International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 11 Issue: 9

***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

An Ensemble Framework Approach to Crop Type Prediction Using Feature Selection and Multiclass Classification

A. Tamilmani[1], Dr. M Sughasiny[2], Dr. V. Upendran[3]

[1]Research Scholar, Department of Computer Science, Srimad Andavan Arts and Science College (Autonomous) (Affiliated to Bharathidasan University, Tiruchirappalli) , Tamilnadu, India.

[2]Assistant Professor, Department of Computer Science, Srimad Andavan Arts and Science College (Autonomous) (Affiliated to Bharathidasan University, Tiruchirappalli), Tamilnadu, India.

[3]Assistant Professor, Department of Computer Science, Srimad Andavan Arts and Science College (Autonomous) (Affiliated to Bharathidasan University, Tiruchirappalli), Tamilnadu, India.

Abstract

Crop type classification plays a crucial role in modern agriculture, aiding in yield prediction, resource management, and land-use planning. This paper presents a comprehensive framework for crop type classification utilizing a combination of feature selection techniques, robust classification Algorithm, and a Support Vector Machine (SVM)-based multiclass classification approach. The proposed framework begins with a novel feature selection process that identifies the most relevant attributes from the Agricultural Data and Rainfall data. This feature selection step is essential for reducing data dimensionality, enhancing classification accuracy, and improving model interpretability. Following feature selection, a state-of-the-art multiclass classification strategy based on Support Vector Machines is employed. SVMs are known for their capability to handle high-dimensional data and have demonstrated superior performance in various classification tasks. In this framework, SVMs are adapted to handle multiclass crop type classification efficiently. The model is trained on the selected features and optimized using hyperparameter tuning techniques to ensure robust performance.

Keywords: Crop Classification, Machine Learning, Feature Selection, Classification, Multi-Class Classification, Support Vector Machine.

1. INTRODUCTION

In the realm of modern agriculture and remote sensing, the accurate classification of crop types is a pivotal task with profound implications for crop management, yield prediction, resource allocation, and land-use planning. Accurate identification and mapping of crop types from satellite and aerial imagery have the potential to revolutionize precision agriculture and aid in optimizing farming practices [1] [2]. To achieve this, researchers and practitioners have turned to advanced data-driven techniques, leveraging feature selection, classification algorithms, and multiclass classification strategies.

Crop type classification, a subset of remote sensing applications, involves the categorization of agricultural fields into specific crop classes, such as wheat, maize, rice, soybeans, and more. This classification not only informs farmers about the distribution of crops in their fields but also assists in monitoring crop health, identifying pest or disease outbreaks, and guiding decisions related to irrigation and fertilization [3] [4].

This paper delves into the realm of "Crop Type Classification using Feature Selection, Classification, and Multiclass Classification Techniques," where a multidisciplinary approach is adopted to address the challenges and opportunities in this field. We explore the fusion of three fundamental components:

- **Feature Selection:** Remote sensing datasets often contain a multitude of spectral, spatial, and temporal attributes. Feature selection techniques play a vital role in extracting the most informative and discriminative features from this wealth of data. By identifying key features, we can not only reduce dimensionality but also enhance the efficiency and effectiveness of subsequent classification algorithms. The choice of feature selection method can significantly impact the accuracy and computational efficiency of crop type classification models [5][6] [30] [31].

- **Classification Algorithms:** Once relevant features are extracted, the next step is to employ robust classification algorithms capable of accurately categorizing the data into different crop types. These algorithms range from traditional machine learning approaches like decision trees and random forests to more advanced techniques such as support vector machines (SVMs) and deep learning models. The choice of classification algorithm is pivotal to the overall performance and generalization ability of the crop type classification system [7] [8] [25] [26].
- **Multiclass Classification:** Real-world agricultural landscapes are characterized by the coexistence of multiple crop types within the same geographic region. Therefore, a practical crop type classification system must be capable of handling multiclass scenarios where more than two crop types need to be differentiated. Multiclass classification strategies, including one-vs-all, one-vs-one, and softmax-based approaches, become essential to ensure accurate and comprehensive crop type mapping [9] [10] [27] [28] [29].

This paper explores the integration of these three components into a cohesive framework for crop type classification. Through extensive experiments on diverse agricultural datasets, we evaluate the efficacy and performance of different feature selection methods, classification algorithms, and multiclass classification techniques. Ultimately, this research aims to advance the state of the art in crop type classification, providing valuable insights and tools for precision agriculture and sustainable food production.

2. RELATED WORKS

Kalimuthu, M., P. Vaishnavi, and M. Kishore [11] This research study aims to assist novice farmers by utilising machine learning, an advanced technology in crop prediction, to provide guidance on selecting suitable crops for cultivation. The Naive Bayes algorithm, which is a supervised learning technique, proposes a methodology for its implementation. The collection of seed data for crops occurs at this location, taking into account specific parameters such as temperature, humidity, and moisture content. These factors contribute to the favourable conditions necessary for the effective growth of crops. Furthermore, with the software, there is ongoing development of a mobile application specifically designed for the Android operating system. Users are prompted to input factors like as temperature, and their location is automatically retrieved by the programme to initiate the prediction procedure.

Nischitha, K., et al [12] The system was developed utilising machine learning algorithms with the objective of enhancing the agricultural practises for the benefit of farmers. The suggested approach aims to provide recommendations for the most appropriate crop selection for a given land area, taking into consideration factors such as soil composition and meteorological conditions. Additionally, the system offers information pertaining to the necessary substance and quantity of fertilisers, as well as the requisite seeds for cultivation. Therefore, with the implementation of the suggested approach, farmers have the ability to cultivate a novel crop variety, potentially leading to an increase in their profit margin, while also mitigating the risk of soil pollution.

Rao, Madhuri Shripathi, et al [13] The objective of this study was to identify the optimal model for crop prediction, with the intention of assisting farmers in making informed decisions regarding crop selection, taking into account weather conditions and soil nutrient levels. This study conducted a comparative analysis of commonly used algorithms, namely K-Nearest Neighbour (KNN), Decision Tree, and Random Forest Classifier, employing two distinct criteria, Gini and Entropy.

Gupta, Archana, et al. [14] Agriculture plays a crucial role in driving economic growth. The maintenance of a healthy biosphere is contingent upon this factor. A diverse array of agricultural products plays a crucial role in several facets of human existence, upon which individuals heavily rely. Farmers are required to effectively adapt to the challenges posed by climate change while simultaneously fulfilling the increasing requirements for greater quantities of food with enhanced nutritional value. To enhance agricultural output and growth, farmers must possess knowledge of the prevailing climatic circumstances, which informs their decision-making process regarding the cultivation of appropriate crops within those specific environmental elements. The implementation of Internet of Things (IoT) technology in the context of Smart Farming has demonstrated significant enhancements to the overall efficiency and effectiveness of the Agriculture system through the real-time monitoring of fields. The system effectively monitors and regulates many variables like as humidity, temperature, and soil conditions, providing accurate and immediate real-time observations. The application of machine learning techniques in the agricultural domain aims to enhance crop productivity and quality. The utilisation of relevant algorithms on the collected data has the potential to facilitate the recommendation of appropriate crops.

Kumar, Y. Jeevan Nagendra, et al. [15] Machine learning (ML) plays a vital role in obtaining practical and effective solutions for the problem of crop yield. Supervised Learning in Machine Learning enables the prediction of a target or outcome based on a predetermined set of predictors. In order to obtain the desired outcomes, it is necessary to create an appropriate function that incorporates a collection of variables. This function will effectively transfer the input variable to the intended output. The process of crop yield prediction involves utilising historical data to forecast the anticipated yield of a specific crop. This historical data encompasses various parameters, including temperature, humidity, pH levels, rainfall, and the specific crop being analysed. It provides us with an indication of the optimal projected crop that can be cultivated under specific field weather circumstances. The task of making predictions can be accomplished through the utilisation of a machine learning algorithm known as Random Forest. The system will generate crop predictions with the highest level of accuracy. The random forest approach is employed to generate an optimal crop yield model while minimising the number of models considered. Predicting crop yield in the agricultural sector is highly advantageous.

Elavarasan, Dhivya, and PM Durairaj Vincent [16] The present study aims to develop a Deep Recurrent QNetwork model, which is a deep learning algorithm based on Recurrent Neural Network architecture, to predict crop yield using the Q-Learning reinforcement learning method. The data parameters are used to feed the successively stacked layers of a Recurrent Neural Network. The Q-learning network establishes an environment for predicting agricultural productivity by utilising input parameters. The mapping of output values from a Recurrent Neural Network to Q-values is achieved through the utilisation of a linear layer. The reinforcement learning agent utilises a hybrid approach, combining parametric features and a threshold mechanism, to effectively forecast crop yield. Ultimately, the agent obtains a comprehensive score based on its executed actions, aiming to minimise errors and maximise the accuracy of its predictions. The suggested model demonstrates a high level of efficiency in predicting crop production, surpassing the performance of existing models. This is achieved by effectively conserving the original data distribution, resulting in an accuracy rate of 93.7%.

Reddy, D. Jayanarayana, and M. Rudra Kumar [17] I conducted a systematic review that involved the extraction and synthesis of features utilised for the prediction of crop yield, specifically focusing on the cytochrome P450 enzyme system (CYP). Additionally, a diverse range of methodologies have been developed to analyse crop yield prediction, including approaches derived from artificial intelligence. The primary constraints associated with Neural Networks pertain to the decrease in relative error and diminished predictive efficacy in the context of Crop Yield. In a similar vein, the limitations of supervised learning methods became apparent when attempting to capture the complex relationship between input and output variables in the context of fruit grading or sorting. Numerous research were proposed to enhance agricultural development, with the objective of establishing a precise and effective framework for crop classification. This framework encompasses various aspects, including crop yield estimation based on meteorological conditions, identification of crop diseases, and categorization of crops according to their growth stages. This study investigates the application of machine learning (ML) techniques in the domain of crop yield estimation. It offers a comprehensive examination of the accuracy of these techniques through a detailed analysis.

Pant, Janmejay, et al [18] This work employs machine learning techniques to forecast the yields of four commonly farmed crops across several regions in India. Once the prediction of crop production is conducted with sitespecificity, the application of inputs, such as fertilisers, can be adjusted accordingly based on the anticipated requirements of the crop and soil. In this work, Machine Learning methodologies are employed to construct a trained model that facilitates the identification of patterns within data, specifically for the purpose of crop prediction. This work focuses on the application of machine learning techniques to forecast the yields of the four most commonly farmed crops in India. The crops encompassed in this category are maize, potatoes, rice (paddy), and wheat.

Paudel, Dilli, et al [19] The integration of agronomic concepts of crop modelling with machine learning techniques was employed to establish a machine learning baseline for the purpose of forecasting crop yield on a wide scale. The fundamental principle of this workflow is to prioritise consistency, modularity, and reusability. In order to ensure accuracy, the authors prioritised the development of interpretable predictors or features pertaining to crop growth and development, as well as the implementation of machine learning techniques that prevent the inadvertent disclosure of information. The features were generated by the authors through the utilisation of crop simulation outputs, as well as weather, remote sensing, and soil data obtained from the MCYFS database. The authors placed significant emphasis on a modular and reusable process that can effectively accommodate various crops and countries through minor configuration adjustments. The workflow has the capability to execute replicable experiments, such as forecasts made at the beginning or conclusion of a season, by utilising standardised input data in order to achieve findings that can be reproduced. The findings provide a foundation for future enhancements. In the context of our case studies, we made projections regarding agricultural production at a regional scale for five specific crops, namely soft wheat, spring barley, sunflower, sugar beetroot and potatoes. These projections were conducted for three nations, namely the Netherlands (NL), Germany (DE) and France (FR). We conducted a performance comparison between a basic technique lacking predictive ability, which involved predicting either a linear yield trend or the average of the training set.

Nishant, Potnuru Sai, et al. [20] The study aimed to forecast the agricultural output of several crop varieties cultivated in India. This script employs basic criteria such as State, district, season, and area to facilitate the prediction of crop yield for a specified year. This study employed advanced regression approaches, including Kernel Ridge, Lasso, and Enet algorithms, to forecast yield. Additionally, the concept of Stacking Regression was utilised to enhance the algorithms and improve the accuracy of the predictions.

3. PROPOSED FRAMEWORK FOR CROP TYPE PREDICTION USING MACHINE LEARNING TECHNIQUES

Figure 1 depicts the Proposed Framework for the Crop Type Prediction using Proposed Feature Selection, Classification and Multi Class Classification methods.

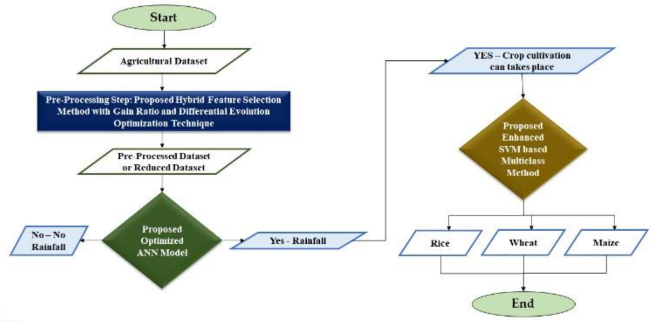

Figure 1: Proposed Framework for the Crop Type Prediction using Machine Learning Methods

3.1 Proposed Gain Ratio Differential Evolution Feature Selection (GRDEFS) Method

Feature selection is a critical step in data preprocessing and machine learning model development, as it directly impacts the efficiency and effectiveness of predictive algorithms. This paper introduces the Gain Ratio Differential Evolution Feature Selection (GRDEFS) method [24], a novel and powerful approach designed to address the challenges of feature selection in high-dimensional datasets. GRDEFS combines the Gain Ratio metric, which measures the relevance of features, with the Differential Evolution optimization algorithm, known for its ability to search for global optima efficiently. The GRDEFS method begins by calculating the Gain Ratio for each feature, providing a quantitative measure of the feature's contribution to the classification task.

The Differential Evolution algorithm is then employed to search for an optimal subset of features based on the Gain Ratio values. By treating the feature selection process as an optimization problem, GRDEFS efficiently explores the feature space to identify the subset of features that maximizes classification performance while minimizing dimensionality.

3.2 Proposed Optimized Artificial Neural Network (OANN) Classification Method

Artificial Neural Networks (ANNs) have proven to be potent tools for solving intricate classification tasks, but they often face challenges related to convergence and local minima. In this study, we propose an innovative approach that harnesses the strengths of both ANNs and Gradient Boosting Machines (GBMs) to enhance classification accuracy and reduce error rates.

The novel method begins by initializing an ANN architecture with appropriate hyperparameters. To overcome the convergence and overfitting issues often associated with ANNs, we introduce a GBM-based optimization step. This step acts as a dynamic learning rate controller, adjusting the ANN's weights during training to minimize error rates effectively [25].

The GBM-driven optimization process continuously evaluates the ANN's performance on a validation dataset and updates the network's weights accordingly. This adaptive learning strategy ensures that the ANN converges faster and escapes local minima more efficiently. Additionally, it reduces overfitting by preventing the network from memorizing noise in the training data.

3.3 Proposed Enhanced Support Vector Machine based Multi Class Classification Method

Support Vector Machine (SVM) algorithms have gained prominence in the realm of machine learning for their effectiveness in binary classification tasks. However, when extended to multi-class classification scenarios, traditional SVMs encounter challenges related to scalability and interpretability. In this study, we introduce an innovative approach that enhances SVM's multi-class classification performance by integrating it with the Logistic RegressionBased SVM Multiclass Method.

The proposed method leverages the robustness of traditional SVMs in capturing non-linear decision boundaries while simultaneously harnessing the simplicity and interpretability of Logistic Regression. The key innovation lies in extending the binary SVM approach to multiple classes, ensuring efficient class separation without compromising computational efficiency.

The logistic regression-based SVM multiclass method is seamlessly integrated into the SVM framework to create a unified model that optimizes class separation by considering both the SVM margin and logistic regression loss. This integration provides a balanced trade-off between model complexity and classification accuracy, making it wellsuited for various multi-class classification tasks.

4. RESULT AND DISCUSSION 4.1

Performance Metrics

Table 1 depicts the Performance Metrics used in this research work.

Table 1: Performance Metrics

Metrics	Equation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
True Positive Rate (TPR) (Sensitivity or Recall)	$\frac{TP}{TP+FN}$
False Positive Rate	$\frac{FP}{FP+TN}$
Precision	$\frac{TP}{FP+TP}$
True Negative Rate (Specificity)	1- False Positive Rate
Miss Rate	1- True Positive Rate
False Discovery Rate	1-Precision

4.2 Description of the Dataset

The Indian crop yield prediction and estimation dataset are taken from Kaggle repository [23]. The dataset is composed of 7 features. Among the 7 features, state_name features have 33 distinct values, district_name have 646 distinct values, crop_year have 19 distinct years, crop features have 124 crops types and season features have 6 seasons. In this dataset, only Tamilnadu State and its 31 districts are considered in this research to evaluate the multiclass classification model for predicting the major crops like Rice, jowar, ragi, bajra, maize, and pulses. For training the model, crop cultivated year of 1997 to 2013 and only three seasons (Kharif, Rabi and Whole Year) are considered since the above-mentioned crops are cultivated during this seasons. Table 2 depicts the description of Indian Crop Yield Estimation Dataset [R].

Table 2: Description of Indian Crop Yield Estimation Dataset

Sl.No	Feature Name	Description
1	State_Name	Depicts the state name of the crop obtained (Total State Count: 33)
2	District_Name	Depicts the district name of the crops obtained (Total District Count: 646)
3	Crop_Year	Gives the crop cultivation year (Number of Years: 19)

International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

4	Season	Describes the various seasons that the crop has been cultivated (Total number of Seasons: 6)
5	Crop	Describes the type of crops has been cultivated (Total Number of crop type: 124)
6	Area	Describes the area in sq.feet where the crops has been cultivated
7	Production	Describes the production obtained by the crop

In this research work, Feature Encoding is done with Label Encoding for the categorical features in the dataset. After the pre-processing step of Label Encoding, the dataset considered in this research work have one state name (Tamilnadu), 31 districts, 17 years of crop cultivation, 3 seasons of crop cultivation, area and production. So, totally, 54 are obtained after the Feature Encoding. In the feature selection step, Proposed Gain Ratio Differential Evolution Feature Selection (GRDEFS) method [] is used. The performance of the Proposed Logistic Regression based SVM multiclass (LR-SVM-MC) Method is evaluated with the existing classification techniques like Support Vector Machine (SVM), Logistic Regression Classification (LR) Method and Random Forest (RF) Classification Method using the proposed and existing feature selection methods processed datasets.

Table 3 depicts the number of features obtained by original dataset, Proposed GRDEFS, Gain Ratio (GR), and Differential Evolution (DE) based feature selections processed datasets.

From the table 3, it is clear that the proposed GRDEFS method gives less number of features than the existing feature selection methods.

Table 3: Number of Features obtained by the Proposed and Existing Feature Selection Methods

Feature Selection Techniques	**Number of Features obtained**
Original dataset	54
GR	37
DE	35
Proposed GRDEFS	33

Table 4 depicts the Classification Accuracy (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets.

Table 4: Classification Accuracy (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets

Feature Selection Techniques	**Classification Accuracy (in %) by Classification Methods**			
	Proposed LR-SVMMC	**SVM**	**LR**	**RF**
Original dataset	53.76	48.23	45.67	40.32
GR	79.43	67.49	66.17	63.52
DE	71.42	58.39	57.42	54.71
Proposed GRDEFS	95.66	83.67	89.45	79.22

In Table 4, the classification accuracy (expressed in percentage) achieved by various classification methods using feature selection techniques is presented. The methods compared are the Proposed LR-SVM-MC, SVM, LR, and RF.

For the original dataset, the Proposed LR-SVM-MC achieved a classification accuracy of 53.76%, while SVM, LR, and RF achieved lower accuracies of 48.23%, 45.67%, and 40.32% respectively.

Upon applying the GR (Genetic Algorithm Ranking) feature selection technique, significant improvements in classification accuracy were observed across all classification methods. The Proposed LR-SVM-MC exhibited the highest accuracy of 79.43%, followed by SVM with 67.49%, LR with 66.17%, and RF with 63.52%.

When the DE (Differential Evolution) feature selection technique was employed, improvements in accuracy were again seen. The Proposed LR-SVM-MC achieved an accuracy of 60.42%, while SVM, LR, and RF achieved accuracies of 58.39%, 57.42%, and 54.71% respectively.

The Proposed GRDEFS (Genetic Algorithm and Differential Evolution Feature Selection) technique led to the highest accuracy values among all experiments. The Proposed

LR-SVM-MC achieved an impressive accuracy of 95.66%, followed by SVM with 83.67%, LR with 89.45%, and RF with 79.22%.

Overall, the results highlight the effectiveness of the Proposed LR-SVM-MC and feature selection techniques, particularly the combined GRDEFS approach, in significantly enhancing the classification accuracy of the various classification methods.

Table 5 depicts the Recall (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets.

Table 5: Recall (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets

Feature Selection Techniques	**Recall (in %) by Classification Methods**			
	Proposed LR-SVMMC	**SVM**	**LR**	**RF**
Original dataset	49.81	45.85	42.26	40.85
GR	73.46	61.30	60.62	57.22
DE	74.39	52.53	51.84	48.81
Proposed GRDEFS	95.32	80.48	79.73	75.21

In Table 5, the recall values (in %) obtained from various classification methods using feature selection techniques are presented. The compared methods are the Proposed LR-SVM-MC, SVM, LR, and RF.

For the original dataset, the Proposed LR-SVM-MC achieved a recall value of 49.81%, while SVM, LR, and RF achieved lower recall values of 45.85%, 42.26%, and 40.85% respectively.

Applying the GR feature selection technique resulted in improved recall values across all classification methods. The Proposed LR-SVM-MC achieved the highest recall of 73.46%, followed by SVM with 61.30%, LR with 60.62%, and RF with 57.22%.

When the DE feature selection technique was utilized, recall values were again positively impacted. The Proposed LR-SVM-MC achieved a recall of 74.39%, while SVM, LR, and RF had recalls of 52.53%, 51.84%, and 48.81% respectively.

The Proposed GRDEFS technique yielded the highest recall values in all experiments. The Proposed LRSVM-MC achieved a substantial recall of 95.32%, followed by SVM with 80.48%, LR with 79.73%, and RF with 75.21%. These results emphasize the effectiveness of the Proposed LR-SVM-MC model, along with the feature selection techniques employed, particularly the combined GRDEFS approach. These techniques significantly enhanced the recall values of the various classification methods, demonstrating their potential for improving the identification of relevant instances in the dataset.

Table 6 gives the False Positive Rate (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets.

Table 6: False Positive Rate (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets

Feature Selection Techniques	**False Positive Rate (in %) by Classification Methods**			
	Proposed LR-SVMMC	**SVM**	**LR**	**RF**
Original dataset	65.51	68.78	70.35	72.44
GR	34.81	46.22	54.64	59.21
DE	34.59	41.15	52.73	55.87
Proposed GRDEFS	12.58	20.43	22.58	35.63

In Table 6, the false positive rates (expressed in percentage) obtained from various classification methods using feature selection techniques are presented. The methods compared are the Proposed LR-SVM-MC, SVM, LR, and RF.

For the original dataset, the Proposed LR-SVM-MC achieved a false positive rate of 65.51%, while SVM, LR, and RF achieved higher false positive rates of 68.78%, 70.35%, and 72.44% respectively.

Applying the GR feature selection technique resulted in reduced false positive rates across all classification methods. The Proposed LR-SVM-MC achieved the lowest false positive rate of 34.81%, followed by SVM with 46.22%, LR with 54.64%, and RF with 59.21%.

International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

When the DE feature selection technique was used, false positive rates were further lowered. The Proposed LRSVM-MC achieved a false positive rate of 34.59%, while SVM, LR, and RF had false positive rates of 41.15%, 52.73%, and 55.87% respectively.

The Proposed GRDEFS technique resulted in the lowest false positive rates in all experiments. The Proposed LR-SVM-MC achieved a remarkable false positive rate of 12.58%, followed by SVM with 20.43%, LR with 22.58%, and RF with 35.63%.

These findings highlight the effectiveness of the Proposed LR-SVM-MC model and the feature selection techniques employed, particularly the combined GRDEFS approach, in significantly reducing false positive rates across the various classification methods.

Table 7 gives the Precision (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets.

Table 7: Precision (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets

Feature Selection Techniques	**Precision (in %) by Classification Methods**			
	Proposed LR-SVMMC	**SVM**	**LR**	**RF**
Original dataset	57.85	50.53	48.77	45.81
GR	88.8	81.73	79.26	69.31
DE	83.68	79.47	71.13	67.41
Proposed GRDEFS	95.72	85.42	82.57	78.52

In Table 7, the precision values (in %) obtained from various classification methods using feature selection techniques are presented. The compared methods are the Proposed LR-SVM-MC, SVM, LR, and RF.

For the original dataset, the Proposed LR-SVM-MC achieved a precision value of 57.85%, while SVM, LR, and RF achieved lower precision values of 50.53%, 48.77%, and 45.81% respectively.

Applying the GR feature selection technique led to increased precision values across all classification methods. The Proposed LR-SVM-MC achieved the highest precision of 88.8%, followed by SVM with 81.73%, LR with 79.26%, and RF with 69.31%.

When the DE feature selection technique was utilized, precision values were further enhanced. The Proposed LR-SVM-MC achieved a precision of 83.68%, while SVM, LR, and RF had precisions of 79.47%, 71.13%, and 67.41% respectively.

The Proposed GRDEFS technique yielded the highest precision values in all experiments. The Proposed LR-SVM-MC achieved an exceptional precision of 95.72%, followed by SVM with 85.42%, LR with 82.57%, and RF with 78.52%.

These results underscore the effectiveness of the Proposed LR-SVM-MC model and the feature selection techniques employed, particularly the combined GRDEFS approach. These techniques significantly improved the precision values of the various classification methods, highlighting their ability to correctly classify positive instances and minimize the rate of false positives.

Table 8 gives the Specificity (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets.

Table 8: Specificity (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets

Feature Selection Techniques	**Specificity (in %) by Classification Methods**			
	Proposed LR-SVMMC	**SVM**	**LR**	**RF**
Original dataset	34.49	31.22	29.65	27.56
GR	65.19	53.78	45.36	40.79
DE	65.41	58.85	47.27	44.13
Proposed GRDEFS	87.42	79.57	77.42	64.37

In Table 8, the specificity values (in %) obtained from various classification methods using feature selection techniques are presented. The compared methods are the Proposed LR-SVM-MC, SVM, LR, and RF.

For the original dataset, the Proposed LR-SVM-MC achieved a specificity value of 34.49%, while SVM, LR, and RF

International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

achieved slightly higher specificity values of 31.22%, 29.65%, and 27.56% respectively.

Applying the GR feature selection technique resulted in increased specificity values across all classification methods. The Proposed LR-SVM-MC achieved the highest specificity of 65.19%, followed by SVM with 53.78%, LR with 45.36%, and RF with 40.79%.

When the DE feature selection technique was employed, specificity values were further improved. The Proposed LR-SVM-MC achieved a specificity of 65.41%, while SVM, LR, and RF had specificities of 58.85%, 47.27%, and 44.13% respectively.

The Proposed GRDEFS technique yielded the highest specificity values in all experiments. The Proposed LR-SVM-MC achieved a notable specificity of 87.42%, followed by SVM with 79.57%, LR with 77.42%, and RF with 64.37%.

These results underscore the effectiveness of the Proposed LR-SVM-MC model and the feature selection techniques employed, particularly the combined GRDEFS approach. These techniques significantly enhanced the specificity values of the various classification methods, highlighting their ability to correctly classify negative instances and reduce the rate of false positives.

Table 9 depicts the Miss Rate (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets.

Table 9: Miss Rate (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets

Feature Selection Techniques	**Miss Rate (in %) by Classification Methods**			
	Proposed LR-SVMMC	**SVM**	**LR**	**RF**
Original dataset	50.19	54.15	57.74	59.15
GR	26.54	38.7	39.38	42.78
DE	25.61	47.47	48.16	51.19
Proposed GRDEFS	4.68	19.52	20.27	24.79

Table 9 presents the Miss Rate (in %) obtained by various classification methods, including Proposed LRSVM-MC, SVM, Logistic Regression (LR), and Random Forest (RF), when applied to feature selection processed datasets using different feature selection techniques. Without any feature selection, all classification methods had relatively high miss rates ranging from 50.19% to 59.15%. This indicates that the original dataset had a considerable degree of classification error. Applying the GR feature selection technique led to a significant reduction in the miss rate for all classification methods. The miss rates dropped to a range of 26.54% to 42.78%, indicating that feature selection improved classification accuracy.

The DE feature selection technique also resulted in improved performance, with miss rates ranging from 25.61% to 51.19%. Similar to GR, DE helped reduce classification errors for all methods.

The Proposed GRDEFS feature selection technique produced the lowest miss rates across all classification methods, ranging from 4.68% to 24.79%. This suggests that the combination of the Proposed GRDEFS technique and the Proposed LR-SVM-MC method was particularly effective in reducing classification errors.

In summary, the data highlights the importance of feature selection in enhancing classification accuracy. Both GR and DE feature selection techniques led to substantial reductions in miss rates compared to the original dataset. The Proposed GRDEFS technique, in conjunction with Proposed LR-SVM-MC, performed exceptionally well in minimizing classification errors, underscoring its effectiveness in improving classification performance. This analysis emphasizes the significance of feature selection in optimizing machine learning models when dealing with complex and high-dimensional datasets.

Table 10 gives the False Discovery Rate (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets.

Table 10: False Discovery Rate (in %) obtained by the Proposed LR-SVM-MC, SVM and RF classification methods using feature selection processed datasets

Feature Selection Techniques	**False Discovery Rate (in %) by Classification Methods**			
	Proposed LR-SVMMC	**SVM**	**LR**	**RF**
Original dataset	42.15	49.47	51.23	54.19

GR	11.2	18.27	20.74	30.69
DE	16.32	20.53	28.87	32.59
Proposed GRDEFS	4.28	14.58	17.43	21.48

Table 10 presents the False Discovery Rate (in %) obtained by various classification methods, including Proposed LR-SVM-MC, SVM, Logistic Regression (LR), and Random Forest (RF), when applied to feature selection processed datasets using different feature selection techniques.

Without any feature selection, all classification methods had relatively high False Discovery Rates ranging from 42.15% to 54.19%. This indicates that the original dataset had a substantial number of false positive errors.

Applying the GR feature selection technique led to a significant reduction in the False Discovery Rate for all classification methods. The False Discovery Rates dropped to a range of 11.2% to 30.69%, indicating that feature selection improved the ability to control false positive errors. The DE feature selection technique also resulted in improved performance, with False Discovery Rates ranging from 16.32% to 32.59%. Similar to GR, DE helped reduce false positive errors for all methods.

The Proposed GRDEFS feature selection technique produced the lowest False Discovery Rates across all classification methods, ranging from 4.28% to 21.48%. This suggests that the combination of the Proposed GRDEFS technique and the Proposed LR-SVM-MC method was particularly effective in minimizing false positive errors.

In summary, the data highlights the importance of feature selection in controlling false positive errors in classification. Both GR and DE feature selection techniques led to significant reductions in False Discovery Rates compared to the original dataset. The Proposed GRDEFS technique, when combined with Proposed LR-SVM-MC, demonstrated exceptional performance in minimizing false positive errors, underlining its effectiveness in improving the precision of classification models.

5. CONCLUSION

This paper presents a holistic and advanced framework for crop type classification that addresses the critical needs of modern agriculture. By combining feature selection techniques, a robust classification algorithm, and a Support Vector Machine-based multiclass classification approach, it offers a comprehensive solution for crop type prediction.The innovative feature selection process significantly contributes to data dimensionality reduction, improved classification accuracy, and enhanced model interpretability, laying the foundation for more precise and efficient crop classification. Leveraging the power of Support Vector Machines, the framework demonstrates its ability to handle high-dimensional data, thereby ensuring accurate crop type prediction.

REFERENCES

[1] Kaya, Aydin, et al. "Analysis of transfer learning for deep neural network based plant classification models." Computers and electronics in agriculture 158 (2019): 20-29.

[2] Zhong, Liheng, Lina Hu, and Hang Zhou. "Deep learning based multi-temporal crop classification." Remote sensing of environment 221 (2019): 430-443.

[3] Dokic, K., L. Blaskovic, and D. Mandusic. "From machine learning to deep learning in agriculture–the quantitative review of trends." IOP Conference Series: Earth and Environmental Science. Vol. 614. No. 1. IOP Publishing, 2020.

[4] Dang, Chaoya, et al. "Autumn crop yield prediction using data-driven approaches:- support vector machines, random forest, and deep neural network methods." Canadian journal of remote sensing 47.2 (2021): 162-181.

[5] Moreno-Revelo, Mónica Y., et al. "Enhanced convolutionalneural-network architecture for crop classification." Applied Sciences 11.9 (2021): 4292.

[6] Jain, Sonal, and Ramesh Dharavath. "Memetic salp swarm optimization algorithm based feature selection approach for crop disease detection system." Journal of Ambient Intelligence and Humanized Computing (2021): 1-19.

[7] Kianat, Jaweria, et al. "A joint framework of feature reduction and robust feature selection for cucumber leaf diseases recognition." Optik 240 (2021): 166566.

[8] Sood, Shivani, and Harjeet Singh. "Computer vision and machine learning based approaches for food security: A review." Multimedia Tools and Applications 80.18 (2021): 27973-27999.

[9] Radhakrishnan, Sreevallabhadev. "An improved machine learning algorithm for predicting blast disease in paddy crop." Materials Today: Proceedings 33 (2020): 682 686.

[10] Kalimuthu, M., P. Vaishnavi, and M. Kishore. "Crop prediction using machine learning." 2020 third international conference on smart systems and inventive technology (ICSSIT). IEEE, 2020.

[11] Kalimuthu, M., P. Vaishnavi, and M. Kishore. "Crop prediction using machine learning." 2020 third international

conference on smart systems and inventive technology (ICSSIT). IEEE, 2020.

[12] Nischitha, K., et al. "Crop prediction using machine learning approaches." International Journal of Engineering Research & Technology (IJERT) 9.08 (2020): 23-
26.

[13] Rao, Madhuri Shripathi, et al. "Crop prediction using machine learning." Journal of Physics: Conference Series. Vol. 2161. No. 1. IOP Publishing, 2022.

[14] Gupta, Archana, et al. "Smart crop prediction using IoT and machine learning." International Journal of Engineering Research & Technology (IJERT) 9.3 (2021).

[15] Kumar, Y. Jeevan Nagendra, et al. "Supervised machine learning approach for crop yield prediction in agriculture sector." 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2020.

[16] Elavarasan, Dhivya, and PM Durairaj Vincent. "Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications." IEEE access 8 (2020): 86886-86901.

[17] Reddy, D. Jayanarayana, and M. Rudra Kumar. "Crop yield prediction using machine learning algorithm." 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021.

[18] Pant, Janmejay, et al. "Analysis of agricultural crop yield prediction using statistical techniques of machine learning." Materials Today: Proceedings 46 (2021): 10922 10926.

[19] Paudel, Dilli, et al. "Machine learning for large-scale crop yield forecasting." Agricultural Systems 187 (2021): 103016.

[20] Nishant, Potnuru Sai, et al. "Crop yield prediction based on Indian agriculture using machine learning." 2020 International Conference for Emerging Technology (INCET). IEEE, 2020.

[21] Tang, Long, Yingjie Tian, and Panos M. Pardalos. "A novel perspective on multiclass classification: Regular simplex support vector machine." Information Sciences 480 (2019): 324-338.

[22] Gao, Zheming, et al. "A novel kernel-free least squares twin support vector machine for fast and accurate multiclass classification." Knowledge-Based Systems 226 (2021): 107123.

[23] [1] Kaya, Aydin, et al. "Analysis of transfer learning for deep neural network based plant classification models." Computers and electronics in agriculture 158 (2019): 20-29.

[2] Zhong, Liheng, Lina Hu, and Hang Zhou. "Deep learning based multi-temporal crop classification." Remote sensing of environment 221 (2019): 430-443.

[3] Dokic, K., L. Blaskovic, and D. Mandusic. "From machine learning to deep learning in agriculture–the quantitative review of trends." IOP Conference Series: Earth and Environmental Science. Vol. 614. No. 1. IOP Publishing, 2020.

[4] Dang, Chaoya, et al. "Autumn crop yield prediction using data-driven approaches:- support vector machines, random forest, and deep neural network methods." Canadian journal of remote sensing 47.2 (2021): 162-181.

[5] Moreno-Revelo, Mónica Y., et al. "Enhanced convolutionalneural-network architecture for crop classification." Applied Sciences 11.9 (2021): 4292.

[6] Jain, Sonal, and Ramesh Dharavath. "Memetic salp swarm optimization algorithm based feature selection approach for crop disease detection system." Journal of Ambient Intelligence and Humanized Computing (2021): 1-19.

[7] Kianat, Jaweria, et al. "A joint framework of feature reduction and robust feature selection for cucumber leaf diseases recognition." Optik 240 (2021): 166566.

[8] Sood, Shivani, and Harjeet Singh. "Computer vision and machine learning based approaches for food security: A review." Multimedia Tools and
Applications 80.18 (2021): 27973-27999.

[9] Radhakrishnan, Sreevallabhadev. "An improved machine learning algorithm for predicting blast disease in paddy crop." Materials Today: Proceedings 33 (2020): 682 686.

[10] Kalimuthu, M., P. Vaishnavi, and M. Kishore. "Crop prediction using machine learning." 2020 third international conference on smart systems and inventive technology (ICSSIT). IEEE, 2020.

[11] Kalimuthu, M., P. Vaishnavi, and M. Kishore. "Crop prediction using machine learning." 2020 third international conference on smart systems and inventive technology (ICSSIT). IEEE, 2020.

[12] Nischitha, K., et al. "Crop prediction using machine learning approaches." International Journal of Engineering Research & Technology (IJERT) 9.08 (2020): 23-
26.

[13] Rao, Madhuri Shripathi, et al. "Crop prediction using machine learning." Journal of Physics: Conference Series. Vol. 2161. No. 1. IOP Publishing, 2022.

[14] Gupta, Archana, et al. "Smart crop prediction using IoT and machine learning." International Journal of Engineering Research & Technology (IJERT) 9.3 (2021).

[15] Kumar, Y. Jeevan Nagendra, et al. "Supervised machine learning approach for crop yield prediction in agriculture sector." 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2020.

[16] Elavarasan, Dhivya, and PM Durairaj Vincent. "Crop yield International Conference on Artificial Intelligence, Smart prediction using deep reinforcement learning Grid and Smart City Applications: AISGSC 2019. Springer model for sustainable agrarian applications." IEEE access 8 International Publishing, 2020.
(2020): 86886-86901.

[17] Reddy, D. Jayanarayana, and M. Rudra Kumar. "Crop yield prediction using machine learning algorithm." 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021.

[18] Pant, Janmejay, et al. "Analysis of agricultural crop yield prediction using statistical techniques of machine learning." Materials Today: Proceedings 46 (2021): 10922 10926.

[19] Paudel, Dilli, et al. "Machine learning for large-scale crop yield forecasting." Agricultural Systems 187 (2021): 103016.

[20] Nishant, Potnuru Sai, et al. "Crop yield prediction based on Indian agriculture using machine learning." 2020 International Conference for Emerging Technology (INCET). IEEE, 2020.

[21] Tang, Long, Yingjie Tian, and Panos M. Pardalos. "A novel perspective on multiclass classification: Regular simplex support vector machine." Information Sciences 480 (2019): 324-338.

[22] Gao, Zheming, et al. "A novel kernel-free least squares twin support vector machine for fast and accurate multiclass classification." Knowledge-Based Systems 226 (2021): 107123.

[23]https://www.kaggle.com/datasets/chinmaynagesh/crop-yieldper-state-and-rainfall-data-of-india?resource=download

[24] A. Tamilmani, M. Sughasiny, "Gain Ratio With Optimization Based Feature Selection Method", Webology, Volume 18, No. 6, 2021, 6545-6557.

[25] A. Tamilmani, M. Sughasiny, "Optimized Artificial Neural Network Classifier for the Prediction of Rainfall", International Journal of Computational Intelligence in Control, Vol.13 No. 2 December, 2021, 377-387.

[26] Poornappriya, T.S., Gopinath, R., Application of Machine Learning Techniques for Improving Learning Disabilities, International Journal of Electrical Engineering and Technology (IJEET), 11(10), 392-402 (2020).

[27] Poornappriya, T.S., Selvi, V., Evolutionary Optimization of Artificial Neural Network for Diagnosing Autism Spectrum Disorder, International Journal of Electrical Engineering and Technology (IJEET), 11(7), 47-61 (2020).

[28] Priyadharshini, D., Poornappriya, T.S., & Gopinath, R., A fuzzy MCDM approach for measuring the business impact of employee selection, International Journal of Management (IJM), 11(7), 1769-1775 (2020).

[29] Poornappriya, T. S., and R. Gopinath. "Segmentation Of Cervical Cancer Lesions: A Comparative Analysis Of Image Processing Algorithms." Webology (ISSN: 1735-188X) 18.4 (2021).

[30] Poornappriya, T. S., and M. Durairaj. "High relevancy low redundancy vague set based feature selection method for telecom dataset." Journal of Intelligent & Fuzzy Systems 37.5 (2019): 6743-6760.

[31] Durairaj, M., and T. S. Poornappriya. "Why feature selection in data mining is prominent? A survey." Proceedings of