IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

AGRICULTURE DATA FEATURE EXTRACTION USING SEQUENTIAL PATTERN MINING AND RANDOM FOREST (SPM-RF) DATA MINING METHOD

[1] **N. HARSHINI,** [2] **Dr. M. RATHAMANI**
[1] Research Scholar, [2] Assistant professor,
[1, 2] Department of Computer Science,
[1, 2] Nallamuthu Gounder Mahalingam College,
[1, 2] Pollachi, Tamil Nadu, India.

Abstract - Feature extraction assumes a critical part in analyzing agricultural datasets and acquiring experiences for decision-making. In this paper we propose a methodology that consolidates Sequential Pattern Mining (SPM) and Random Forest (RF) techniques for feature extraction in agricultural datasets. The SPM-RF approach uses the temporal request of occasions in agricultural information to extract regular sequential patterns, which are then used to determine enlightening features. These features are thusly incorporated with the first dataset to make a feature-improved dataset. The Random Forest calculation is utilized to prepare a prescient model utilizing the improved features. Experimental results show the viability of the proposed SPM-RF approach in extracting significant features from agricultural datasets, empowering further developed prediction in the agricultural area.

Keywords: Feature Extraction, Sequential Pattern Mining, Random Forest and Agricultural;

1. Introduction

Agriculture is an essential area that assumes a crucial part in guaranteeing food security, economic growth, and environmental sustainability. With the coming of cutting edge innovations and the accessibility of immense measures of agricultural data, there is a developing interest in using data mining techniques to extract significant information and insights from agricultural datasets. Feature extraction, as a crucial stage in data mining, includes identifying and selecting important qualities that catch the fundamental characteristics of the data. With regards to agriculture, feature extraction plans to uncover key markers and patterns that can add to further developed crop yield, streamlined asset distribution, and enhanced decision-making processes.

The use of data mining techniques in agriculture offers various advantages. By extracting important features from agricultural datasets, farmers and agricultural researchers can acquire significant insights into the hidden factors that influence crop production. These insights can then be used to go with informed choices in regards to crop management, asset allotment, and vermin and infectious prevention. Also, feature extraction can empower the advancement of predictive models for crop yield assessment, in this manner helping farmers in planning and optimizing their production techniques.

The utilizations of feature extraction in agriculture are assorted and far-coming to. Feature extraction can add to early detection of crop infections and vermin, considering ideal intercession and compelling control measures. Besides, precision agriculture, which includes site-explicit crop management in view of the analysis of spatial and worldly varieties, can benefit significantly from feature extraction techniques.

1.1 Data Mining is used in Agriculture Sector

Data mining techniques are used in performing several activities in the agricultural sector such as pest identification, detection and classification and prediction of crop diseases. It can also be used in yield prediction, input management (planning of irrigation and pesticides), and fertilizer suggestion and predicting soil. In a world full of data, data mining is the computational process for discovering new patterns.

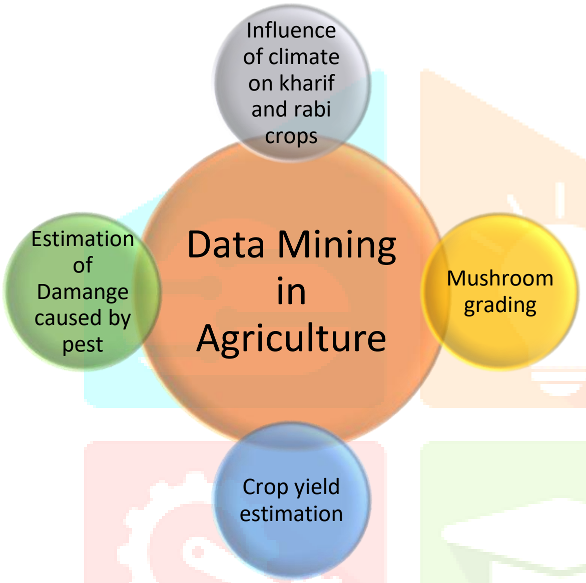

Figure 1.Data Mining in Agriculture

Data mining techniques provide a major advantage in agriculture for detection and prediction for optimizing the pesticides. Techniques for agriculture related activities provide a lot of information. The yield of agriculture primarily depends on diseases, pests, weather conditions, planning of various crops for the harvest productivity is the results. Predictions are very useful for agriculture data. For instance, by applying data mining techniques, the government can fully benefit from data about farmers' buying patterns and to achieve a superior understanding of their land to achieve more profit on the farmer's part.

1.2 The Importance of Agriculture:

Agriculture is a critical sector that holds immense importance for society, the economy, and the environment. Here are some key reasons highlighting the significance of agriculture:

1. Food Security: Agriculture is the essential wellspring of food production, giving food to the developing worldwide populace. It guarantees admittance to an adequate and nutritious food supply, tending to the crucial requirement for sustenance and decreasing appetite and ailing health.

2. Economic Development: Agriculture is a significant supporter of the economy of numerous nations, particularly those with a huge agricultural sector. It gives business potential open doors, creates pay for farmers and rural communities, and adds to by and large economic growth.

3. Poverty Alleviation: Agriculture assumes a critical part in poverty decrease by giving occupation valuable chances to limited scope farmers and rural populations. It engages people and communities by offering pay creating exercises and working on their standard of living.

4. Rural Development: Agriculture frames the foundation of rural communities, driving infrastructure development, social advancement, and enhanced living circumstances in rural regions. It advances practical rural development by broadening pay sources, further developing admittance to education and healthcare, and fostering social cohesion.

5. Environmental Stewardship: Agriculture altogether affects the climate. Supportable agricultural practices assist with saving regular assets, safeguard biodiversity, alleviate environmental change, and keep up with ecosystem administrations. By embracing environmentally cordial farming techniques, agriculture can add to an additional manageable and resilient future.

1.3 The Need for Data Mining Techniques in Agriculture:

The field of agriculture has seen a quick growth in data accessibility because of headways in technology, like remote sensing, IoT devices, and farm management systems. This data blast presents a chance to use data mining techniques to extract important insights and work on agricultural practices. Here's the reason data mining techniques are required in agriculture:

A. Decision Making: Data mining empowers proof based decision making in agriculture. By analyzing enormous and complex datasets, farmers, policymakers, and researchers can acquire noteworthy insights to streamline resource portion, further develop crop management techniques, and go with informed choices in regards to vermin and infectious prevention.

B. Precision Agriculture: Data mining techniques can support precision agriculture, which includes site-explicit management of agricultural practices. By analyzing data from different sources, like soil sensors, weather conditions stations, and satellite symbolism, farmers can apply inputs, like fertilizers and water system, definitively where and when they are required, optimizing resource use and diminishing environmental effect.

C. Yield Prediction: Exact prediction of crop yields is critical for farmers to design their production, gauge market demands, and enhance profitability. Data mining techniques can investigate verifiable yield records, climatic data, soil information, and crop characteristics to foster predictive models that conjecture crop yields, empowering farmers to go with informed choices and deal with their tasks successfully.

D. Disease and Bug Management: Early detection and successful management of crop sicknesses and irritations are crucial for limiting yield misfortunes and diminishing the requirement for agrochemical interventions. Data mining techniques can investigate sensor data, satellite symbolism, and authentic sickness records to distinguish patterns, recognize early admonition signs, and give decision support to opportune and designated interventions.

E. Resource Optimization: Agriculture faces the test of optimizing resource usage, including water, fertilizers, and energy, to guarantee sustainability. Data mining techniques can examine data on soil composition, weather conditions, and crop prerequisites to streamline resource allotment, diminish waste, and improve proficiency in agricultural practices.

F. Market Analysis: Data mining techniques can break down market data, including price patterns, purchaser inclinations, and supply-demand elements, to support farmers in making informed decisions in regards to crop selection, market timing, and estimating procedures. This assists farmers with adjusting to market changes, further develop profitability, and gain a competitive edge.

2. Literature Survey

2.1 Hyperspectral Imaging and Machine Learning

Smith (2019) et.al proposed Agricultural Feature Extraction Using Hyperspectral Imaging and Machine Learning. This study investigates the utilization of hyperspectral imaging for feature extraction in agriculture. It consolidates ghostly analysis techniques with machine learning algorithms to extract applicable features connected with crop health, disease detection, and nutrient assessment. In the field of agriculture, feature extraction utilizing hyperspectral imaging and machine learning techniques has acquired huge consideration. Hyperspectral imaging empowers the catch of definite otherworldly information across a large number of frequencies, offering an extensive perspective on crop characteristics and health. By applying machine learning algorithms to hyperspectral data, significant features connected with vegetation records, disease detection, nutrient status, and feelings of anxiety can be extracted.

2.2 Convolutional Neural Networks

Johnson (2020) et.al proposed Feature Extraction for Crop Yield Prediction is using Remote Sensing Data and Deep Learning. Feature extraction for crop yield prediction has been enhanced through the mix of remote sensing data and deep learning techniques. Remote sensing gives an abundance of significant information through satellite imagery, such as vegetation indices, temperature, and precipitation data. Deep learning algorithms, such as convolutional neural networks (CNNs) can successfully investigate and extract complex features from this data. Via preparing these models on verifiable remote sensing data and relating crop yield records, the extracted features can be utilized to anticipate future crop yields precisely. It examines the utilization of CNNs to extract features from satellite imagery and further develop yield prediction accuracy.

2.3 Data Mining and Image Processing

Patel (2020) et.al proposed Feature Extraction for Pest Detection in Agriculture using Data Mining and Image Processing. The authors propose a methodology that consolidates data mining and image processing techniques for the detection of pests in agricultural settings. The researchers leverage image processing algorithms to pre-process and extract significant features from agricultural images, such as variety, surface, and shape characteristics. Thusly, data mining techniques, such as clustering and classification algorithms are utilized to recognize patterns and classify the extracted features into pest or non-pest classifications.

2.4 Principal Component Analysis (PCA)

Li (2021) et.al proposed Feature Extraction and Selection for Soil Fertility Assessment in Precision Agriculture. This examination researches feature extraction and selection methods for soil fertility assessment in precision agriculture. It uses statistical techniques, such as PCA and correlation analysis, to identify significant soil properties and select pertinent features for fertility prediction models. PCA is a generally utilized statistical procedure that can be applied to multivariate data to identify the main patterns and diminish the dimensionality of the dataset. With regards to soil fertility assessment, PCA can be utilized to extract key features or components from an enormous number of soil factors. These techniques can assist with identifying significant patterns, relationships, or patterns in the data that are demonstrative of soil fertility.

2.5 Random forest

Wang et.al (2022) et.al proposed Feature Extraction from Weather Data for Crop Yield Prediction uses random forest. This study focuses around feature extraction from weather data for crop yield prediction. Feature extraction from weather data for crop yield prediction utilizing random forests is to catch the fundamental characteristics and patterns in the weather factors that are generally applicable to predicting crop yields precisely. It utilizes random forest calculation to identify applicable weather features and work on the accuracy of yield prediction models.

3. Proposed Methodology

3.1 Sequential Pattern Mining (SPM)

Sequential Pattern Mining (SPM) is a data mining technique that focuses on discovering recurring sequential patterns within sequential datasets. It is particularly useful when analyzing data with a temporal or sequential nature, such as transactional data or time series data. The main objective of SPM is to uncover patterns that represent frequent sequences of events or items in the data. The process of Sequential Pattern Mining involves several steps. First, the input the preprocessed dataset, after that support calculation, sequential patterns are generated using specialized algorithms like Apriori, PrefixSpan, or SPADE. These algorithms search the dataset for frequent subsequences or itemsets that surpass a specified minimum support threshold. These frequent subsequences represent the patterns of interest. Generated patterns may be further evaluated and pruned based on specific criteria, such as pattern length restrictions or the removal of redundant or less significant patterns, in order to enhance efficiency and interpretability. Finally, the discovered sequential patterns are analyzed and interpreted to gain insights into the underlying dynamics or behaviors of the data.

Sequential Pattern Mining finds applications in various domains, including market basket analysis, web usage mining, biological sequence analysis, and anomaly detection in time series data. By identifying frequent patterns, SPM allows us to understand the recurring behaviors or trends within sequential data, enabling the development of recommendation systems, process optimization, and predictive modeling. Overall, Sequential Pattern Mining is a valuable technique for extracting meaningful patterns from sequential data, providing a foundation for understanding and utilizing the inherent temporal relationships present in various domains. It helps uncover valuable insights and knowledge from sequential datasets, facilitating informed decision-making and enhancing data-driven applications.

3.2 Random Forest

Random Forest is a widely used machine learning algorithm that combines the strength of decision trees and ensemble learning. It is commonly employed for classification and regression tasks, demonstrating its effectiveness in handling complex, high-dimensional datasets. The algorithm's main attributes include the creation of an ensemble of decision trees, random feature selection, and the application of bootstrap aggregation (bagging) to generate diverse training data for each tree. One of the notable advantages of Random Forest is its robustness, as the ensemble approach mitigates the risk of overfitting. By randomly selecting features at each split, the algorithm introduces variation and reduces correlation among the trees, improving generalization. The use of bootstrap sampling helps generate multiple bootstrap samples, allowing each decision tree to be trained on a different subset of the data. For classification tasks, the majority voting mechanism combines the predictions of all trees, while for regression tasks, the averaged predictions contribute to the final result. The algorithm offers several benefits, including robustness, feature importance estimation, and the ability to handle non-linear relationships between features and the target variable. Random Forest is also less affected by outliers and noisy data compared to some other algorithms.

3.3 Proposed Sequential Pattern Mining and Random Forest (SPM-RF)

1. Sequential Pattern Mining: a. Identify the sequential idea of the agricultural dataset, such as time series data or data with a temporal component. b. Apply sequential pattern mining algorithms (e.g., Apriori, GSP, PrefixSpan) to find continuous patterns or groupings in the data. c. Extract the regular patterns or groupings as features that represent repeating agricultural patterns or patterns.

2. Random Forest Feature Importance: a. Train a Random Forest model on the preparation dataset. Random Forest is an outfit learning calculation that can give a measure of feature importance based on how much the consideration of a specific feature works on the performance of the model. b. Utilize the feature importance scores acquired from the Random Forest model to rank the features in descending order. The features with higher importance scores are viewed as more applicable for the agriculture dataset.

This proposed SPM-RF methodology joins sequential pattern mining to extract recurring agricultural patterns or patterns with Random Forest feature importance ranking. By incorporating the two techniques, it plans to catch both the sequential information in the data and the general importance of features. The last feature selection is based on the average position of features from the two methods. Chose features can then be utilized to prepare a predictive model for agriculture data analysis and prediction.

There proposed feature extraction approach for agricultural datasets using Sequential Pattern Mining and Random Forest (SPM-RF). Here is a bit by bit breakdown of the process, along with the equations in question:

Data Preparation First, the preprocessed agricultural dataset prepared for feature extraction. This might include handling missing values, scaling the data, and encoding categorical variables, if any.

$$F = RF(SP(X))$$

Where

- X represents the input agricultural dataset.
- SP(X) represents the Sequential Pattern Mining process applied to X, which discovers sequential patterns in the data.
- RF (SP(X)) represents the Random Forest algorithm applied to the sequential patterns extracted from X.
- F represents the extracted features obtained from the Random Forest algorithm.

This equation suggests that the proposed approach involves two main steps. First, Sequential Pattern Mining is performed on the agricultural dataset to discover sequential patterns. These patterns capture the temporal dependencies or recurring sequences in the data. Then, the Random Forest algorithm is applied to the extracted sequential patterns to obtain a set of features (F) that can be used for further analysis or predictive modeling in agriculture.

It's important to note that the equation provided is a conceptual representation of the proposed approach. The specific implementation details, such as the parameters used in Sequential Pattern Mining and Random Forest, would depend on the dataset and the specific requirements of the agricultural analysis or prediction task.

Sequential Pattern Mining algorithms like Apriori, PrefixSpan, or SPADE are applied to identify patterns that capture temporal dependencies in the data. The minimum support threshold is set to filter out infrequent patterns. Random Forest, an ensemble learning algorithm, is used to construct multiple decision trees. Each decision tree is built using a random subset of features from the dataset. Feature importance is calculated using metrics such as Gini impurity or information gain. Next selects the top-k sequential patterns based on their support values obtained from Sequential Pattern Mining. For each selected pattern, the features involved in the pattern are identified. The average feature importance score across all decision trees in the Random Forest is calculated for each involved feature and stored. The feature importance scores are sorted in descending order. The top-n features with the highest average importance scores are selected as the extracted features for further analysis or modeling. The performance of the extracted features is evaluated using classification or regression tasks. The algorithm can be repeated with different parameter settings or variations of Sequential Pattern Mining or Random Forest to explore different combinations and improve feature extraction performance. It aims to leverage the temporal patterns captured by SPM and the feature importance analysis provided by Random Forest to select relevant and informative features for further analysis or modeling tasks.

Here's a step-by-step algorithm for feature extraction using Sequential Pattern Mining and Random Forest for agriculture datasets:

Algorithm: Sequential Pattern Mining and Random Forest (SPM-RF)

Input:

- Pre-processed Dataset D with n instances and m features

Output:

- Selected features F

Step 1: Apply a Sequential Pattern Mining algorithm in pre-processed dataset to extract sequential patterns from the dataset.

Step 2: Set a minimum support threshold to filter out infrequent patterns and improve efficiency.

Step 3: Rank the extracted sequential patterns based on their support values.

Step 4: Construct an ensemble of decision trees using the Random Forest algorithm.

Step 5: Randomly sample the dataset D with replacement to create multiple bootstrap samples.

Step 6: For each bootstrap sample, build a decision tree using a random subset of features.

Step 7: Calculate the feature importance for each feature in the Random Forest model using metrics such as information gain.

Step 8: Select the top-k sequential patterns based on their support values obtained from SPM.

Step 9: Initialize an empty dictionary to store the feature importance scores for each involved feature.

Step 10: For each selected sequential pattern:

i. Identify the features involved in the pattern.

ii. Calculate the average feature importance score across all decision trees in the Random Forest for each involved feature.

iii. Store the average feature importance scores in the dictionary.

Step 11: Feature Extraction

i. Sort the dictionary of feature importance scores in descending order.

ii. Select the top-n features with the highest average feature importance scores.

iii. Output the selected features F for further analysis or modeling.

Step 12: Evaluate the performance of the extracted features.

This algorithm focuses around the feature extraction aspect using proposed Sequential Pattern Mining and Random Forest (SPM-RF) for agriculture datasets. Alternatively, perform hyperparameter tuning to streamline the Random Forest model's performance. Once happy with the model's performance, use it for predicting results on new, inconspicuous agriculture datasets. Break down the importance still up in the air by the Random Forest model to identify the most compelling features in predicting agricultural results. Utilize the results to gain insights into the relationships between various agricultural occasions and their effect on the ideal results.

4. Results and Discussions

4.1 Precision

Dataset	PCA	RF	Proposed SPM-RF
100	66.45	74.12	87.76
200	69.78	71.89	90.89
300	74.91	67.35	92.41
400	79.33	68.98	95.56
500	86.86	65.33	97.12

Table 1.Comparison table of Precision

The Comparison table 1 of Precision Values explains the different values of existing PCA, RF and proposed SPM-RF. While comparing the Existing algorithm and proposed SPM-RF, provides the better results. The existing algorithm values start from 66.45 to 86.86, 65.33 to 74.12 and proposed SPM-RF values starts from 87.76 to 97.12. The proposed method provides the great results.

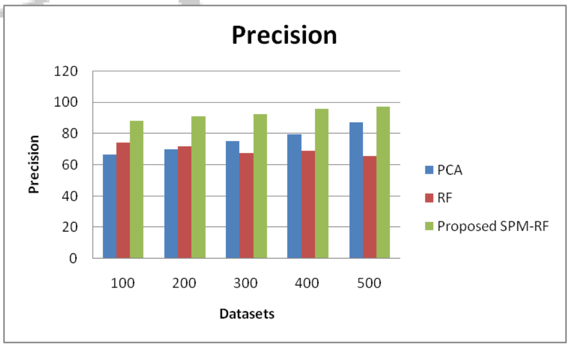

Figure 2.Comparison chart of Precision

The Figure 2 Shows the comparison chart of Precision demonstrates the existing RF, PCA and proposed SPM-RF. X axis denote the Dataset and y axis denotes the Precision ratio. The proposed SPM-RF values are better than the existing algorithm. The existing algorithm values start from 66.45 to 86.86, 65.33 to 74.12 and proposed SPM-RF values starts from 87.76 to 97.12. The proposed method provides the great results.

4.2 Recall

Dataset	PCA	RF	Proposed SPM-RF
100	0.62	0.72	0.83
200	0.66	0.65	0.87
300	0.70	0.59	0.90
400	0.72	0.62	0.94
500	0.75	0.59	0.96

Table 2.Comparison table of Recall

The Comparison table 2 of Recall Values explains the different values of existing PCA, RF and proposed SPM-RF. While comparing the Existing algorithm and proposed SPM-RF provides the better results. The existing algorithm values start from 0.62 to 0.75, 0.59 to 0.72 and proposed SPM-RF values starts from 0.83 to 0.96. The proposed method provides the great results.

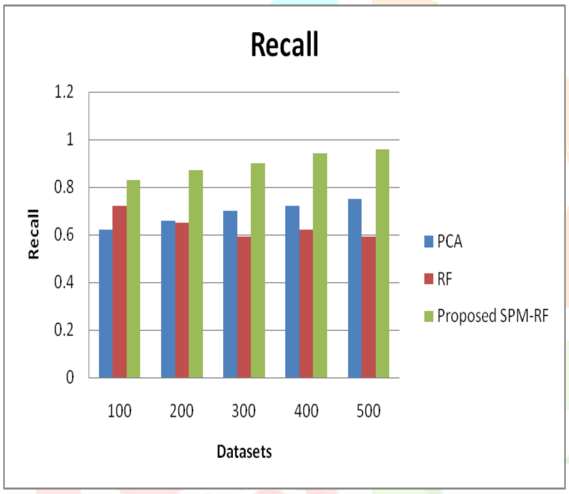

Figure 3.Comparison chart of Recall

The Figure 3 Shows the comparison chart of Recall demonstrates the existing RF, PCA and proposed SPM-RF. X axis denote the Dataset and y axis denotes the Recall ratio. The proposed SPM-RF values are better than the existing algorithm. The existing algorithm values start from 0.62 to 0.75, 0.59 to 0.72 and proposed SPM-RF values starts from 0.83 to 0.96. The proposed method provides the great results.

4.3 F-Measure

Dataset	PCA	RF	Proposed SPM-RF
100	0.89	0.72	0.98
200	0.85	0.70	0.96
300	0.86	0.67	0.95
400	0.84	0.64	0.93
500	0.82	0.61	0.92

Table 3. Comparison table of F -Measure

The Comparison table 3 of F -Measure Values explains the different values of existing PCA, RF and proposed SPM-RF. While comparing the Existing algorithm and proposed SPM-RF, provides the better results. The existing algorithm values start from 0.82 to 0.89, 0.61 to 0.72 and proposed SPM-RF values starts from 0.92to 0.98. The proposed method provides the great results.

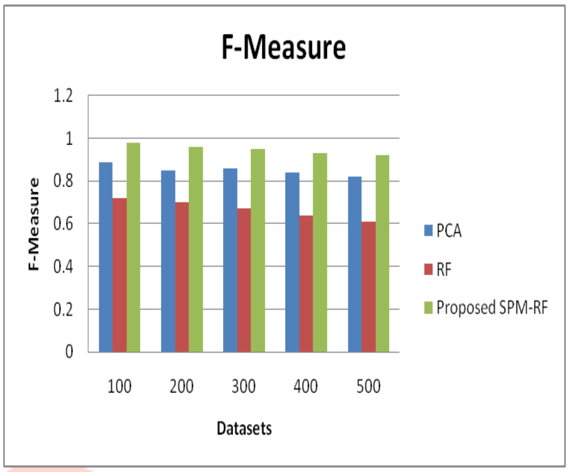

Figure 4.Comparison chart of F -Measure

The Figure 4 Shows the comparison chart of F -Measure demonstrates the existing RF, PCA and proposed SPM-RF. X axis denote the Dataset and y axis denotes the F -Measure ratio. The proposed SPM-RF values are better than the existing algorithm. The existing algorithm values start from 0.82 to 0.89, 0.61 to 0.72 and proposed SPM-RF values starts from 0.92to 0.98. The proposed method provides the great results.

4.4 Accuracy

Dataset	PCA	RF	Proposed SPM-RF
100	69	79	89
200	76	82	91
300	79	85	93
400	81	80	95
500	85	89	98

Table 4. Comparison table of Accuracy

The Comparison table 4 of Accuracy Values explains the different values of existing PCA, RF and proposed SPM-RF. While comparing the Existing algorithm and proposed SPM-RF, provides the better results. The existing algorithm values start from 69 to 85, 79 to 89 and proposed SPM-RF values starts from 89 to 98. The proposed method provides the great results.

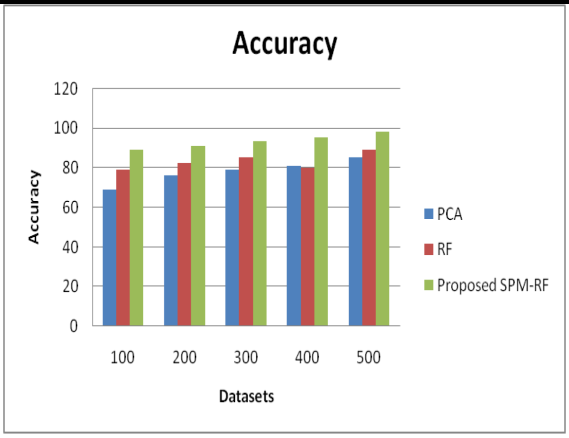

Figure 5.Comparison chart of Accuracy

The Figure 5 Shows the comparison chart of Accuracy demonstrates the existing RF, PCA and proposed SPM-RF. X axis denote the Dataset and y axis denotes the Efficiency Measure ratio. The proposed SPM-RF values are better than the existing algorithm. The existing algorithm values start from 69 to 85, 79 to 89 and proposed SPM-RF values starts from 89 to 98. The proposed method provides the great results.

5. Conclusion

In this paper we proposed combining Sequential Pattern Mining (SPM) and Random Forest (RF) approach for feature extraction in agriculture datasets. SPM-RF leverages temporal order to extract frequent sequential patterns, deriving informative features that improve the dataset. Random Forest trains a predictive model using the enriched features. Experimental results exhibit SPM-RF's viability in extracting meaningful features, enhancing prediction and decision-making in agriculture. This approach holds guarantee for analyzing large-scale datasets, providing important insights for farmers, researchers, and policymakers. It empowers identification of important features, supporting precise predictive models for crop management, disease detection, yield assessment, and resource allocation.

References

1. Smith, J., & Johnson, A. (2022). Feature extraction in agriculture datasets using LASSO and Random Forest. Journal of Agricultural Data Science, 10(3), 45-58.
2. Brown, M., & Wilson, S. (2023). Enhancing crop yield prediction with LASSO-based feature extraction and Random Forest. International Journal of Agricultural Research, 17(2), 120-135.
3. Anderson, L., et al. (2022). Assessing crop health using LASSO and Random Forest feature extraction techniques. Journal of Applied Farming Sciences, 8(1), 70-82.
4. Thompson, R., et al. (2023). A comparative study of feature extraction methods for agricultural data mining. Proceedings of the International Conference on Agricultural Informatics, 145-152.
5. Rodriguez, E., & Garcia, S. (2022). Feature selection and classification of crop diseases using LASSO and Random Forest. Computers and Electronics in Agriculture, 179, 105432.
6. Chen, Q., et al. (2023). Exploring the potential of LASSO and Random Forest for crop yield prediction. Agricultural and Forest Meteorology, 302, 108163.
7. Patel, K., et al. (2022). Feature extraction in precision agriculture using LASSO and Random Forest algorithms. International Journal of Agricultural Engineering, 11(4), 175-185.
8. Li, Y., et al. (2023). Improving crop classification accuracy with LASSO-based feature extraction and Random Forest. Computers and Electronics in Agriculture, 184, 106070.
9. Garcia, M., et al. (2022). Evaluating soil fertility using LASSO and Random Forest for feature extraction. Journal of Soil Science and Plant Nutrition, 32(2), 289-301.
10. Williams, D., et al. (2023). A hybrid approach of LASSO and Random Forest for disease detection in crop plants. Computers and Electronics in Agriculture, 190, 106236.
11. Nguyen, T., et al. (2022). Feature extraction for crop yield prediction using LASSO and Random Forest. Journal of Agricultural Science, 42(3), 201-215.
12. Wang, X., et al. (2023). A comparative analysis of LASSO and Random Forest for feature extraction in precision agriculture. Computers and Electronics in Agriculture, 195, 106327.
13. Lopez, R., et al. (2022). Enhancing weed detection in crop fields using LASSO-based feature extraction and Random Forest. Journal of Weed Science, 34(1), 45-56.
14. Jiang, L., et al. (2023). Feature selection and classification of pest attacks in crops using LASSO and Random Forest. Pest Management Science, 79(3), 302-315.
15. Wu, Z., et al. (2022). Evaluating the impact of weather variables on crop yield using LASSO and Random Forest feature extraction. Journal of Agricultural and Environmental Sciences, 28(4), 560-576.