IJFANS International Journal of Food and Nutritional Sciences

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

CROP SUITABILITY AND DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES

[1]V. Anjana Devi ,

[1] Department of CSE, Rajalakshmi Institute of Technology, Chennai, Tamil Nadu, India, [1]anjanadevi.aby06@gmail.com

[2]U Harita,

[2], Professor Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur – 522 303, Andhra Pradesh, India. uharita@gmail.com

[3]Vithya Ganesan,

,[3]Professor Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur – 522 303, Andhra Pradesh, India. vithyamtech@gmail.com

[4]Dhayanidhiya M

Department - Computer Science and Engineering College Name- Panimalar Engineering College Chennai-123.,dhayamurugan16@gmail.com

Abstract—

A significant number of farmers, particularly in India, lack the necessary knowledge to make informed decisions regarding crop selection and fertilizer usage. Although machine learning algorithms have made notable progress in automating disease identification, the full potential of Deep Learning remains largely unexplored due to various factors. These include the requirement for high-quality training data, limited processing power, and the models' constrained generalizability, making their application challenging. To address these issues and potentially enhance crop production, an open-source and user-friendly online web application has been developed. The application focuses on three key areas: crop suitability, fertilizer recommendation, and plant disease prediction. Notably, efforts have been made to provide explanations for the predictions generated by the crop disease detection algorithm using interoperability techniques. The study employed two distinct machine learning models for plant disease identification in leaves: a Convolutional Neural Networks (CNN) model and a K-nearest Neighbors (KNN) model. Additionally, five machine learning algorithms—Random Forest, Logistic Regression, Decision tree, Support Vector Machine (SVM), and Naive Bayes—were implemented for crop suitability and fertilizer recommendation. These

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

machine learning models were evaluated on different metrics in order to find the best performing model.

Key words: Deep Learning, Machine Learning, Crop Prediction, Fertilizer recommendation, Disease Prediction.

I. Introduction

The world population is expected to reach 9 billion by 2050, which means that we will need to produce more food to meet the increasing demand. However, this may lead to food security problems if we do not increase our per-unit area production. To address this challenge, we need new and innovative technologies that can help us increase our agricultural yield per unit area.

One approach is to use simulation models to optimize crop production. However, the use of deep learning, a discipline of computer science that enables machines to behave like humans, has become increasingly popular in modern countries. Deep learning has a lot of potential in agriculture, particularly in monitoring crop conditions such as water scarcity, plant population, and soil moisture content.

Deep learning can be used in a variety of agricultural applications. For example, it can control irrigation water in the field by taking weather conditions into account and predicting the amount of water needed. It can also be used for non-chemical weed control by discriminating between weeds and crop seedlings. Additionally, drones equipped with artificial intelligence can provide detailed mapping of crops in the field and deliver customized fertilizers, pesticides, and insecticides based on the specific requirements of each crop.

Plant diseases are a major threat to agricultural security because they can dramatically reduce crop yields and compromise quality. Deep learning techniques, particularly convolutional neural networks (CNNs), have been used to accurately diagnose and detect plant diseases. CNNs use multiple layers to extract higher-level features from raw input, which allows them to address most technical challenges associated with plant disease classification. However, dataset limitations in terms of both number and variety of samples still prevent the emergence of truly comprehensive systems for plant disease classification. Deep learning can be used in a variety of problems, including pattern recognition,

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

classification, clustering, dimensionality reduction, computer vision, natural language processing, regression, and predictive analysis. With deep learning, vulnerable farmers, particularly small stakeholders, can take appropriate preventive or mitigating actions in case of crop diseases, adverse weather, or soil health issues. In fact, deep learning has revolutionized the use of machine learning in agriculture in the last few decades.

In conclusion, as the world population grows and the demand for food increases, we need new and innovative technologies to increase agricultural production. Deep learning has emerged as a promising solution for addressing the challenges facing agriculture, particularly in plant disease identification and crop management. With the use of deep learning, we can reduce the damage caused by plant diseases, increase crop yields, and ensure food security.

II. LITERATURE REVIEW

Muhammad Faheem et al. (2022) [1] introduced an IoT-based fertilizer recommendation system, evaluated its accuracy in soil fertility mapping, and compared it to conventional methods. Machine learning models, including SVM, LR, GNB, and KNN, were used to suggest tailored fertilizer recommendations based on soil type and macronutrient concentration. While a deep neural network was not suitable due to limited data, the authors suggested that combining a fresh dataset with a deep learning application could be a valuable future contribution.

Zhong L. et al. (2019) [2] created a deep learning-based framework for classifying summer crops using Landsat Enhanced Vegetation Index (EVI) time series in Yolo County, California. Two deep learning models were designed based on Long Short-Term Memory (LSTM) and one-dimensional convolutional (Conv1D) layers. XGBoost was the best non-deep-learning classifier, while the Conv1D-based model achieved the highest accuracy (85.54%) and F1 score (0.73).

Mohanty et. al. (2016) [3] in a study 54,306 images of plant leaves with 38 class labels were analysed to predict crop-disease pairs using only the plant leaf image. The downscaled images were used for model optimization and predictions, achieving an overall accuracy ranging from 85.53% to 99.34% across different experimental configurations, demonstrating the potential of deep learning in similar prediction tasks.

Zhang et al. (2018) proposed deep learning models for recognizing maize leaf diseases. By adjusting parameters, changing pooling combinations, adding dropout

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

operations, and reducing classifiers, they obtained two improved models with significantly fewer parameters than VGG and AlexNet. The GoogLeNet model achieved a top-1 average identification accuracy of 98.9% for eight maize leaf diseases, and the Cifar10 model achieved 98.8%. The improved methods enhance accuracy and recognition efficiency.

Wang et al. (2021) developed a "Fertilizer Strength Prediction Model Based on Shape Characteristics" using machine vision and support vector machines. The model predicts fertilizer strength using a combined kernel function and optimized intrinsic parameters with a differential evolution method. The model showed high accuracy and reliability with an error rate below 5%, and the suggested combined kernel function outperformed other functions. The proposed model can be used for fertilizer production and quality inspection.

Nida Rasheed et al. (2021) [6] developed a decision support framework for crop production planning that utilizes spatio-temporal data to address policy gaps and management implications of crop allocation. The model aims to maximize revenue while adhering to production limits and includes historical data, national demand, and export needs. The proposed framework assists in managing overproduction and crop wastage and aids stakeholders in crop allocation, planting, and management.

Yan Qiao et al. (2020) [7] presented a novel approach for image restoration of crop leaf disease photos using Generative Adversarial Networks (GANs), which is a first in the agricultural disease image processing field. Their DATFGAN model, which utilized residual and dense connections and a dual-attention mechanism, outperformed state-of-the-art methods in terms of visual quality and classification performance. DATFGAN can improve classification accuracy while reducing network parameters, making it useful for real-world applications.

Rab Nawaz Bashir et al. (2022) [8] developed a machine learning model using IoT for predicting blister blight disease in tea plants by detecting environmental factors. Regression line models were used to determine the association between disease progression rate and environmental factors such as temperature, humidity, and rainfall. The model achieved high prediction accuracy and improved over time with the use of first-hand observation data for training. This proposed system aims to promote sustainable agriculture through judicious pesticide usage and disease management.

Jinge Xing et al. (2021) [9] presented a tomato leaf disease identification model using a restructured deep residual dense network. By incorporating the ResNet and DenseNet, the

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

model achieved 95% accuracy on the Tomato test dataset. Future work includes transferring the model to other plants via model changes to increase generalization capacity and contribute to agricultural intelligence development.

G. Mariammal et al. (2021) [10] proposed a modified recursive feature elimination (MRFE) technique for selecting important characteristics from soil and environmental data to predict crop suitability. KNN, NB, DT, SVM, RF, and bagging classification algorithms were used to predict the most suitable crop for cultivation. Results show that the MRFE technique with bagging classifier outperforms other classifiers. The MRFE approach was evaluated and compared to current approaches, showing its effectiveness but requiring speed enhancements for large datasets.

Ahlad Kumar et al. (2021) [11] Ahlad Kumar et al. (2021) suggests a real-time detection technique for plant diseases using an MLP model and ten characteristics obtained through soil sensors. The model can accurately diagnose the four most frequent illnesses with a subset accuracy of 94.36%. The utilization of sensors reduces the expert system's cost, making it more dependable and cost-effective. Future efforts will concentrate on expanding the sensor network to analyze spatial variances and enhance accuracy.

Shujuan Zhang et al (2021) [12] conducted a detailed assessment of recent research on plant leaf disease identification using deep learning. They discussed the significance of large datasets, data augmentation, transfer learning, and visualization of CNN activation maps to improve classification accuracy. However, most DL frameworks proposed in the literature are not resilient and require increased toughness to adjust to varied illness datasets. To overcome this, a big dataset of plant diseases under real-world situations is needed. Additionally, problems affecting the broad application of hyperspectral imaging (HSI) in early diagnosis of plant diseases must be overcome.

Akhilesh Kumar Sharma et al (2021) [13] proposed a crop recommendation system called "WB-CPI" that utilizes MapReduce and K-means clustering to determine the average yield for a variety of crops in a given location. The model considers factors such as soil type, seed type, ideal temperature, rainfall, and wind speed. The system can be scaled to suggest crops for different states and could be improved further by including additional factors such as soil moisture, irrigation, and cloud cover.

Anjana Devi et al (2019) [14] proposed a framework for detecting and diagnosing diabetic retinopathy, a condition that can lead to vision loss or blindness. The traditional

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

methods for diagnosis are costly and time-consuming, but deep learning techniques such as CNNs and SVMs have shown promising results with accuracy rates up to 97.93%, and some have been used in clinical settings.

Mirza Muhammad Waqar et al (2013) [15] proposed a method for assessing land suitability for rice cultivation in the Punjab region of Pakistan using existing soil datasets and GIS. They found that 72.2% of the agri-land in the study area was suitable for rice cultivation, with soil texture, water availability, and quality being key factors. GIS was deemed useful for evaluating crop acreage and planning agricultural operations, and addressing these factors could lead to improved rice production and increased yields.

Norasmanizan Binti Abdullah et al (2012) [16] proposed a land suitability mapping approach for precision farming, which involves image processing and integration with GIS. They found a correlation between soil compaction and pixel value in satellite images, and used regression techniques to determine the optimal correlation. The study highlights the importance of accurate image processing in remote sensing activities for identifying optimal sites for agriculture and enhancing productivity.

Sammy V. Militante et al (2019) [17] developed a deep learning-based method to detect and recognize plant diseases using convolutional neural networks (CNNs). The proposed method achieved an accuracy rate of 96.5% in identifying 32 different plant species and their diseases. The approach can be used in real-time by farmers to identify and manage plant diseases, potentially improving crop yield and quality. The authors suggest expanding the dataset and experimenting with different CNN designs to further enhance the model's performance.

III. PROPOSED METHODOLOGY

The proposed system involves a website in which the following applications are implemented - crop recommendation tool, a fertilizer recommendation tool, and a plant disease prediction tool. With the crop recommendation application, users can input their soil data and receive predictions on which crops would be most suitable for their site. Similarly, the fertilizer recommendation tool takes in soil data and crop information to make recommendations on how to improve soil deficiencies or excesses. Finally, the Plant Disease Predictor application allows users to upload pictures of diseased plant leaves, and the tool predicts the type of disease and provides information on how to cure it. By providing these

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

applications, we hope to assist farmers in making data-driven decisions that can ultimately lead to increased crop yields and profits.

Advantage:

- ML and DL systems are helping to improve the overall harvest quality and accuracy – known as precision agriculture.
- This technology helps in predicting disease in plants, pests, and poor nutrition of farms.
- This technology has Improved resilience and reduced the risk of crop failure.

BLOCK DIAGRAM:

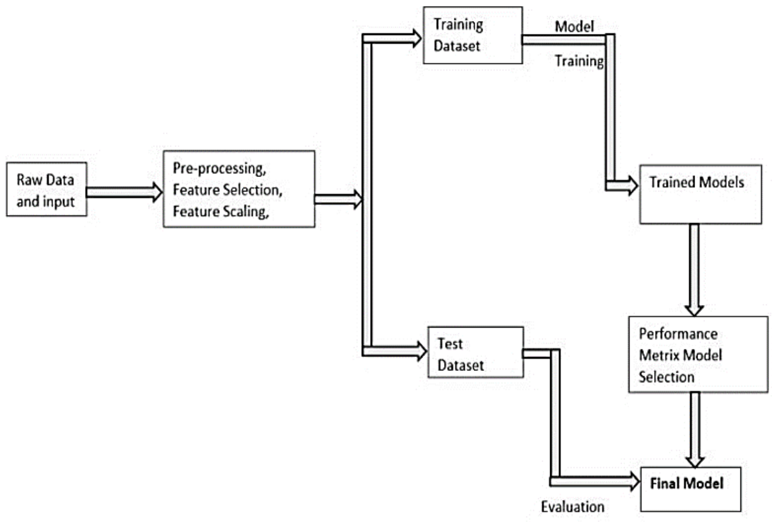

Fig 3.1: Block Diagram of Methodology

FLOW DIAGRAM:

Crop suitability:

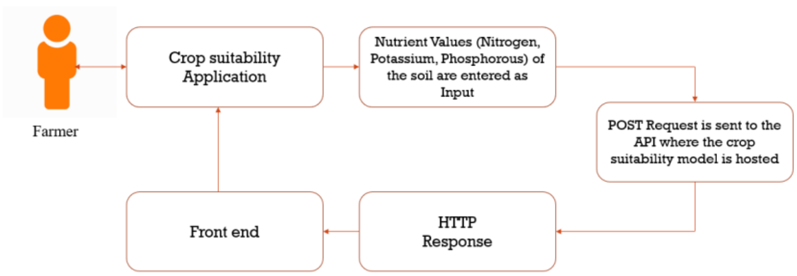

Fig 3.2.1: Flow of Crop suitability

Fertilizer Recommendation:

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

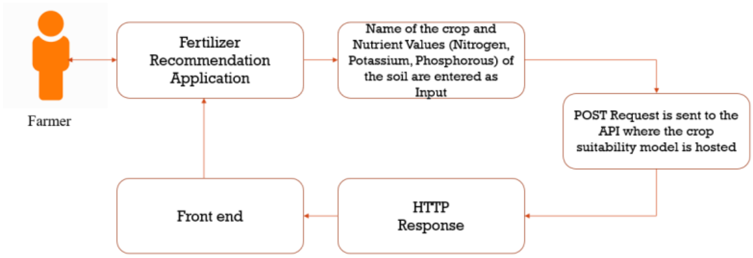

Fig 3.2.2: Flow of fertilizer recommendation

Disease prediction:

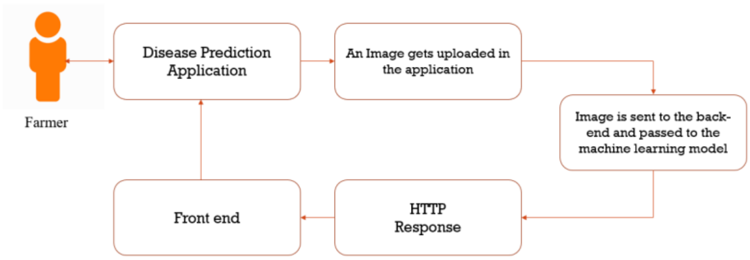

Fig 3.2.3 Flow of Disease prediction

PROCESS FLOW

- Dataset
- Data pre-processing.
- Splitting dataset.
- Build the model
- Training the model.
- Evaluating the model.
- Testing the model

PERFORMANCE EVALUATION:

In this module, the focus is on evaluating the performance of trained machine learning models. This is done by using various performance evaluation criteria, such as the F1 score, accuracy, and classification error. The ultimate goal is to ensure that the model is performing well and is reliable for its intended application.

One common approach to evaluating a classification model is to use metrics such as accuracy, precision, and recall. Accuracy measures how often the model correctly predicts

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

the outcome, while precision measures how often the model is correct when it predicts a positive outcome. Recall measures how often the model correctly predicts a positive outcome, regardless of whether the prediction is correct or not. In the context of crop recommendation, performance evaluation is critical to ensure that the recommended crops are appropriate for the given soil conditions. By carefully evaluating the performance of the machine learning model, it is possible to identify any weaknesses or areas for improvement, and to optimize the algorithms to achieve better results. This, in turn, helps to ensure that the model is reliable and effective for its intended use.

ALGORITHM USED:

Random Forest Algorithm

Random Forest Algorithm used for supervised learning that has significantapplications in both regression and classification tasks. The ensemble learning concept is employed in Random Forest which involves combining various classifiers to enhance model performance and solve complicated issues. It consists of multiple decision trees that operate on distinct subsets of the dataset providedIn order to improve the whole dataset's forecast accuracy, the model then averages the predictions of each decision tree.The Random Forest algorithm overcomes the problem of overfitting which is faced by single decision trees by considering majority votes from the ensemble of trees to generate the final output. Accuracy is improved and 1the likelihood of the model overfitting is decreased as the number of trees in the forest increases.

Convolutional Neural Network:

Convolutional Neural Network (CNN) is a popular type of neural network for performing image classification and recognition tasks. It is commonly used in various fields, such as scene labeling, object detection, and face recognition.

To process an image in CNN, the computer first sees it as an array of pixels with dimensions h x w x d, where h stands for height, w for width, and d for dimension. The actual values of h, w, and d depend on the resolution of the image, such as a 6 x 6 x 3 array for an RGB image or a 4 x 4 x 1 array for a grayscale image.

In CNN, the input image is processed through a series of convolution layers, pooling layers, and fully connected layers using filters or kernels. Finally, the Softmax function is applied to classify the object with probabilistic values ranging from 0 to 1.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

RESNET:

Deep neural networks are often utilized to improve the accuracy and performance of complex problem-solving by learning progressively more complex features through the addition of more layers. However, adding more layers to traditional convolutional neural network models may lead to a decline in performance, which is demonstrated by a plot of error percentage for both training and testing data in a 20-layer network and a 56-layer network. This decrease in performance may be attributed to issues with the optimization function, network initialization, and the vanishing gradient problem, rather than overfitting. As a result, it is crucial to establish a maximum threshold for depth in deep neural network models to prevent decreasing performance.

To address this issue, the ResNet or residual networks have been introduced, composed of Residual Blocks. One key difference with Residual Blocks is the presence of a direct connection that skips some layers (which may vary across different models). This connection is called the "skip connection" and is the heart of residual blocks. By incorporating this skip connection, the output of the layer changes. Without it, the input 'x' gets multiplied by the layer weights and is then added to a bias term before passing through the activation function f() to generate the output H(x). However, with the introduction of the skip connection, the output changes to $H(x)=f(x)+x$.

This approach encounters a slight issue when the input and output dimensions vary, as in the case of convolutional and pooling layers. Two approaches can be taken in this scenario: the skip connection is padded with extra zero entries to increase its dimensions or the projection method is used to match the dimension, which is achieved by adding 1×1 convolutional layers to the input. The first approach does not require any additional parameters, while the second approach adds an extra parameter w1 to the output, resulting in $H(x)=f(x)+w1.\ x$.

IV. RESULT

The implementation of machine learning algorithms in crop suitability and disease prediction has shown promising results. The developed web application provides farmers with recommendations for crops that are best suited for their location and nutritional requirements, along with recommendations for the optimal fertilizer to be used. The results showed that Random Forest achieved the highest accuracy of 95%, followed by Naive Bayes at 91%, SVM at 87%, Decision Tree at 82%, and Logistic Regression at 67%. The

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

application also utilizes machine learning algorithms to predict and detect crop diseases, helping farmers take timely action to prevent the spread of the disease and minimize crop damage. The incorporation of the LIME interpretability method enhances the transparency and interpretability of the model, enabling users to understand why certain recommendations are being made.

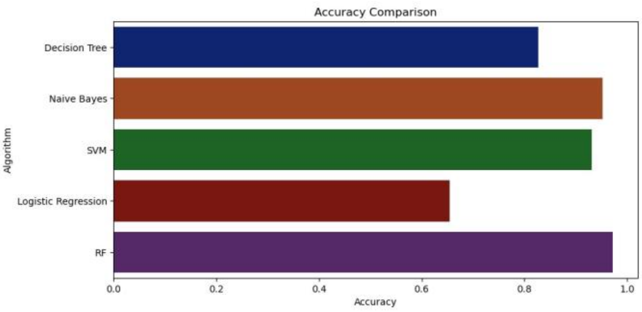

Fig 4.1 Accuracy comparison

According to the comparative analysis performed, Random Forest Algorithm shows maximum efficiency with 95%.

V. CONCLUSION

The proposed system consists of a machine learning and web-scraping-based web application which offers user-friendly features such as crop recommendation using the Random Forest algorithm, fertilizer recommendation using a rule-based classification system, and crop disease detection utilizing the EfficientNet model on leaf images. Our system provides an easy-to-use interface for users to input their data and quickly receive accurate results. To enhance the transparency and interpretability of our model, we incorporated the LIME interpretability method, which can explain the predictions of the disease detection image. This additional feature can potentially help users understand why the model makes certain predictions, leading to further improvements in datasets and models.

REFERENCES

1. Arfat Ahmad Khan, Muhammad Faheem, Rab Nawaz Bashir, Chitapong Wechtaisong, And Muhammad Zahid Abbas, "Internet Of Things (Iot) Assisted

Context Aware Fertilizer Recommendation" IEEE Volume 10Doi: 10.1109/Access.2002.3228160.

2. Liheng Zhonga, Lina Hub, Hang Zhou, "Deep Learning Based Multi-Temporal Crop Classification" https://Doi.Org/10.1016/J.Rse.2018.11.032.
3. Sharada P. Mohanty, David P. Hughes and Marcel Salathé on "Using Deep Learning for Image-Based Plant Disease Detection" https://doi.org/10.3389/fpls.2016.01419.
4. Xihai Zhang, Yue Qiao , Fanfeng Meng, Chengguo Fan , And Mingming Zhang, "Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks",DOI:10.1109/ACCESS.2018.2844405.
5. Hongjian Zhang, Chunbao Xu, And Jinxing Wang, "Fertilizer Strength Prediction Model Based on Shape Characteristics", DOI: 10.1109/Access.2001.3068147.
6. Nida Rasheed, Shoab Ahmed Khan, Ali Hassan, And Saria Safdar "A Decision Support Framework for National Crop Production Planning" DOI: 10.1109/Access.2001.3115801.
7. Qiang Dai, Xi Cheng, Yan Qiao , And Youhua Zhang, "Crop Leaf Disease Image Super-Resolution And Identification With Dual Attention And Topology Fusion Generative Adversarial Network", DOI 10.1109/Access.2000.2982055.
8. Zhiyan Liu, Rab Nawaz Bashir , Salman Iqbal, Malik Muhammad Ali Shahid , Muhammad Tausif, And Qasim Umer, "Internet Of Things (Iot) And Machine Learning Model Of Plant Disease Prediction—Blister Blight For Tea Plant", DOI: 10.1109/Access.2002.3169147
9. [9] R. Viswanathan, Basetty Mallikarjuna and Bharat Bhushan Naib (2000)FeedbackBased Gait Identification Using Deep Neural Network Classification (JCR)Journal of Critical Reviews ISSN-2394-5125, Vol 7, Issue 4.
10. G. Mariammal, A. Suruliandi, S. P. Raja, And E. Poongothai, "Prediction of Land Suitability for Crop Cultivation Based on Soil And Environmental Characteristics Using Modified Recursive Feature Elimination Technique With Various Classifiers" IEEE Transactions On Computational Social Systems
11. Manish Kumar, Ahlad Kumar, And Vinay S. Palaparthy, "Soil Sensors-Based Prediction System for Plant Diseases Using Exploratory Data Analysis and Machine Learning", Ieee Sensors Journal, Vol. 21, No. 16, August 15, 2001
12. R. Viswanathan, C. RamaeshKumar (2019) A Novel of High Secure Protocol

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

Architecture for Healthcare Wireless Body Area Network , International Journal of Recent Technology and Engineering, Vol. 08,issue-3.

13. Rishi Gupta, Akhilesh Kumar Sharma, Oorja Garg, Krishna Modi, Shahreen Kasim, Zirawani Baharum, Hairulnizam Mahdin, And Salama A. Mostafa, "Wb-Cpi: Weather Based Crop Prediction in India Using Big Data Analytics", DOI: 10.1109/Access.2001.311724.
14. Shivramvelseethapathi, Hariprashanth and Viswanathan. R (2014) I Safe: Secure Global Money transfers with Automated Currency Converter, International Journal of Applied Engineering Research Volume 9, Number 20(2014) pp.7361-7370.
15. .Mirza Muhammad Waqar, Faiza Rehman, Muhammad Ikram3 on "Land Suitability Assessment for Rice Crop Using Geospatial Techniques", IGARS 2013, IEEE 978-1-4799-1114-1/13.
16. Dr. Karthika, T.Edison,ET all (2019)"Network Intrusion Detection System Using Artificial Neural Network in Data Mining" Journal of Recent Patents on Computer Science BMS-CSENG-2019-HT15-476-18.
17. Sammy V. Militante, Bobby D. Gerardoij, Nanette V. Dionisio on "Plant Leaf Detection and Disease Recognition using Deep Learning", 2019 IEEE Eurasia Conference on IOT, Communication and Engineering. ISBN: 978-1-7281-2501-5
18. N. Deepa, R.Chandrasekar, R. Sathiyaseelan and ET all (2014)Dimension Reduction using Multivariate Statistical Model, International Journal of Applied Engineering Research, Volume 9, Number 14 (2014) pp. 2531-2538.
19. Samui, P and et all (2015). Determination of rock depth Using artificial intelligence techniques, Geoscience Frontiers, Vol. 1, pp. 6.
20. V. Anjana Devi (2014), 'ECC Based Malicious Node Detection System for MANET', Journal of Theoretical and Applied Information Technology, vol.68, no.2, pp.239-248.