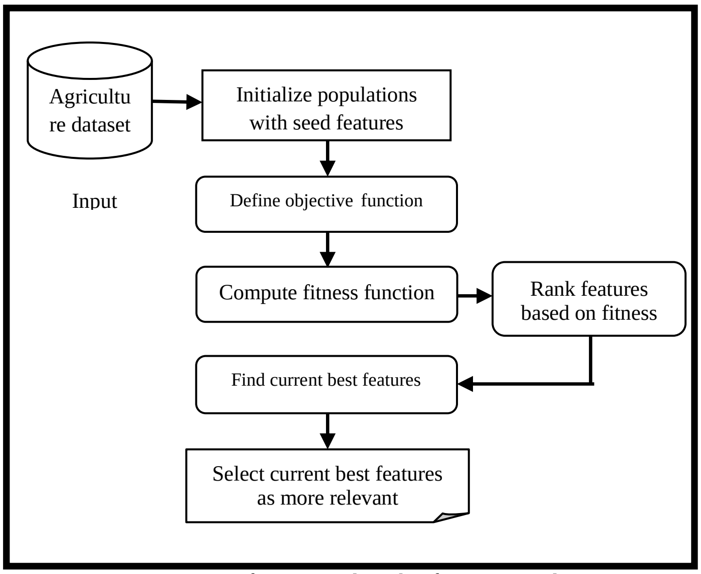

Figure 2 Processes of SS-CFA Algorithm for Feature Selection

Figure 2 demonstrates the flow processes of SS-CFA Algorithm to improve the feature selection performance with a minimal time complexity. As presented in above figure, SS-CFA Algorithm first get agriculture dataset i.e. Soybean Dataset as input which includes many number of features represented as '$\beta_i = \beta_1, \beta_2, \dots, \beta_n$'. Here, '$n$' denotes the total number of features available in input dataset. Then SS-CFA Algorithm initialize the population of 'n' hosts nests with seed features. Followed by, SS-CFA Algorithm define the objective function 'OF' using below expression,

$$OF = arg\,max\{x_{\beta_i}\} \quad (1)$$

Here, objective function is to choose the feature with higher similarity value for seeds classification whereas 'x_{β_i}' denotes the Sørensen–Dice similarity value of feature 'β_i' and '$\arg max$' helps to find the features with higher similarity value. Then, fitness of the each feature 'β_i' is computed as follows,

$$F_{\beta_i} = x_{\beta_i} \quad (2)$$

From the above equation (2), 'F_{β_i}' denotes fitness function of 'i^{th}' feature whereas 'x_{β_i}'indicates similarity value. On the contrary to existing works, the Sørensen–Dice Indexing is applied in SS-CFA Algorithm for determining the similarity between the features using below,

$$x_{\beta_i} = \log_{10} 10^2 * a \quad (3)$$

$$a = \frac{\beta_i \cap \beta_j}{|\beta_i| * |\beta_j|} \quad (4)$$

The above equation (4) is substituted to the equation (3) to acquire final similarity results,

$$x_{\beta_i} = \log_{10} 10^2 * \left(\frac{\beta_i \cap \beta_j}{|\beta_i| * |\beta_j|}\right) \quad (5)$$

From equation (5), 'x_{β_i}' represents a Sørensen–Dice similarity value, 'β_i' and 'β_j' denotes two features in the input dataset and 'a' indicates the ratio of mutual dependence between two features. The intersection symbol '$\cap$' represents a mutual dependence and '$|\beta_i|$'and '$|\beta_j|$' denotes the cardinalities of the two features. The Sørensen–Dice similarity value 'x_{β_i}'is ranges between 0 and 1. If the two features are similar, then the output value is '1'. When the two features are dissimilar, the output value is '0'.

After determining the fitness value, the current best solutions (i.e. relevant features) are selected by assigning the rank. The SS-CFA Algorithm assigns rank to each feature 'β_i' based on their fitness value using below mathematical formulation,

$$R \rightarrow \{\beta_1, \beta_2, \beta_3, \dots, \beta_n\} \quad (6)$$

From equation (6), 'R' indicates a rank. Based on the rank assigned, SS-CFA Algorithm chooses the current best features (i.e. features with higher rank) for increasing the seeds classification accuracy.

The algorithmic step of SS-CFA is explained in below.

// **Sørensen–Dice Similarity Based Cuckoo Feature Selection Algorithm**
Input: Agriculture Dataset 'AD'; Number of Features '$\beta_i = \beta_1, \beta_2, .., \beta_n$'
Output: Enhanced feature selection accuracy
Step 1:Begin
// Feature Selection
Step 2: Initialize the population with number of features 'β_i'
Step 3: For each 'β_i'
Step 4: Define objective function '$OF = arg\,max\{x_{\beta_i}\}$' using (1)
Step 5: Determine fitness function 'F_{β_i}' using (2)
Step 6: Rank 'β_i' based on 'F_{β_i}' using (6)
Step 7: Find current best β_i
Step 8: Select current best feature as relevant to perform seed classification
Step 9: End for
Step 10:End

Algorithm 1 Sørensen–Dice Similarity Based Cuckoo Feature Selection

Algorithm 1 depicts the step by step processes of SS-CFA. By using the above algorithmic steps, SS-CFA initialize the population of 'n' hosts nests with features and consequently defines objective function. Then, SS-CFA computes fitness value for each features using Sørensen–Dice Similarity measurement. After that, SS-CFA ranks the features according to their fitness value. Finally, SS-CFA chooses the features with higher rank as current best solutions to efficiently classify the seeds. From that, SS-CFA significantly selects the more significant features in input dataset with higher accuracy and minimal time consumption as compared to conventional works.

B. Gaussian Neuro-Fuzzy Classifier

The Gaussian Neuro-Fuzzy (GNF) Classifier is designed in SCFS-GNFC Technique with aiming at enhancing classification performance of seeds. On the contrary to state-of-the-art works, GNF Classifier is introduced by combining the Gaussian membership function in Neuro-Fuzzy classification. The GNF Classifier is designed to categorize the seed data as normal or abnormal and thereby predicting the seed growth in the agriculture field.

In GNF Classifier, fuzzy classification rule 'r_i' shows the relation between the input seed data features and classes which is mathematically represented as,

'r_i: $if\ d_{m1}\ is\ b_{i1}$' and '$..\ d_{mj}\ is\ b_{ij}$ ... and '$d_{mn}\ is\ b_{in}$ then class is C_k (7)

From equation (7), 'd_{mj}' denotes the 'j^{th}' feature of 'm^{th}' seed data and 'A_{ij}' indicates the fuzzy set of the 'j^{th}' feature in the 'i^{th}' rule; and 'C_k' point outs 'k^{th}' label of class. Here, 'A_{ij}' is determined with the help of membership function.

In GNF Classifier, the feature space is partitioned into multiple fuzzy subspaces with help of fuzzy if-then rules. These fuzzy rules are represented by a network structure. The GNF Classifier is a multilayer feed-forward network. The GNF Classifier in SCFS-GNFC Technique contains input, fuzzy membership, fuzzification, de-fuzzification, normalization, and output layers. Besides, GNF Classifier contains multiple inputs and multiple outputs. Figure 3 shows structure of GNF Classifier for agriculture seed classification.

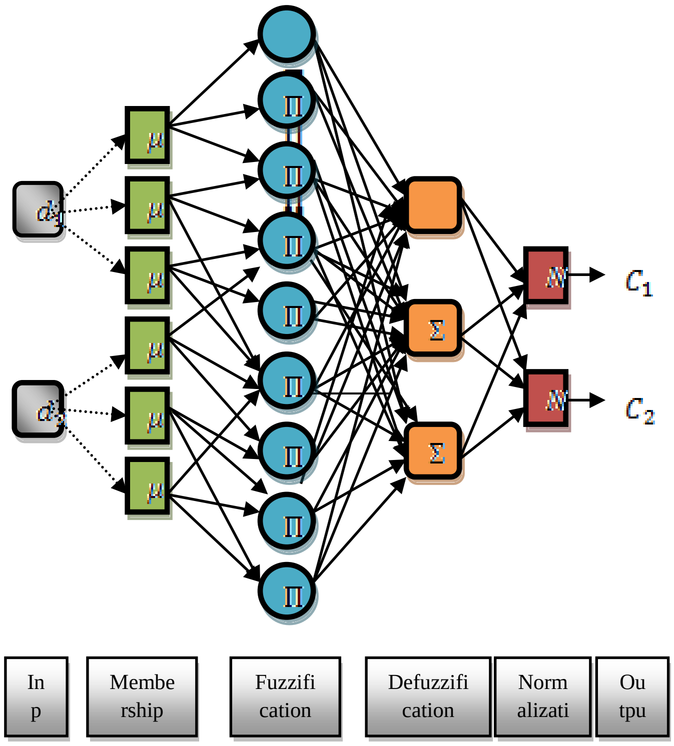

Figure 3 Structure of GNF Classifier

Figure 3 depicts the processes of GNF Classifier to enhance the accuracy of seed classification in agriculture field where input layer get the number of seed data 'd_i' from soybean dataset and then forward it to membership layer.

In Membership layer, the membership function of all input seed data is discovered. On the contrary to existing works, a Gaussian membership function is used in GNF Classifier. Because, Gaussian membership function has fewer parameters and smoother partial derivatives for parameters. Thus, Gaussian membership function is mathematically obtained as,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

$$M_{ij}(d_{mj}) = \exp\left(-\frac{(d_{mj}-c_{ij})^2}{2\sigma_{ij}^2}\right) \quad (8)$$

From the above equation (8), '$M_{ij}(x_{mj})$' represents the membership grade of 'i^{th}' rule and 'j^{th}' feature whereas 'd_{mj}' indicates 'm^{th}' seed data and 'j^{th}' feature. Here, 'c_{ij}' and 'σ_{ij}' denotes center and width of Gaussian function.

In fuzzification layer, each node produces a signal corresponding to the degree of fulfillment of the fuzzy rule for input seed data 'd_m'. This represents firing strength of a fuzzy rule with respect to a seed data to be classified. The firing strength of the 'i^{th}' rule is mathematically determined as,

$$\varepsilon_{im} = \prod_{j=1}^{N} M_{ij}(d_{mj}) \quad (9)$$

From the above equation (9), 'n' point outs the number of seed features. In de-fuzzification layer, weighted outputs are measured. If a rule controls a particular class region, the weight among that rule output and the particular class is larger than the other weights. Otherwise, the class weight is lower. The weighted output for input 'm^{th}' seed data that belongs to the 'k^{th}' class is mathematically estimated as,

$$\beta_{mk} = \sum_{i=1}^{n} \varepsilon_{im}\varphi_{ik} \quad (10)$$

From the above equation (10), 'φ_{ik}' represents the degree of belonging to 'k^{th}' class that is controlled by 'i^{th}' rule and 'n' indicates the number of rules.

In normalization layer, the outputs of network is normalized, because the summation of weights may be larger than '1' in some circumstance.

$$n_{mk} = \frac{\beta_{mk}}{\sum_{l=1}^{k}\beta_{ml}} \quad (11)$$

From the above equation (11), 'n_{mk}' indicates the normalized value of the 'm^{th}' seed data that belongs to the 'k^{th}' class and 'K' is the number of classes.

Followed by, the class label for 'm^{th}' seed data is obtained by the maximum 'n_{mk}' value using below expression,

$$c_m = \max_{k=1,2,..,K}\{n_{mk}\} \quad (12)$$

From the above equation (12), 'c_m' represent the class label of the 'm^{th}' seed data in input agriculture dataset. By using the above processes, GNF Classifier classifies input seed data as normal or abnormal with higher accuracy.

The algorithmic process of Gaussian Neuro-Fuzzy Classifier is described in below,

Algorithm 2 Gaussian Neuro-Fuzzy Classifier

Algorithm 2 shows the step by step processes of GNF Classifier. By using the above algorithmic steps, GNF Classifier accurately categorizes the each input seed data

// Gaussian Neuro-Fuzzy Classifier Algorithm
Input: Agriculture Dataset 'AD'; Number of Seed Data '$d_i = d_1, d_2, .., d_N$'; Selected Relevant Features
Output: Improved Classification Accuracy
Step 1:Begin
Step 2: Input layer takes 'd_i' as input
Step 3: For each seed data 'd_i'
Step 4: Compute Gaussian membership function '$M_{ij}(d_{mj})$' using (8)
Step 5: Measure firing strength of rule 'i' with respect to 'd_i' to be classified 'ε_{im}'using (9)
Step 6: Calculate weighted output for 'm^{th}' data that belongs to 'k^{th}' class 'β_{mk}' using(10)
Step 7: Normalize value of input seed data that belongs to the 'k^{th}' class 'n_{mk}' using (11)
Step 8: Classify seed data as normal or abnormal using (12)
Step 9: End For
Step 10: End

into normal or abnormal classes with a lower amount of time utilization. This helps for SCFS-GNFC Technique to effectively find the seed quality and growth in agriculture field as compared to existing works.

IV. EXPERIMENTAL SETTINGS

In order to measure the performance of proposed, SCFS-GNFC Technique is implemented in Java Language using agriculture dataset i.e. Soybean Dataset. This dataset is obtained from UCI Machine Learning Repository. The Soybean Dataset contains 35 attributes and 307 instances. From these 35 attributes (i.e. features), SCFS-GNFC Technique selects optimal number of features for finding seed disease through classification. The SCFS-GNFC Technique takes different number of seed data in the range of 30-300 from Soybean Dataset to perform experimental evaluation. The performance of SCFS-GNFC Technique is measured in terms of feature selection time, classification accuracy and error rate and compared with two existing methods [1] and [2].

V. RESULTS AND DISCUSSIONS

In this section, the experimental result of SCFS-GNFC Technique is compared with two state-of-the-art works namely Multilayer perceptron network classifier [1] and Hybrid Kernel based Support Vector Machine (H-SVM) [2]. The efficiency of SCFS-GNFC Technique is determined with the aid of below tables and graphs.

A. Performance Measure of Feature Selection Time

Feature Selection Time 'FST' determines an amount of time required to select the relevant features from an input dataset. The feature selection time is measured using below mathematical expression,

$$FST = n * t(SSF) \quad (13)$$

From equation (13), '$t(SSF)$' represent time utilized for choosing single features as relevant or

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

irrelevant whereas 'n' denotes total number of features considered for experimental process. The feature selection time is evaluated in terms of milliseconds (ms).

Sample Mathematical Calculation for Feature Selection Time

- *Proposed SCFS-GNFC Technique*: Number of features is 5 and the time taken for selecting one feature is 0.9 ms, then the $FST = 5 * 0.9\ ms = 5\ ms$
- *Existing Multilayer perceptron network classifier:* Number of features is 5 and the time needed for choosing one feature is 1.2 ms, then the $FST = 5 * 1.2\ ms = 6\ ms$
- *Existing H-SVM*: Number of features is 5 and the time employed for selecting one feature is 1.5 ms, then the $FST = 30 * 1.5\ ms = 8\ ms$

In order to evaluate the time complexity involved during process of feature selection for seed classification, SCFS-GNFC Technique is implemented in Java language by considering varied number of features in the range of 5-35. When accomplishing the experimental evaluation using 30 features from soybean dataset, SCFS-GNFC Technique attains 14 ms feature selection time whereas conventional works Multilayer perceptron network classifier [1] and H-SVM [2] gets 15 ms, 17 ms respectively. From the above get experimental results, feature selection time using SCFS-GNFC Technique is lower as compared to other state-of-the-art works Multilayer perceptron network classifier [1] and H-SVM [2]. The comparative result analysis of feature selection time is depicted in below Table 1.

TABLE 1 TABULATION FOR FEATURE SELECTION TIME

Number of features (n)	Feature Selection Time (ms)		
	SCFS-GNFC	Multilayer perceptron network classifier	H-SVM
5	5	6	8
10	6	8	10
15	8	11	12
20	10	13	14
25	12	14	15
30	14	15	17
35	15	18	20

Figure 4 Experimental Result of Feature Selection Time versus Number of Features

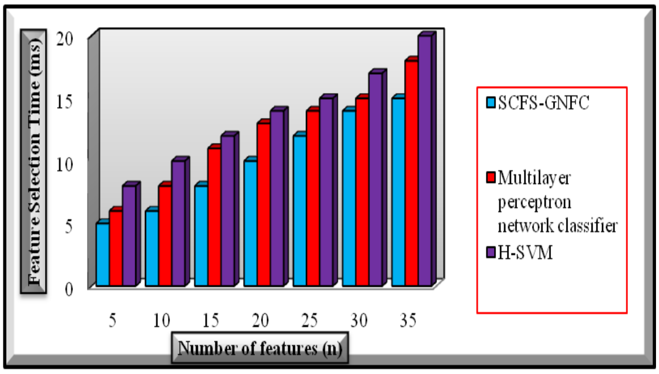

Figure 4 presents the impact of feature selection time with respect to diverse number of features using three methods namely SCFS-GNFC Technique, Multilayer perceptron network classifier [1] and H-SVM [2]. As demonstrated in the above graphical representation, SCFS-GNFC Technique provides minimal feature selection time for efficient seed classification as compared to existing Multilayer perceptron network classifier [1] and H-SVM [2] respectively. This is due to application of Sørensen–Dice Similarity Based Cuckoo Feature Selection (SS-CFA) algorithm in SCFS-GNFC Technique on the contrary to existing algorithm. With application of SS-CFA algorithm, SCFS-GNFC Technique find outs the features with higher rank as current best solutions to significantly classify the seeds with a lower amount of time consumption. This supports for SCFS-GNFC Technique to employ minimal time to choose the relevant features from an input dataset as compared to other conventional works. Hence, SCFS-GNFC Technique reduces the feature selection time by 19 % when compared to existing Multilayer perceptron network classifier [1] and 29 % when compared to existing H-SVM [2].

B. Performance Measure of Classification Accuracy

Classification accuracy estimates the ratio of a number of seed data that are accurately classified as normal or abnormal seed to the total number of seed data. The classification accuracy is mathematically obtained as,

$$CA = \frac{Z_{CC}}{N} * 100 \tag{14}$$

From equation (14), 'Z_{CC}' indicates number of seed data that are exactly classified and 'N' designates a total number of seed data taken for simulation process. The classification accuracy is determined in terms of percentage (%).

Sample Mathematical Calculation for Classification Accuracy

- *Proposed SCFS-GNFC Technique*: Number of seed data is properly classified are 27 and the total number of seed data is 30. Then the CA $= \frac{27}{30} * 100 = 90\%$
- *Existing Multilayer perceptron network classifier*: Number of seed data is correctly classified is 24 and the total number of seed data is 30. Then the CA $= \frac{24}{30} * 100 = 80\ \%$
- *Existing H-SVM*: Number of seed data is accurately classified are 22 and the total number of seed data is 30. Then the CA $= \frac{22}{30} * 100 = 73\%$

For measuring the accuracy of seed classification in agriculture field, SCFS-GNFC Technique is implemented in Java language with different number of seed data in the range of 30-300. When performing the experimental work using 270 seed data from soybean dataset, SCFS-GNFC Technique obtains 96 % classification accuracy whereas state-of-the-art works Multilayer perceptron network classifier [1] and H-SVM [2] acquires 91 %, 85 % respectively. From these experimental results, classification accuracy using SCFS-

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

GNFC Technique is very higher when compared to other conventional works Multilayer perceptron network classifier [1] and H-SVM [2]. The experimental result analysis of classification accuracy is portrayed in below Table 2.

TABLE 2 TABULATION FOR CLASSIFICATION ACCURACY

Number of seed data (N)	Classification accuracy (%)		
	SCFS-GNFC	Multilayer perceptron network classifier	H-SVM
30	90	80	73
60	88	83	75
90	91	87	77
120	93	83	78
150	92	84	75
180	94	85	79
210	91	85	72
240	95	89	73
270	96	91	85
300	97	93	89

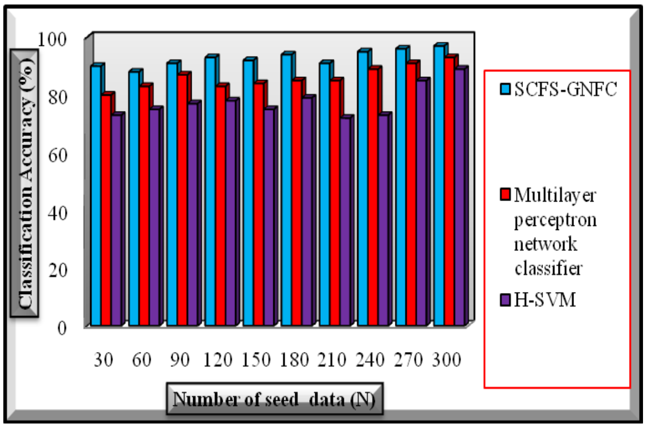

Figure 5 Experimental Result of Classification Accuracy versus Number of Seed Data

Figure 5 shows the impact of classification accuracy based on various numbers of seed data using three methods namely SCFS-GNFC Technique, Multilayer perceptron network classifier [1] and H-SVM [2]. As depicted in the above graphical diagram, SCFS-GNFC Technique provides higher accuracy for seed categorization in agriculture field when compared to existing Multilayer perceptron network classifier [1] and H-SVM [2] respectively. This is because of application of Gaussian Neuro-Fuzzy (GNF) Classifier in SCFS-GNFC Technique on the contrary to conventional algorithms. In GNF Classifier, Gaussian membership function and Neuro-Fuzzy classification are combined in order to enhance the performance of seed categorization. This helps for SCFS-GNFC Technique to increase ratio of a number of seed data that are accurately classified as normal or abnormal when compared to other state-of-the-art works. Therefore, SCFS-GNFC Technique enhances the classification accuracy by 8 % when compared to existing Multilayer perceptron network classifier [1] and 20 % when compared to existing H-SVM [2].

C. Performance Measure of Error Rate

Error rate 'ER' calculates the ratio of a number of seed data that are incorrectly classified to the total number of seed data. The error rate is mathematically measured using below,

$$ER = \frac{Z_{IC}}{N} * 100 \qquad (15)$$

From equation (15) 'Z_{IC}' point outs a number of seed data are inaccurately classified whereas 'N' indicates a total number of seed data. The error rate is measured in terms of percentage (%).

Sample Mathematical Calculation for Error rate

- *Proposed SCFS-GNFC Technique*: Number of seed data is incorrectly classified are 3 and the total number of seed is 30. Then the $ER = \frac{3}{30} * 100 = 10\%$
- *Existing Multilayer perceptron network classifier:* Number of seed data is wrongly classified is 6 and the total number of seed data is 30. Then the $ER = \frac{6}{30} * 100 = 20\%$
- *Existing H-SVM*: Number of seed data is mistakenly classified are 8 and the total number of seed data is 30. Then the $ER = \frac{8}{30} * 100 = 27\%$

To determine the error rate involved during the process of seed classification in agriculture field, SCFS-GNFC Technique is implemented in Java language with help of various number of data in the range of 30-300. When conducting the experimental process using 210 seed data from soybean dataset, SCFS-GNFC Technique gets 9 % error rate whereas existing works Multilayer perceptron network classifier [1] and H-SVM [2] obtains 15 %, 28 % respectively. Accordingly, error rate using SCFS-GNFC Technique is very lower when compared to other conventional works Multilayer perceptron network classifier [1] and H-SVM [2]. The performance result analysis of error rate is demonstrated in below Table 3.

Table 3 Tabulation for Error rate

Number of seed data (N)	Error rate (%)		
	SCFS-GNFC	Multilayer perceptron network classifier	H-SVM
30	10	20	27
60	12	17	25
90	9	13	23
120	7	18	23
150	8	16	25
180	6	15	21
210	9	15	28
240	5	11	27
270	4	9	15
300	3	7	11

Figure 6 Experimental Result of Error Rate versus Number of Seed Data

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

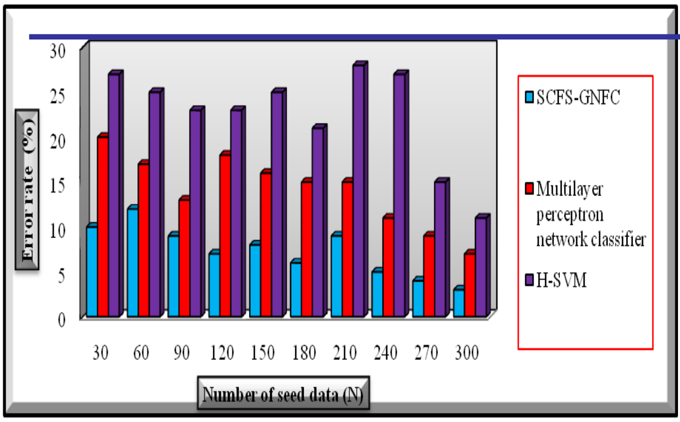

Figure 6 illustrates the impact of error rate with respect to different numbers of seed data using three methods namely SCFS-GNFC Technique, Multilayer perceptron network classifier [1] and H-SVM [2]. As presented in the above graphical figure, SCFS-GNFC Technique provides lower error rate for accurately classify seeds as normal or abnormal in agriculture field as compared to conventional Multilayer perceptron network classifier [1] and H-SVM [2] respectively. This is owing to application of Gaussian Neuro-Fuzzy (GNF) Classifier in SCFS-GNFC Technique on the contrary to state-of-the-art algorithms. With the support of GNF Classifier algorithmic process, SCFS-GNFC Technique correctly classifies each input seed data into normal or abnormal classes by using Gaussian membership function result. This assists for SCFS-GNFC Technique to minimize the ratio of a number of seed data that are wrongly classified when compared to other state-of-the-art works. As a result, SCFS-GNFC Technique reduces the error rate by 49 % when compared to existing Multilayer perceptron network classifier [1] and 68 % when compared to existing H-SVM [2].

VI. CONCLUSION

An effective SCFS-GNFC Technique is designed with the goal of increasing the performance of seed data classification in agriculture field. The goal of SCFS-GNFC Technique is achieved with the application of SS-CFA algorithm and GNF Classifier. The proposed SCFS-GNFC Technique attain higher accuracy and takes lower amount of time for selecting the key features from an input agriculture dataset as compared to state-of-the-art works. Moreover, proposed SCFS-GNFC Technique obtained enhanced performance for categorizing the seed data into corresponding classes (i.e. normal or abnormal) with a minimal amount of time consumption as compared to conventional algorithms. From that, SCFS-GNFC Technique also reduces the ratio of data that are incorrectly classified as normal or abnormal for efficient seeds quality measurement as compared to existing works. The effectiveness of SCFS-GNFC Technique is evaluated in terms of classification accuracy, feature selection time and error rate and compared with two existing works. The experimental result depicts that SCFS-GNFC Technique provides a better performance with an improvement of classification accuracy and minimization of feature selection time for attaining enhanced seed growth in agricultural field when compared to state-of-the-art works.

REFERENCES

[1] Ke-ling TU, Lin-juan LI, Li-ming Yang, Jian-hua Wang, Qun Sun, “Selection For High Quality Pepper Seeds By Machine Vision And Classifiers”, Journal of Integrative Agriculture, Elsevier, Volume 17, Issue 9, Pages 1999-2006, September 2018

[2] K. Aditya Shastry, H.A. Sanjay and G. Deexith, “Quadratic-Radial-Basis-Function-Kernel for classifying multi-class agricultural datasets with continuous attributes”, Applied Soft Computing, Elsevier, Volume 58, Pages 65–74, September 2017

[3] P. Kumar, R. Prasad, V. N. Mishra, D. K. Gupta and S. K. Singh, “Artificial Neural Network for Crop Classification Using C-band RISAT-1 Satellite Datasets”, Russian Agricultural Sciences, Volume 42, Issue 3–4, Pages 281–284, 2016

[4] N. Deepa, K. Ganesan, “Decision-making tool for crop selection for agriculture development”, Neural Computing and Applications, Springer, Volume 31, Issue 4, Pages 1215–1225, April 2019

[5] Kristina Koenig, Bernhard Höfle, Martin Hämmerle, Thomas Jarmer, Bastian Siegmann and Holger Lilienthal, “Comparative classification analysis of post-harvest growth detection from terrestrial LiDAR point clouds in precision agriculture”, ISPRS Journal of Photogrammetry and Remote Sensing, Elsevier, Volume 104, Pages 112–125, 2015

[6] Jharna Majumdar, Sneha Naraseeyappa, Shilpa Ankalaki, “Analysis of Agriculture Data Using Data Mining Techniques: Application of Big Data”, Journal of Big Data, Springer, Volume 4, Issue 20, December 2017

[7] N. Deepa, K. Ganesan, “Hybrid Rough Fuzzy Soft Classifier Based Multi-Class Classification Model for Agriculture Crop Selection”, Soft Computing, Springer, Pages 1–17, November 2018

[8] Wayne Goodridge, Margaret Bernard, René Jordan, Reanne Rampersad, “Intelligent diagnosis of diseases in plants using a hybrid Multi-Criteria decision making technique”, Computers and Electronics in Agriculture, Elsevier, Volume 133, Pages 80–87, 2017

[9] K.D.Yesugade,Aditi Kharde, Ketki Mirashi, Kajal Muley,Hetanshi Chudasama, “ Machine Learning Approach for Crop Selection based on Agro-Climatic Conditions”, International Journal of Advanced Research in Computer and Communication Engineering, Volume 7, Issue 10, Pages 103-106, October 2018

[10] V. Mancini, S. Murolo, G. Romanazzi, “Diagnostic methods for detecting fungal pathogens on vegetable seeds”, Wiley Online Library, Volume 65, Issue 5, Pages 691-703, June 2016

[11] Mohamed Rafik N. Qureshi, Ram Karan Singh, Mohd. Abul Hasan, “Decision support model to select crop pattern for sustainable agricultural practices using fuzzy MCDM”, Environment, Development and Sustainability: A Multidisciplinary Approach to the Theory and Practice of Sustainable Development, Springer, Volume 20, Issue 2, Pages 641-659, 2018

[12] Douglas F.Pereira, Pedro H.Bugatti, Fabricio M.Lopes, André L.S.M.Souza, Priscila T.M.Saito, “Contributing to agriculture by using soybean seed data from the tetrazolium test”, Data in Brief, Elsevier, Volume 23, Pages 1-6, April 2019

[13] Xia Xin, Yinglang Wan, Wenjun Wang, Guangkun Yin, Eric S. McLamore & Xinxiong Lu, “A real-time, non-invasive, micro-optrode technique for detecting seed viability by using oxygen influx” Scientific Reports, Volume 3, Pages 1-6, 2015

[14] Srinivasan SP, Shanthi DS, “A seed yield estimation modelling using classification and regression trees (CART) in the biofuel supply chain”, Journal of Biomedical Imaging and Bioengineering, Volume 1, Issue 1, Pages 15, 2017

[15] Archana Chaudhary, Savita Kolhe and Raj Kamal, “A hybrid ensemble for classification in multiclass datasets: An application to oilseed disease dataset”, Computers and Electronics in Agriculture, Elsevier, Volume 124, Pages 65–72, 2016

[16] Romulo Santelices Moya, Sergio Espinoza Meza, Carlos Magni Díaz, Antonio Cabrera Ariza, Sergio Donoso Calderon and Karen Peña-Rojas, “Variability in seed germination and seedling growth at the intra- and interprovenance levels of Nothofagus glauca (Lophozonia glauca), an endemic species of Central Chile”, New Zealand Journal of Forestry Science, Volume 47, Issue 10, Pages 1-9, 2017

[17] Tekalign TujoG, Dileep Kumar G, Elifenesh YitagesuD, Meseret GirmaB, "A Predictive Model to Predict Seed Classes using Machine Learning", International Journal of Engineering Research & Technology (IJERT), Volume 6, Issue 08, Pages 334-344, August - 2017

[18] Zhan Zhao,Mingzhi Jin, Chunjie Tian, Simon X.Yang, "Prediction of seed distribution in rectangular vibrating tray using grey model and artificial neural network", Biosystems Engineering, Elsevier, Volume 175, Pages 194-205, November 2018

[19] S.Bhuvaneswari, T.Pramananda Perumal, B.Jagadhesan, "An analysis and impact factors on Agriculture field using Data Mining Techniques", International Journal of Business Intelligents, Volume 05, Issue 01, Pages 41-44, June 2016

[20] Lei Shi, Xinming Ma, Qiguo Duan and Mei Weng, "Agricultural Data Classification Based on Rough Set and Decision Tree Ensemble", Sensor Letters, Volume 10, Issue 1, Pages 271-278, 2012