AN EFFECTIVE SEGMENTATION AND LIMO CLASSIFICATION FOR PADDY DISEASE DETECTION USING DEEP LEARNING

A Thesis submitted

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN
COMPUTER SCIENCE AND ENGINEERING

By

Mr. ANANDHAN.K
Registration Number – 18SCSE3010014

Supervisor

Dr. AJAY SHANKER SINGH
Professor
SCSE

GALGOTIAS UNIVERSITY
UTTAR PRADESH, INDIA

May 2023

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis, entitled "**AN EFFECTIVE SEGMENTATION AND LIMO CLASSIFICATION FOR PADDY DISEASE DETECTION USING DEEP LEARNING**" in fulfillment of the requirements for the award of the degree of Doctor of Philosophy in the Faculty of Computer Science and Engineering and submitted in Galgotias University, Greater Noida is an authentic record of my own work carried out during a period from Aug 2018 to May 2023 under the supervision of **Dr. AJAY SHANKER SINGH**.

The matter embodied in this thesis has not been submitted by me for the award of any other degree of this or any other University/Institute.

Mr. ANANDHAN.K

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

Dr. AJAY SHANKER SINGH
Supervisor
School of Computing Science and Engineering

The Ph.D. Viva-Voice examination of ______________________________ Research Scholar has been held on __________________.

Sign of Supervisor   Sign of External Examiner

ABSTRACT

Rice crop disease detection and its diagnosis methods are vitally important for the agriculture field to be sustainable. For that many researchers finding solutions to minimize or avoid the rice plant disease to take the best yield for formers. Because this disease led to a more than 38% yearly drop in paddy production. Due to a lack of awareness and digital knowledge in fast identifying and best remedy for rice crop diseases. In that, automated and artificial intelligence (AI) based rice crop disease detection and prevention method is a key research solution needed for the current agriculture field. The internet of things (IoT), has plenty of opportunities and contributing a vital role in wireless networks, especially in the last 15 years. Using IoT in the agriculture industry is growing up rapidly as it receives complex contextual information about water irrigation, crop disease detection, fertilizer utilization, and soil rate. Various crop disease detection methods need more accuracy and dimensionality corrections. Disease detection is indispensable for agriculture to be maintainable. Meantime automated rice plant disease detection systems also face various problems to detect diseases in the current situation. Regular machine learning-based image-wise disease detection methods are following preprocessing input values, necessary feature extraction, image segmentation, and disease classifications steps.

The proposed research work provides solutions for the above-mentioned problems and requirements with a novel approach which is the combination of the laurent series with intelligent multidimensional object optimization (LIMO classification framework) based on generative adversarial network (GAN) and swarm intelligence optimal classification through cognitive attribute selection (SIOC-CAS) to recognize various types of crop diseases in an agricultural field. Through this proposed research work IoT nodes are senses the values of a field, and crop, and the gathered information is shared with the processing unit with base station communication. Multi-objective and cognitive learning-based routing (MOCLEAR) protocol supports choosing optimal paths for data transmission betterment. After MOCLEAR protocol communication, receives crop data such as image then data reduction and dimensional sets creation to improve the eminence of the input values for segmentation and classification. Then, image segmentation using GAN combined with cognitive residual convolution network (CRCNet) is modified to segment values from input images. After receiving segment input images perform under feature

extraction and image classification with significant attributes. Generally, a corpus of attributes is extracted to better characterize a pattern recognition problem. Attribute selection plays a vital role in image classification to the complexity of handling many features and increases the possibility of prediction rate. From a broader perspective, this stage greatly influences the recognition process, in terms of both predictive performance and the design of a computationally simple classifier.

The process of attribute selection involves two steps such as attribute subset search and subset evaluation. The problem of attribute selection demands a Hamming search space with every possible combination of attributes as points in it. Introducing a new framework SIOC-CAS is an attribute selection-based segmentation with heterogeneous multimodalities data fusion (HMDF). SIOC-CAS model is a combination of flower pollination optimization (FPO) and particle swarm optimization (PSO). HMDF model collects numerical feature values through sensor values from temperature, location, image, and sound. Collected values combined with visual image features finally fusion with concatenated and dense feature vector layers for better accuracy. The predictive performance of classifiers with selected attributes is analyzed to validate the generality of proposed methodologies as a pre-processing technique for classification. Finally, the three proposed cognitive approaches for attribute selection and SIOC-CAS segmentation along with one-to-all random population initialization are applied to a pattern recognition problem in the field of agriculture. An engineering approach to solving the problem of disease identification in rice leaves is drawn and the results are analyzed.

Disease spots are segmented from the rice leaf to extract color and shape attributes characterizing it. As a pre-processing step, highly predictive attributes selected by the proposed cognitive approaches are fed as input to the classifier for disease recognition. Wrapper-based attribute selection, one of its types, employs a classifier to evaluate the attribute subset obtained during the process of search in the problem space. Support vector machine (SVM) is employed in this research work to evaluate the predictive performance of attribute subsets chosen during the process of attribute selection. Researchers have explored the use of metaheuristic randomized algorithms to find an optimal attribute subset from the problem space than an exhaustive evaluation of all possible combinations of attributes. Particle swarm optimization (PSO), an efficient search technique, has been extensively explored by researchers for attribute selection. Similarly, flower pollination algorithm (FPA), with its global walk capability explores the problem space efficiently in search of an optimal solution. Both PSO and FPA are employed in this research work as search techniques to implement the wrapper approach for attribute selection.

The problem of attribute selection demands a hamming search space with every possible combination of attributes as points in it. Both the optimization algorithms are population-based and the individuals in the population traverse the problem space in search of an optimal attribute subset. Owing to hamming space, there is no problem of divergence. Nevertheless, if the leader of the population such as individual with the best solution vector doesn't improve itself, during the process of searching for an optimal attribute subset, the problem of pre-mature convergence occurs, resulting in a sub-optimal solution. In such a situation, a re-initialization strategy is necessary to restructure the leader and re-direct the individuals in the population to another position in the problem space to continue the search thereon. From the literature, it is understood that researchers have solved the problem of pre-mature convergence by re-directing the leader to a random position in the problem space. In this thesis, heuristic-based re-initialization strategies are proposed to handle the problem of premature convergence. In PSO, if the leader of the population remains unchanged for a pre-fixed number of runs, the re-initialization strategy, named Greedy reset, re-directs the leader of the population to another position in the problem space with one attribute less compared to the recently stuck local optimum solution. Pre-mature convergence to a sub-optimal point in the problem space also occurs, when the velocity of individuals in the population stagnates.

Localized random mistakes are introduced to overcome the problem of velocity stagnation in PSO. PSO with Greedy reset and localized random mistakes is used as the subset identifier and SVM as the subset evaluator to implement wrapper-based attribute selection methodology. In FPA, a re-initialization mechanism named Greedy crossover is introduced to find a new position in the problem space on pre-mature convergence to a sub-optimal solution in the problem space. The lack of explicit crossover in FPA is the drive to design Greedy crossover on the stagnation of the leader in the problem space. FPA with Greedy crossover and SVM are engaged to design the wrapper-based attribute selection methodology. Owing to partial optimization in PSO, a re-initialization strategy named Greedy leap is designed such that, the stagnated leader of the population is reset to another point in the problem space using the concept of flower pollination. Levy's walk on global pollination between the leader and the individuals in the population is the underlying principle of the Greedy leap mechanism. PSO with metaheuristic-inspired Greedy leap strategy and SVM are employed to develop the wrapper approach for attribute selection. Both PSO and FPA are population-based metaheuristic search techniques. Population initialization is the first step in the search process. It is an important task that aids the individuals in the population to cover the entire problem space and begin the search process from all directions.

A One to all random population initialization technique is devised such that the individuals of the population represent every dimension of the problem space. Improvement in predictive accuracy on applying attribute selection and reduction in the dimensionality of attribute space is calculated as the performance indices with proposed optimized algorithms. Also, attribute selection techniques using other metaheuristic algorithms as search strategy is reviewed and compared with the proposed methodologies. Finally, attributes selected using proposed methodologies are fed as input to six traditional classifiers. The predictive performance of classifiers with selected attributes is analyzed to validate the generality of proposed methodologies as a pre-processing technique for classification. Finally, the three proposed wrapper approaches for attribute selection along with One-to-all random population initialization applied to a pattern recognition problem in the field of agriculture. An engineering approach to solving the problem of disease identification in rice leaves is drawn and the results are analyzed.

Disease spots are segmented from the rice leaf to extract color and shape attributes characterizing it. As a pre-processing step, highly predictive attributes selected by the proposed wrapper approaches are fed as input to the classifier for disease recognition. For texture classification approach using gray level co-occurrence matrix (GLCM) that considers overlap and edge pixel extract essential attributes are revised by the GAN network, which is trained by laurent with intelligent multidimensional object optimization (LIMO). The proposed Laurent series with IMO is newly formulated by integrating the Laurent series with intelligent multidimensional object optimization (IMO) algorithms. To increase efficiency the new and best computational time created such as SIOC+LIMO framework which is a combination of SIOC and LIMO models.

Here all preprocessing steps are controlled by the cognitive advisor. For segmentation and better feature selection used the SIOC model. Also, the LIMO model is used for intelligent classification and optimal outcomes. The proposed SIOC+LIMO based GAN network provides effective and improved performance metrics with overall precision, recall, F1 score, accuracy, sensitivity, and specificity values are 93.8%, 93.9%, 93.8%, 93.97%, 93.3%, and 93.97% respectively in evaluation with existing crop diseases detection.

ACKNOWLEDGEMENT

Working as an Assistant Professor and doing research for the degree of Ph.D. at Galgotias University was a quite magnificent and challenging experience for me. In all these years, many people directly or indirectly contributed to shaping my career. It was hardly possible for me to complete my doctoral work without the precious and invaluable support of these personalities.

I would like to extend my sincere thanks and heartfelt respect to my guide Dr. Ajay Shanker Singh, Ph.D., Professor, School of Computing Science and Engineering, for his valuable guidance, constant inspiration that enabled me in shaping this thesis in the present form.

I must owe a special debt of gratitude to Hon'ble Chancellor Mr. Suneel Galgtotia, Mr. Dhruv Galgotia, CEO, and Hon'ble Vice-Chancellor Dr. K. Mallikharjuna Babu, Galgotias University for their valuable support throughout my research work.

I express my sincere thanks to Dr. Munish Sabharwal, Dean, School of Computing Science & Engineering, for his continuous guidance and support during my research work and to all faculty of the School of Computing Science & Engineering who helped me a lot in my course of research work and all those who stood behind me.

I thank the almighty for giving me the opportunity, strength and determination to upgrade myself and for completing this work. My deepest gratitude goes to my wife Mrs. Nithya, my daughter Ms. Harshitha A.N, and my parents for their unflagging love and support throughout this research work. Finally, I heartfully thank my family members and my dear friends, who have extended their support and kindness intimately.

ANANDHAN.K

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACO	Ant Colony Optimization
ADC	Analog Digital Conversion
AG	Accuracy Gain
AI	Artificial Intelligence
ANN	Artificial Neural Network
AOMDV	Adhoc On-demand Multiple Distance Vector
BPH	Brown Plant Hopper
BS	Background Subtraction
CAGR	Compound Annual Growth Rate
C-GAN	Conditional Generative Adversarial Network
CNG	Chaotic Number Generator
CNN	Cognitive Neural Network
CNN	Convolution Neural Network
DBN	Deep Brief Network
DGCU	Deep Gated Current Unit
DIP	Digital Image Processing
DL	Deep Learning
DNN	Deep Neural Network
DR	Dimensionality Reduction
DRN	Dilated Residual Network
DSN	Deep Stacking Network
FAO	Food and Agricultural Organization
FoM	Figure of Merit
FPO	Flower Pollination Optimization

GA	Genetic Algorithm
GAN	Generative Adversarial Network
GF	Gaussian Filtering
GLCM	Gray Level Co-occurrence Matrix
GWHB	Grey Wolf Hunting Behavior
HIS	Hyper Spectral Imaging
HMDF	Heterogeneous Multimodalities Data Fusion
HSV	Hue Saturation Value
IBEF	India Brand Equity Foundation
IMO	Intelligent Multidimensional Object Optimization
IoT	Internet of Things
IRRI	Indian Rice Research Institute
KNN	Knowledge Nearest Neighbour
LGP	Local Gradient Pattern
LIMO	Laurent series with Intelligent Multidimensional Object Optimization
LR	Logistic Regression
MA	Machine Automation
MCFN	Multi-Context Fusion Network
MCLD	Multi-Crops Leaf Disease
MDFC ResNet	Multi-Dimensional Features Compensation ResNet
ML	Machine Learning
MLP	Multi-Layer Perceptron
MOCLEAR	Multi-Objective and Cognitive Learning based Routing

NB	Naive Bayes
NNBP	Noisy Net Back propagation
NP	Non-deterministic Polynomial
OPF	Optimum-Path Forest
PDbE	Probabilistic Distribution-based Entropy
PRNG	Pseudo Random Number Generator
PSO	Particle Swarm Optimization
RCNN	Recurrent Convolutional Neural Network
RF	Random Forest
RL	Reinforcement Learning
RSB	Rice Stem Borer
SAE	Stacked Auto Encoder
SBS	Sequential Backward Search
SCARNN	Sine Cosine Algorithm with RNN
SFFS	Sequential Floating Forward Search
SFS	Sequential Forward Search
SIOC-CAS	Swarm Intelligence Optimal Classification through Cognitive Attribute Selection
SMOTE	Synthetic Minority Over-Sampling Technique
SVM	Support Vector Machine
TLB	Tea Leaf Blight
VGG	Visual Geometry Group
WF	Weiner Filter

LIST OF PUBLICATIONS

INTERNATIONAL JOURNALS

- Anandhan.K, Ajay Shanker Singh, and K.Thirunavukkarasu, "Enhanced Hybrid Neural Networks (CoAtNet) for Paddy Crops Disease Detection and Classification", Revue d'Intelligence Artificielle (RIA), Lavoisier, Vol.36, No.5, Oct.2022, pp.671-679, ISSN: 0992-499X (print); 1958-5748 (online), Canada. https://doi.org/10.18280/ria.360503 (Scopus)

- Anandhan.K, Ajay Shanker Singh, and K.Thirunavukkarasu, "An Automatic Rice Plant Disease Detection Model Built With Unstructured Data Using IMDT Tiling and CNN Cognitive Object Detection", International Journal on Recent and Innovation Trends in Computing and Communication, Auricle Global Society of Education and Research, Vol. 10, No. 12, Dec. 2022, pp. 65-75, ISSN: 2321-8169 (Online). https://doi.org/10.17762/ijritcc.v10i12.5887 (Scopus)
- Anandhan.K, Ajay Shanker Singh, and K.Thirunavukkarasu, "An Effective Scalable LIMO Classification and MOCLEAR Segmentation algorithms for Rice Plant Disease Detection System", Neural Computing and Applications, ISSN: 0941-0643, E-ISSN: 1433-3058, Springer. (SCI-Accepted)

- Anandhan.K, Ajay Shanker Singh, and Munish Sabharwal, "Rice Plant Disease Detection using Swarm Intelligence based Optimized Classification with cognitive attribute selection", International Journal of Systems Science (IJSS), ISSN:0020-7721, E-ISSN:1464-5319, Taylor & Francis. (SCI-Submitted)

- Anandhan.K, Dr.Ajay Shanker Singh, and K.Thirunavukkarasu, "Hybrid Neural Networks for Paddy Crops Disease Detection", International Journal of Advanced Science and Technology, Vol. 29, No. 8, pp.6521-6538. http://sersc.org/journals/index.php/IJAST/article/view/37956. (Scopus)

INTERNATIONAL CONFERENCES

- Anandhan.K, Dr.Ajay Shanker Singh, and K. Thirunavukkarasu, "Comprehensive Study: Machine Learning & Deep Learning Algorithms for Paddy Crops," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE, ISBN: 978-1-6654-1703-7, 2021, pp. 1-7. doi: 10.1109/ICRITO51393.2021.9596189. (IEEE)

- Anandhan.K and Dr.Ajay Shanker Singh, "Detection of Paddy Crops Diseases and Early Diagnosis Using Faster Regional Convolutional Neural Networks," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), IEEE, 2021, pp. 898-902. doi: 10.1109/ICACITE51222.2021.9404759. (IEEE)

PATENT FILED

- Anandhan.K and Dr.Ajay Shanker Singh, Invention Title: "Plant Disease Identification System Through Implementing Image Processing", Type: Indian Patent, Status: Published, Application Filling Date: 27/02/2022, Publication Date: 11/03/2022, Application Number: 202211010476.
  Source Link:
  https://ipindiaservices.gov.in/PublicSearch/PublicationSearch/PatentDetails

CHAPTER 1

INTRODUCTION

In this chapter, we will discuss major plant diseases, especially rice plant diseases, as well as their categories along with their symptoms, disease types, and affected ranges and issues around them. The chapter also focused on how image processing is an effective tool with a convolutional neural network (CNN) for the identification of rice diseases. Apart from CNN with image processing applications of deep learning with image processing are also discussed.

1.1 Rational

This research work has adopted advanced deep learning as a tool for the identification of rice diseases using images. This research work is focused on designing a hybrid deep-learning model to increase the efficiency of the diagnosis process.

1.2 Overview

Major financial losses in the global agricultural sector are caused by plant diseases. To stop the spread of disease and enable efficient management techniques, it is crucial to monitor plant growth and find pathogens early. Rice production has recorded impressive growth from 20.58 million tons in 1950-1951 to 104.86 million tons in 2014-2015, which is almost 5 times. China managed to produce over 148 million tons of milled rice in the harvest time year of 2020–2021, more than almost any other nation. According to figure 1.1, India placed second that year with 122 tons of metric milled rice. Global rice production has increased steadily but at a slower pace from 400 million to 477 million tons over the past 15 years.

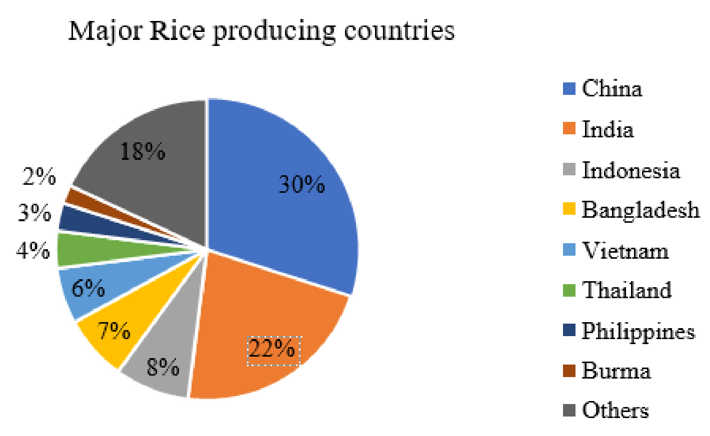


Figure 1.1 Major rice-producing countries (%)

1.3 Plant Diseases

A plant frequently contracts an infection when a causal agent continuously disturbs it, causing abnormal body mechanisms to interfere with the normal formation, growth, feature, or other functional areas of the plant [1]. As per their root issue, whether infectious as well as non-infectious, plant diseases can be broadly categorized. Infectious agents can procreate both inside and outside of their hosts, and they can also be distributed from one host to another. Poor growing conditions, such as extreme temperatures, poor oxygen as well as moisture relationships, and necessary minerals, are the root cause of non-communicable plant diseases. Non-infectious causes are not transmissible because they are not living things that multiply inside of their hosts. When they affect a significant crop, plant diseases can be harmful. A lot of plant germs are more sensitive to a specific kind of plant or group of plants. Occasionally, a plant pathogen may emit toxins that are dangerous to people. Similar to the production of livestock, crops are frequently grown in spots that are more concentrated over a large area of land. They become susceptible to pathogen risk as a result. Whole grains are a good example of this. The majority of the daily calories consumed worldwide are produced by the crops of rice and wheat. The economic system and the population's nutritional intake could be harmed by any pathogen that targets these two plants. One or more of the fundamental conditions for a healthy, safe human life are impacted by plant diseases because they interfere with the existence, adequate growth, and production of all kinds of plants. This has been the case ever since humans stopped depending on the fruit and native animals and instead stopped, took control of, and began farming more than 6000 years ago.

Food and feed crops have historically been regularly destroyed. Malnourishment, famine, migration, and the deaths of people and animals, many of which are well-documented in history, has caused it all. Similar outcomes are observed yearly in the growth of agricultural communities, which provide a living for families and nations [2]. Loss of food consumption and production contributes to economic losses and higher prices in highly developed societies. However, it must be kept in mind that any loss of food or food caused by plant diseases means that it is little accessible in the international economy. Rich people and wealthy nations will be able to purchase such food wherever that is available due to the inadequate prices and supply of food, whereas many poor people will suffer as a result of this loss and go hungry. People decide what germs and pathogen species will rule by choosing which plants to resist. Humans influence the amount of first and second inoculums that are available to target plants through their cultural practices, as well as the chemical and biological controls they can use. By delaying or accelerating planting or harvesting, planting in brought-up beds or open areas, covering plant areas with chemical compounds before it rains, managing moisture in storage areas, and other practices, they also change the natural effect on disease development. The timing of human activities related to plant growth and protection may inadvertently influence different combinations of these elements to some extent, having a significant impact on both the affected rate in a single plant and the total affected plants. Although the human half of the tetrahedron disease is occasionally substituted for the time portion, it should be viewed as a distinct fifth that both intrinsically & extrinsically influences the growth of rice plant diseases and its affected level [3].

1.4 Plant Diseases Fundamental

Plant disease research is important because it causes the loss of the plant. Plant science basic has three main purposes:

- Learn about the living, non-living and natural causes of plant diseases.
- Study methods for developing bacterial infections.
- Study the interactions between plants and pathogens.

1.4.1 Various rice plant diseases

The majority of the nation is affected by this paddy disease, particularly West Bengal, Uttar Pradesh, Orissa, Andhra Pradesh, Tamil Nadu, Punjab, Bihar, Chhattisgarh, etc. In Tamil Nadu, this illness is very prevalent. It is a disease spread by seeds. One of the most common rice diseases is bacterial blight such as a viral disease. Millions of hectares of rice are contaminated each year, and serious epidemics can result in crop losses of up to 75%. These diseases are widespread in rice paddies in California and can occur wherever rice is soaked in water. Seed rot and seed disease often lead to the formation of similar stalls. Signs of seed rot and seed disease appear shortly after seeding [4]. Table 1.1 illustrated some common rice diseases.

Table 1.1 Common rice plant diseases, pathogen, symptoms, and sample images

Disease	Pathogen	Symptoms	Sample Image
Rice Blast	Fungi	● Lesions were found on parts of the plant.	

Rice bacterial Blight	Bacteria	● Yellow is hundulated marginal ● Drying and curling of leaves	
Brown spots	Fungi	● Small, circular, yellow-brown lesions	
False Smut	Fungi	● The appearance of velvet / yellow spore ● Reduction in grain weight ● Reduced seed germination	
Node Blast or Neck blast	Fungi	● Blackish or brownish lesions on nodes ● Poor quality of grain.	
Paddy stemborer	-	● Presence of egg mass (brown colored) near the tip of a leaf. ● The presence of a caterpillar bore in the central portion of the shoot.	
Rice tungro virus (RTSV, RTBV)	-	● Partially filled grains ● Leaves become yellow or orange-yellow	

BLB	Bacteria	● Yellow is hundulate marginal necrosis ● Drying and curling of leaves	
BPH	-	● Oval-shaped lodging and dry spots on the bottom of the plant. ● Hopper burns appearance	
Hispa	-	● The mining of the grubs will be seen on the leaves. ● Generally, the plants are affected in the young stage	
Leaf Blast	-	● The symptoms are narrow reddish-brown wide bands. ● Sometimes the lesions are at the edges	
Sheath Blight	Fungus	● Filled with empty grain, similar to the bottom panicles.	

1.4.2 Rice diseases and their symptoms

Numerous factors, including bacteria, insect pests, and a rare natural condition, can lead to paddy diseases. Plant parts can be harmed above or below the soil surface by parasitic, adenoviruses, viral or nematode plant germs. This section describes paddy diseases according to their symptoms. The reader should comprehend the kind of image analysis and the element required to diagnose the illness [5].

Rice blast: In south Gujarat and middle Gujarat, rice blasting has been a serious disease for ten years. At whatever stage of plant development, from sprouting to harvest, the pathogen targets all aerial sections of plants. The illness manifests as grain spots, node explosion, neck rash, seed decay, and leaf blight. Large spindle-shaped lesions with grayish-gray areas as well as brown genes that considerably inhibit plant growth and seeding are produced during leaf growth. The neck tissue or infected node decayed and lightened. A node or neck rupture, which strikes pre- or post-flowering as well as grain development and causes a major reduction in grain level and product value, is regarded as an active phase of the infection. A dark brown spot was caused by grain infection.

Rice bacterial blight: Enlarged lesions on the leaf tip that are many inches in length and turn yellow from white as a result of bacterial action are among the symptomatology.

Brown Spot: The rice plant's leaves develop this disease. Circular to elliptic lesions having dark brown spots is the disease's common symptom.

False smut: A silvery white patch on leaves of rice that later turns into orange dust is termed false smut. The disease cycle persists late into the planting season and is to blame for the current reduction in direct yield. False smut is caused in the area with rainfall and strong humidity; extremely nitrogen-rich land.

Neck blast: It happens in areas with low soil moisture, frequent and prolonged rainfall, and cooler temperatures during the day. In northern rice, a large alteration between day and night temperatures is foremost to the development of leaf dew, and the overall cool temperatures favor the disease. Ice can explode at all stages of growth. However, leaf inflammation tends to decrease as plants mature and grow resilient to adult plant diseases.

Paddy stemborer: Stem borer can damage rice at any stage of the crop from spring to ripening. They eat tillers and cause dead hearts or dry out of the central tiller, during plant growth; and cause whiteheads in the reproductive stage. Rice stem borer (RSB) can be found in all rice-growing areas. At the time of planting, the most common sign of damage is a dead heart, while in the flowering stage, it causes a white head. The larvae also produce small holes in the stem & insert feces into it, which is visible when the stem opens.

Tungro: It results in sterile or partially filled grains, stunted growth, diminished tiller numbers, and discolored leaves. Tungro spreads to other grassy weeds typically occurring in rice paddies, wild rice relatives, and cultivated rice.

BLB: Among the most devastating diseases to affect cultivated rice is bacterial leaf blight, also known as the bacterial blight of rice. A million hectares of rice are contaminated every year, and serious epidemics can result in crop losses of up to 75%.

BPH: A life form of planthopper that feeds on rice plants is the brown planthopper.

Hispa: In addition, the existence of grassy weeds in and around rice fields and other ports promotes the growth of insects. Damage is also a result of the extremely fertile field. Heavy rains, particularly during the rainy season or the first rainy season, are accompanied by unusually little rain, a short period with little variation in the daylight hours temperature, and high relative humidity, which results in a large number of insects, and the rainy season makes it more plentiful.

Leaf blast: The most harmful of all fungal infections in Southern Africa, blast disease, also recognized as rice rotten neck, is brought on by the fungus.

Sheath blight: The fungus rhizoctonia solani kuhn, which afflicts rice grass, is the reason for the rice plant pathogens known as sheath blight. Sclerotia that land on the grass before the actual harvesting period causes this disease. The leaf blades will become sore. On the planet, there are several types of plant illnesses.

These diseases may result in a decline in the quality of agricultural products and significant loss of returns and even threaten food safety. Prompt detection and diagnosis of herbal diseases are the foundation for disease control steps. Plant diseases are normally recognized and diagnosed by agricultural technicians based on visual recognition in-field. This method requires a high level of professional competence and extensive expertise and requires numerous staff. Pathogen diagnosis by disease detection requires more technical expertise and satisfactory laboratory conditions. Pathogens were quickly developed based on molecular biological techniques and the diagnostic results of these approaches could be obtained. These methods should, however, be carried out by technical and professional staff and cannot be carried out in the field. In addition, these approaches are temporary and costly. An easy and fast method of identification of plants with high identifying accuracy is therefore very important. The acquisition, processing, and transmission of information on plants have been significantly influenced by data technology with biocomputing. The main factor arising from the progress of visual technology and the popularity of digital goods is now recognition of plant diseases. Studies have been published on the identification and automatic evaluation of image-built plant conditions [6-10]. Automated computer-based identification and diagnosis based on plant disease symptoms may provide agricultural technicians and farmers with rapid and accurate disease information, thereby reducing dependency on farm technicians.

1.4.3 Acceptable crop growth situation/environment

- Excess nitrogen aggravates the disease severity.
- In the eastern and southern regions of the country, moderate temperatures are found to be ideal for growing rice all year round. Therefore, two or three rice crops are grown annually in the eastern and southern countries.
- More than eighty percent with a 20-32°C highly recommendable situation for rice crops.
- In the northern and western parts of the country, where rainfall is high and winter temperatures are very low, only one rice crop is grown from May to November [8].

1.4.4 Rice plant disease formation

The rice disease cycle is presented in figure 1.2.

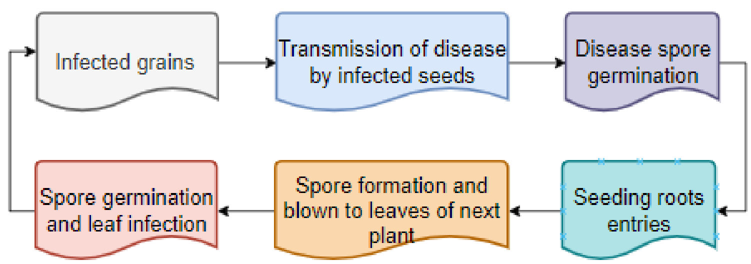


Figure 1.2 Rice plant disease cycle

Leersiahexandra as well as Echinochloacolonum are examples of collateral hosts on which the fungus can also survive. The brown spot fungus is typically found in regions where rice cultivation has a long history. Older lesions and infected debris generate airborne spores that can spread the infection.

1.5 Plant Pathology

Plant pathology or the study of diseases in plants, has been an area of interest for hundreds, if not thousands of years. Plant pathology is relevant in one way or another to nearly all living things, whether it be from direct or indirect impact. Diseases in plants can cause entire fields of crops to perish leading to food shortages. They can also result in serious health effects for both humans and animals, and cost millions of dollars to manage every year. For these reasons, the need to study diseases in plants is not unique to one particular field of study, and motivations for understanding this branch of pathology are abundant and widespread.

Like diseases in humans, plant pathology is also not a trivial field of study. Plant pathologists, or experts on plant diseases, need to understand countless variables about the diseases themselves, as well as prevention/management techniques. Plant diseases can come from several different organisms, including but not limited to fungi, bacteria, viruses, and even other parasitic plants. Oftentimes, these disease-causing organisms infect specific hosts and have complex life cycles, making it even more difficult to study and limit plant diseases.

1.6 Agricultural Land Management

Agricultural lands in India roughly 78% of people are directly or indirectly linked to agronomy. The economy of India is heavily dependent on the agricultural sector and most of the Indian economy is in this sector. The cultivation of crops and fruits for many years is changing our financial status. Many plants are planted across the country in India and rice, wheat, and potato have affected the high level of popularity amongst all crops and fruits. Maize, peach, grape, and strawberry plantation is showing a growing trend, which is quickly strengthened by the invitation of this crop. Thus, farmers' interest in growing such crops and fruits is higher than in the previous decline. Fish is very common and highly tested for vitamin A products. In the hemisphere of the northern and southern regions, fish grows at colder temperatures. Peach has been invented in China for the first time, since spreading to Asia, Europe, Spain, Mexico, and the US. The fruit tree size is very short of peach. The peach farmer has suffered various diseases that have decreased the production rate of the peach and caused enormous losses. Several diseases are present, such as pseudomonas syringae, crown gall (agrobacteiumspp), scab (cladosporium carpophlum), bacterial spot (xanthomonas campestris), brown red (moniline fructicola), and rust (tranzschelia discolour) illness.

The blades on the bottom of the tree turn violet in the center during the bacterial spot. The surface of the fruit introduces tiny, circular spots with green color when faced with Scab disease and its size gradually increases to a dark yellow halo. The skin and tissue of fruit lose color during brown red disease. The top and bottom of the peach are yellow-green in angular form. The new leaf is red and the leaf is unregulated and raised as leaf curl disease. It becomes red with a yellow hue. Plant diseases come in a range of different forms, originating from a wide variety of parasitic organisms and conditions, each of which displays many different defining symptoms of the disease. For these reasons, it is important to classify diseases by many factors to help with their management. The majority of diseases are caused by parasitic organisms, and the main types of organisms that can cause diseases in plants consist of fungi, oomycetes, bacteria, phytoplasmas, viruses, viroids, nematodes, and other parasitic plants. Fungi are living organisms that spread via spores and often reproduce both sexually and asexually. They have several different sub-species each with its specific characteristics, including a wide variety of complex life cycles and structures that help them reproduce and survive. Oomycetes are organisms that are very similar to fungi in terms of structure and reproduction, though are entirely separate in the evolutionary tree. Bacteria on the other hand are microscopic organisms, single-celled, that are also capable of causing disease.

Viruses and viroids, differing from bacteria, rely entirely on hosts for reproduction. Nematodes, also known as roundworms, are a small type of animal that often survives as parasites in plants and other animals. Last but not least, plants themselves can be parasites of other plants, obtaining part if not all of their nutrients from their plant hosts [11]. With this broad range of disease-causing organisms and a much wider variety of species, plant pathologists have a significant challenge in being able to classify infections initially from observation. To do this, plant pathology experts use what are known as signs and symptoms to assist in uniquely identifying diseases. A symptom of a disease is a visible identifier on the host which is reacting to some pathogen. These can include spots on leaves, stunted growth, plant wilt, rotting, and discoloration, as well as many other reactions.

Figure 1.3 A rice leaf infected with a fungus (anthracnose), displays lesions as a symptom of the disease.

On the other hand, a sign is a direct observation of the pathogen itself, as opposed to the reaction caused by the pathogen. This can include observing bacterial colonies under a microscope, spotting fungus on a leaf figure 1.3, or finding roundworms in crops. Frequently, symptoms are much more obvious than signs, while signs help to more accurately classify the specific pathogen. Oftentimes in the field, however, a pathologist may use a symptom to identify a disease before confirming the presence of signs in a laboratory. Various papers surveyed in this thesis will cover a wide range of plant disease types, with drastically different symptoms. The data set gathered and modeled for the latter portion of this paper will specifically look at strawberries and common diseases that impact strawberry growth every year. The focus of this thesis will not go into the intricacies of various disease types but instead look at the possibility of disease detection through computational models. Nonetheless, the variability discussed in plant diseases is still important to recognize, as different disease identifiers are important for computational model detection [12].

1.7 Impact of Plant Diseases

Often overlooked, plant diseases are a severe economic and health concern across the globe, making detection and prevention an incredibly important area of study. Just like epidemics exist among us as humans, every year disease outbreaks result in significant losses in crops worldwide. Plant diseases often account for approximately 20% to 40% of crops being lost each growing season depending on the infection and product. Fusarium Head Blight, a specific strain of fungus that infects wheat species, accounted for approximately 870 million dollars in crop losses alone between the years 1998 and 2000. As a whole, crop losses due to disease cost the agricultural economy billions of dollars every single year.

Furthermore, the impact of plant diseases goes far beyond economic losses. While plant diseases in developed countries can be fairly well managed at a price, often developing countries lack the money and resources to take preventative measures against initial infection, or effectively separate healthy plants from the diseased after infection is apparent. One such pathogen that is a variant of the common disease wheat stem rust is currently a significant problem in countries within Africa and the middle East. Other diseases can be less noticeable but just as dangerous, leading to serious health effects in humans when consumed [13].

1.7.1 Impacts of pesticide use on the environment

It is well-documented that chemical pesticides have damaged many beneficial insect species and soil microbes shown in figure 1.4. These chemical fertilizers and pesticides are responsible for causing health hazards among people and livestock.

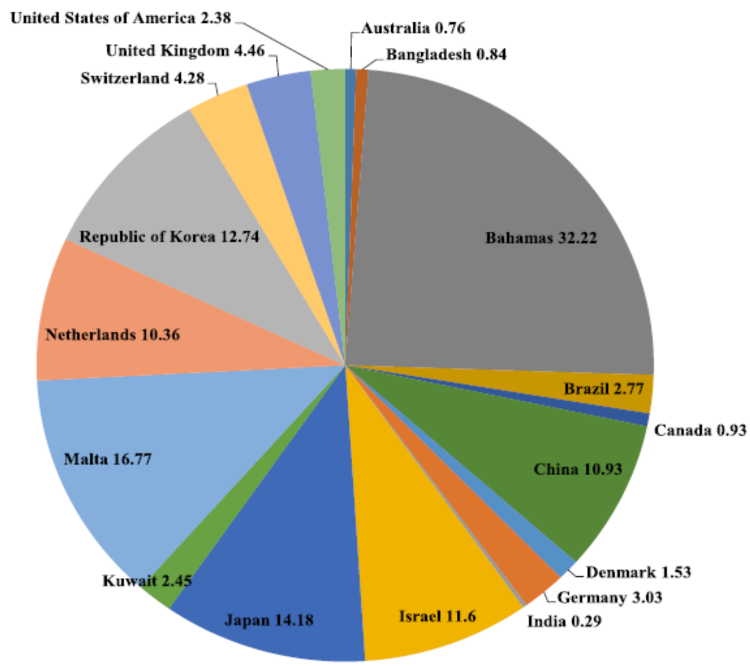


Figure 1.4 Consumption of pesticides in the world (Tones)

Organic farming systems or eco-friendly agriculture are based on specific standards precisely formulated for food production and aimed at achieving agroecosystems, which are socially and ecologically sustainable. According to existing research, sustainability refers to the substitution of biological inputs and knowledge, aimed at minimizing the cost of production without affecting the yields.

1.7.2 Threshold limit of chemical pesticides

Present information says that the complex interactions of agrochemical and various field management practices may interfere with the buildup of predators, parasites, and soil-beneficial microbial populations. If pesticides are used in an uncontrolled way, then the ecological balance between pests and natural enemies would be hampered. The activity of natural enemies in the ecosystem implies their useful role as bio-control agents for the protection of ecological stability and as biological indicators of the health of the agroecosystem. Therefore, restricted use of pesticides and organic farming is emphasized so that the indigenous predators, parasites, and pathogens that exist in the ecosystem could be preserved for sustainable crop protection and also provide a solution for their better use under organic farming or integrated pest management ensuring a healthier pesticide-free product in India [14].

Being an organic state, Tamilnadu, Department of agriculture and horticulture provides many biopesticides such as trichoderma harzianum, pseudomonas fluorescence, Trichoderma hamatum, trichoderma viride, trichoderma virens, bacillus subtilis, beauveria bassiana, etc. for disease and insect management. Therefore, environmentally safe and eco-friendly methods of disease management practices are needed in Uttarakhand's agriculture and horticulture. In this context, biological control is an alternate strategy for disease management in different crops. Biological control agents for sustainable agriculture are now being used in many developed countries for combating diseases to improve quality food production.

1.7.3 Reasons for limited use of bio-control agents

Although bio-control agents are eco-friendly bio-fungicides they do not reach every farmer because they may not take action as quickly as chemicals, since their populations take time to multiply. In most cases, they have a narrower target range. Most of the bio-agents are not broad-spectrum products. It may have a shorter shelf life if not stored properly. The availability of bio-agents is limited compared to chemical pesticides in the local market, while chemical pesticides are easily available and bio-control agents tend to be more difficult to use compared to chemicals [15].

1.8 Introduction Rice Plant Disease Computational Control and Detection Techniques

1.8.1 Image processing

IP is a technique for performing operational processes on images in an attempt to improve them or retrieve relevant important data from them. Identifying designs, recognizing borders, reducing distortion, quantifying elements, and generating statistics for texture assessment or image clarity are all examples of image processing activities. Image analysis is a technique for identifying patterns and characteristics in photographs.

Pattern identification is employed in variability of applications, comprising handwriting assessment, image detection, and computer-assisted diagnostics. Image processing is a form of software technology that permits users to examine, evaluate, and retrieve data from photographs. Color modification of photographs is one of the fastest-growing technologies, and it's changed dramatically over time. Image enhancement is the process of transforming a picture into a digitalized format and then executing actions on it to create a better image or retrieve valuable details. Typically, an image analysis system takes images as two-dimensional inputs that are subjected to predetermined signal processing procedures [16].

1.8.2 Various steps involved in digital image processing

DIP is a method for transforming physical image convert into digital data, which is then used to perform operations on the data to create a better image or retrieve crucial information. It is a signal distribution technique where an image, including a video frame or an image, serves as the input, and an image or an image characterizes serves as the output. Typically, an image analysis system treats images as two-dimensional signals to which signal processing procedures have previously been applied. Figure 1.5 depicts some of the common procedures involved in image processing implementations.

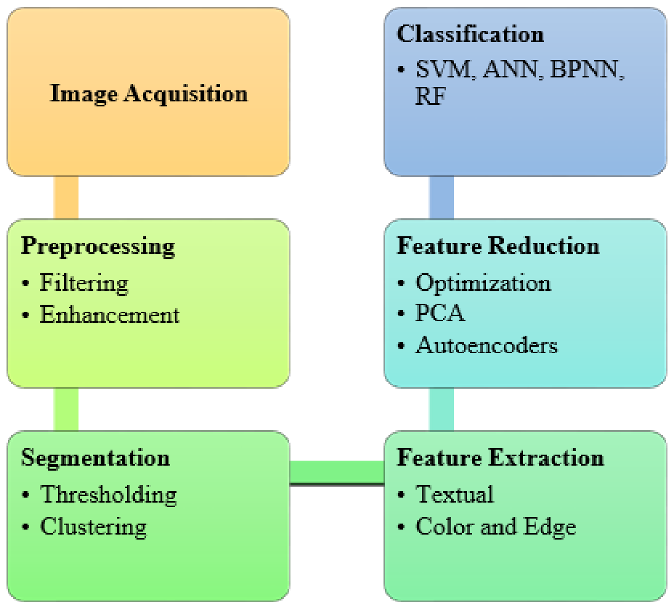


Figure 1.5 General steps for image processing

Image acquisition: Pre-processing, transformation, and resizing of the image acquiring phase. A dataset of images dedicated to rice disease information. Images from the farms are recorded utilizing a digital camera and converted to digital format so that digital image analysis processes can be conducted.

Image pre-processing: Image pre-processing involves correcting image data captured by cameras that contain defects in geometrical parameters and pixel intensity values. The following are typically common image pre-processing stages.

- Contrast stretching: Image color-based stretching method implemented for context and its regularity. Different stretching models are used to extend the image and coloring range dynamically.
- Noise filtering: Noise Filtering is a technique for removing unwanted data from an image. It's also utilized to eliminate several kinds of distortion from photos. This characteristic is primarily participative.
- Changing the histogram: The histogram is very important in image enhancements.

- Image enhancement: Basic image sharpening technique with restoration model.
- Image segmentation: When it comes to image fragmentation, it's not only just about separating things from the backdrop; it can also be about separating distinct areas. Watershed segmentation is one technique for separating such areas. Methods for segmenting an image divide it into its component portions or elements. In general, automated image fragmentation is complex processing in DIP. A robust separation approach helps the operation get a lot closer to solving photographic challenges that need single-item identification [17]. In-plant disease identification, image segmentation can be quite useful. The term image segmentation refers to the process of partitioning the input image into small segments based on the required attributes. The main objective of fragmentation is evaluating image data to retrieve pertinent characteristics. The image segmentation process can execute in two ways: first depending on inconsistencies and second depending on resemblance. The first method involves splitting an image depending on abrupt variations in intensity values, such as by border recognition. In the second approach, images are segregated depending on predetermined criteria, such as thresholding utilizing Otsu's technique.
- Feature extraction: To retrieve characteristics from synthetic aperture radar images, retrieval of characteristics approaches has been established. Color, geometry, and texture are the three domains from which characteristics are retrieved. The hue is significant as it can distinguish one disorder from another. Moreover, every sickness may have a unique geometry, allowing the system to distinguish between disorders based on shape characteristics. Area, axis, and angle are a few geometrical characteristics.
- Image classification: Finds the values of disease and classifies with various classifiers and provides optimal solutions.

1.8.3 Image processing techniques

Image processing techniques can be utilized to improve farming activities by increasing procedure precision and uniformity while minimizing the amount of human supervision required by agricultural producers. It frequently provides versatility and efficiently replaces the peasant’s ability to make perceptual decisions. The image analysis terms used in agricultural applications are illustrated in figure 1.6.

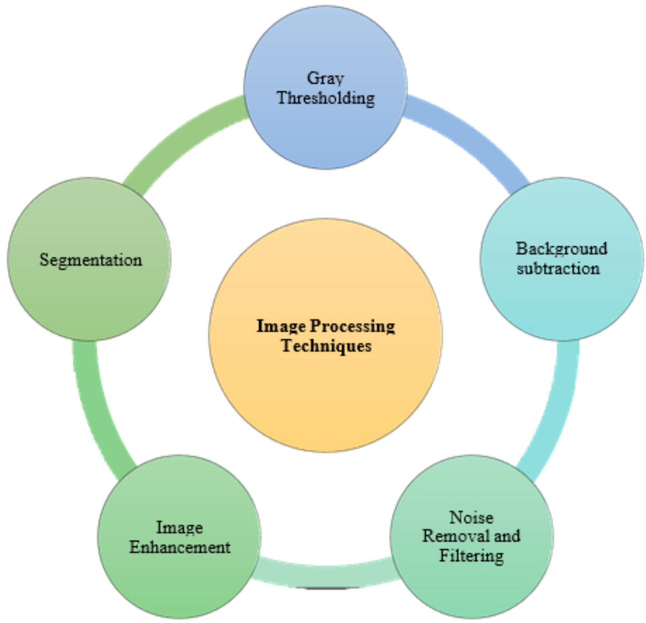


Figure 1.6 Image processing techniques

Gray thresholding: Grey level thresholding is a basic search table that divides gray levels in an image into one of two groups: those below and those over a user-defined threshold. Thresholding is just one of several ways to make a binary mask for an image. These masks are utilized to limit future processing to a specific area of an image. This process divides an input image into two lessons: one for pixels with values below an analyst-defined gray level, and another for pixels with values beyond this standard [18].

The effective techniques used were global gray-level thresholding for image segmentation. When used with pre-processing techniques like background illumination corrections and top hat filtration, in which the object and background classes are separated in grayscale, it is highly effective. The fundamental operation of the grey-level threshold is shown in figure 1.7. Grey-level thresholding is used to identify the pixels in an image that best represent an item using feature extraction methods.

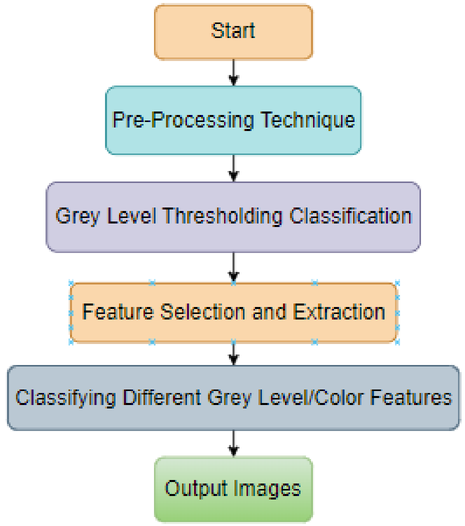


Figure 1.7 Block diagram of grey thresholding process [19]

Background subtraction: Background subtraction is a frequently utilized approach for utilizing stationary camera systems to generate a frontal overlay. As the title implies, BS computes the frontal mask by subtracting the existing frame from a backdrop model containing the stationary component of the image or more broadly, anything that may be called a backdrop based on the apparent scene's features. A common technique for identifying moving objects in a series with still images from static cameras is background subtraction. The fundamental idea behind this method is to identify moving objects. It is commonly accomplished by identifying the foreground items in a video frame.

Background modeling consists of two main steps.

- Background initialization: The very primary stage is to measure a preliminary model of the backdrop.
- Background update: The model is upgraded in the subsequent stage, in addition to responding to any alterations in the scene.

Several implementations do not require knowledge of the entire components of the series; also, subsequent assessment is centered on a portion of the series since attention is drawn to the specific items of images in the forefront. Entity positioning is performed after all the preparatory stages have been completed, which include deionizing and morphological analysis, and this foreground identification is utilized there. All current identification algorithms are dependent on modeling the image's backdrop, that is, setting the backstory and detecting alterations. When there are structures, reflections, and dynamic things in the backdrop, identifying the right backdrop might be tricky. All of the strategies presume that the static items will change coloring and brightness over the period while establishing the backdrop.

Noise removal and filtering: Noise is created by a wide range of elements, including several exogenous circumstances in transmitting systems and atmospheric conditions, and comprises noise types such as gaussian, poisson, blurred, speckle, and salt-and-pepper noise. Noise removal methods are becoming significant aspects of medical photography uses, as well as the most widely utilized filtration are the median filter, gaussian filter, and weiner filter, which provide optimum results for related noises. This information is gathered by employing filtration and choosing the appropriate filtration & histogram, magnitude, and image resolution provided by these filters.

Median filter: In digital image analysis, the most well-known order-statistic filtration system. The Median Filter uses a preset filtering screen dimension to determine the result of neighboring pixels.

Weiner filter: Conventional filtration is constructed with certain frequency responsiveness, and the weiner filter is an excellent illustration of this method. This filtering method utilizes a distinct technique for filtration.

Image enhancement: An image improvement is something that renders it simpler or quicker to understand an image graphically. In many circumstances, like with low-pass filtering, the improved image may appear to be inferior to the actual, although this augmentation was most probably done to aid the translator in seeing low spatial frequency characteristics among many of the regular high-frequency noise present in an image. In addition, a certain implementation is enhanced. This improvement could be unsuitable for another use, necessitating the use of an alternative augmentation. Figure 1.8 displays the fundamental working diagram for image enhancement.

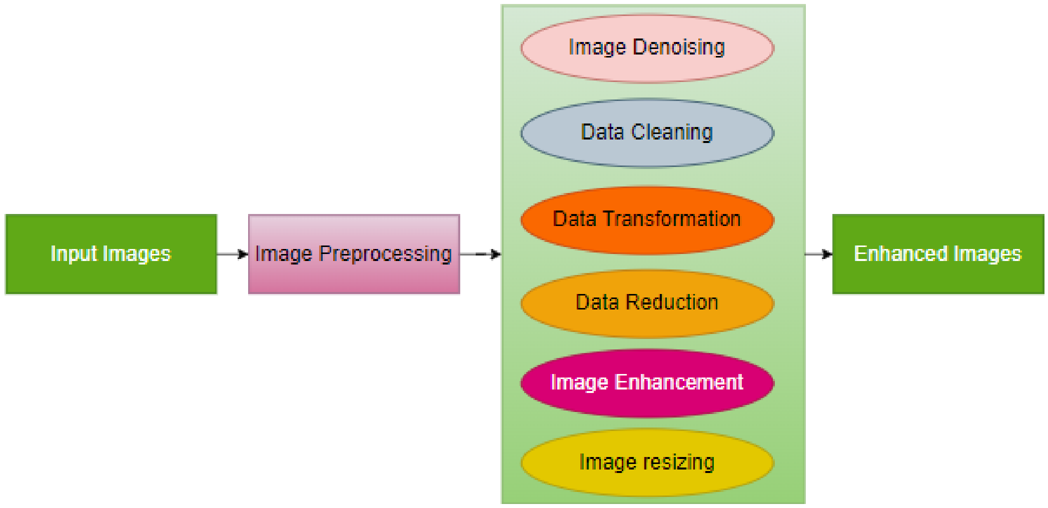


Figure 1.8 Block diagram of image enhancement process [38]

Two broad classifications can be used to categorize image enhancement methods.

- Spatial domain: Image improvement that breaks a picture into consistent pixels as per spatial coordinates with a specific quality. The techniques in the spatial area operate explicitly on pixels.
- Frequency domain: A spatial domain increase derived by using the fourier transform. Pixels are handled in clusters and also indirect means in the frequency realm.

Smooth and sharpening, noise reduction, deblur photos, contrast correction, brightening a picture, and grayscale picture histogram normalization are a few instances of picture improvement [20].

Segmentation: The image segmentation method is requirements-based segments forming on given input images numerous sections termed image segments to reduce the image complexity and variety future editing or assessment of the picture easier. In layman's terms, segmentation is the process of allocating tags to pixels. For entity identification, the picture must be delivered as input. Instead of analyzing the entire picture, the detectors might be fed a segmented section. This prevents the detectors from analyzing the entire picture, resulting in a faster implication period. The image segmentation process begins with any image being provided as input. The image is then processed by the image segmentation engine according to the specific image segmentation algorithm and divided into several image objects as shown in the image. Figure 1.9 shows various image objects are nothing more than the valuable data extracted from the image. Super pixels are the common term used to describe these image elements or image segments.

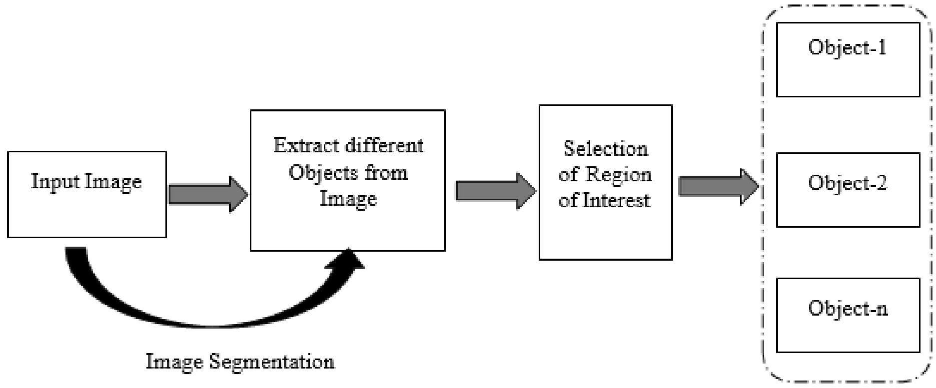


Figure 1.9 Regular image segmentation process block diagram

1.8.4 Approaches to image segmentation

- Similarity method: this approach calculates and compares the similarity pixel values in image pixels for image segmentation. Image segmentation method using machine learning (ML) and convolution neural network techniques playing major in DIP.
- Discontinuity method: This approach calculates and compares the intensity rate of the image being discontinuous with every pixel. This discontinuity method applies to all elements of an image and creates optimized segmentation.

Table 1.2 Both positive and negative aspects of image processing techniques

Technique	Positive aspects	Negative aspects
Gray thresholding	1. Low complexity. 2. Multiple criteria at the same time and give every good result with less noise. 3. Connected regions are guaranteed.	1. Sensitive to noise. 2. The computational time and resources needed to compute the new data set. 3. The new data takes up space in your memory image.
Background subtraction	1. Different thresholds are selected for each pixel. 2. This pixel threshold by time. 3. Provide fast recovery. 4. Objects can blend into the background with outer the	1. Cannot deal with sudden drastic lightning changes. 2. Initializing the Gaussian is important.

	current background model.	
Noise removal and filtering	1. Unnecessary pixel rate and image quality reduction factors identification. 2. Threshold-based conversion model.	1. Highly expensive. 2. Reducing image quality, edges, and clarity. 3. Parameter adjustment.
Image enhancement	1. It contains slow contrast and dark regions. 2. It is the most effective technique for grey-scale images.	1. It is how complex processes are in the description of other techniques. 2. Have lower accuracy.
Segmentation	1. Easy implementation and very widely used in medical Image processing. 2. Work well when the edges in an image are prominent.	1. Fails to separate the pixel accurately into the suitable region. 2. Not good for the extraction of ROIs and lead to the high false positive and false negative result.

1.9 Image Processing in Agriculture

Picture processing has also been proven to be a useful machine-sensing approach in the agricultural industry. Photography technologies with various spectrums, which include infrared, hyperspectral imaging, and remote sensing, were effective in identifying vegetative indexes, canopy measuring, field mapping, and so forth with improved precision.

Picture processing: These methods were applied to farming by gathering photos utilizing remote sensing technology utilizing airplanes or satellite systems, which would then be treated and evaluated utilizing computer systems. Image processing technologies have been established to address numerous challenges in the domains of farming as a result of newer technology breakthroughs in picture capturing and information processing. Picture analysis can help with vegetative assessment, watering, and fruit picking, among other things.

As a result, image enhancement has now been increasingly employed in agribusiness as a noninvasive method for analyzing agricultural products. This has been aided by advancements in digital picture capture equipment and software for picture modification [21]. Imaging technologies have been established to address numerous challenges in agricultural production as a result of newer technology breakthroughs in picture capturing and information analysis. In farming, image processing technologies can also be applied in the various domains.

Crop management: The crop management process is hectic as it consists of crop growth monitoring, disease detection, irrigation, nutritional measurement, etc. Therefore, for such large-scale management, image processing tools and techniques simplify the management issues. Weed detection: A significant problem in an agricultural investigation is the identification and classification of weeds. Identification of weed species for their control and the detection of harm to plants or crops require the categorization of weeds. Identification of nutrient deficiencies: Image processing techniques have been used to identify nutritional deficiencies and different plant contents from the leaves as well as the skin of products. Quality inspection: Image processing as well as ML is employed to enhance and preserve the standard of fruits and vegetables as well as for the categorization of farm commodities. Geographic information systems (GIS) and algorithms for segmenting color and texture are used in the prediction of land evaluation.

1.10 Deep Learning Model

ML as well as AI models is called deep learning to imitate how people learn particular types of information. DL is a key factor for data science technology that includes facts and predicting models. Growers may also monitor their crops from any place on the globe. This AI-assisted intelligent agricultural system is extremely effective. Plants need water to develop properly.

1.10.1 Deep learning techniques used in agriculture

A smaller database is needed for training the algorithm in machine learning, although a larger database is needed for training the algorithm. Even though learning requires less time than evaluation, low-end workstations seem to be adequate for having to process machine learning. On the other hand, it requires higher computational resources. In the field of biological sciences, images are used to convey relevant data and knowledge. In the biological and agricultural fields, DIP and image analysis technologies are critical. The classification of plant leaves and the cultivation of healthy plants are critical aspects of agricultural automation. The disease primarily infects the color and conditions of the roots, flowers, fruits, leaves, and buds of a plant. More plants are affected by this disease which results in a low yield of the crop. The technological development which relates to the processing of images has grizzled the interest of the people and has created a specific way for improving crop yield shown in figure 1.10. As we all know, agriculture has a huge impact on the economy of the country.

The number of fertilizers and pesticides used to kill bacteria on the crops also impacts the growth of the crop. Some individuals find it to be hard to spot the disease of the plant with their eyes. And if the disease is detected, they cannot find the correct treatment. If the disease is detected on the crops at an initial stage, then this might reduce heavy damage to the crops. For this, they need to identify the additional technique for the classification of the plant by new research ways. Deep learning and machine learning provide useful results to detect the plant leaves whether they are healthy or infected by a disease, with the help of image processing. Leaf's classification finding has a major role in the agriculture field where it indirectly affects the economy of the country. They need to provide proper care for identifying and analyzing the plant which has healthy leaves or bacterial leaves in the initial stage. The new technology of image processing has made it possible for classifying plants with minimum intervention from humans.

Figure 1.10 Rice plant leaves

DL as a subset of ML has worked with great success in image classification and object detection. Therefore, agricultural research has been moving towards solutions designed for DL. DL strategies have been achieved with high-quality agricultural activities including crop/weed control, fruit harvesting, and leaf crop recognition. Facilitating the task of identifying and classification of an object, deep learning is performed on the meta-learning method. The CNN was one of the first modern-day imaging techniques. The work aims for introducing the learning method for classifying the healthy and bacterial leaves which mainly focuses on classifying the plant along with the images of the leaf. Comparing CNN with other machine learning algorithms gives us a planned technique and the new solution about the pepper plant gives the planned technique that has a better rate of classification other than the methods that exist. The classification solution gives whether the leaf is healthy or bacterial.

1.10.2 Deep auto-encoder

An auto-encoder is a neural network that combines encoding and decoding types. Primary information is delivered into encoding devices, which generate characteristics, which are then sent into the decoding system, which reconstructs the information using the derived characteristics. The dispersion of the encoder and decoder is significantly minimized when teaching the deep auto-encoder system. It is not controlled data when the encoder extracts features and the decoder reconstructs them. Noise removal auto-encoders and sparse auto-encoders are two kinds of deep auto-encoders in which scientists are interested.

1.10.3 Recurrent neural networks

RNNs are a NN model created for successive input and are particularly useful for time series prediction issues. The prior state's output is supplied into the present state in that kind of network, although in conventional network models, all input and output are autonomous of one another. In RNNs, the concealed state retains sequential data. Conventional RNNs could only accommodate sequences of a certain length.

1.10.4 Convolutional neural networks

CNN had already demonstrated impressive accomplishments in the realm of computer detection. It has lately been used in other areas as well. CNNs are used in a lot of modern human-computer interfaces. Since CNNs can locate characteristic regions, they have a benefit against feed-forward networks. As a result, it may retrieve characteristics and activities. Therefore, this research work has adopted deep learning because the conventional method of detecting rice diseases relies primarily on the visual observations of experienced farmers or rice specialists in the field. This necessitates continuous expert supervision, which may be extremely expensive in big farms. Farmers may have to travel across considerable distances, particularly in some developing nations, to talk with agricultural experts, which is undoubtedly time consuming, expensive, predictably slow, and cannot be done across a wide variety of rice. Nevertheless, a new method for identifying and diagnosing plant diseases is now available because of the rapid developments of image processing and pattern recognition technology.

1.11 Problem Background

Image processing techniques could be used to alter the exterior visual effect of infected leaves [22-25]. But on the other side, disease symptoms vary from plant to plant. Every illness has unique characteristics. Symptoms of illness can differ by a condition in terms of shape, size, and color [26-28]. Workers are also sometimes perplexed and therefore unable to make an informed judgment when it comes to pesticide choices. This paper briefly discusses a deep learning strategy for solving the problem of autonomous rice plant detection of diseases as well as classification [29-31].

1.11.1 Flat spot problems

Current artificial intelligence, computer vision, machine learning, and neural networks have created huge advancement in other disciplines such as medical, chemical, mechanical, remote sensing, and agriculture. In this research work, the automatic rice crop disease diagnosis system is studied and implemented using a neural network for faster diagnostic purposes. The agricultural sector is going to face enormous challenges in the future. The food and agriculture organization of the USA made a survey and stated that agriculture yield must extend by 70% to feed 960 million people that reside on the earth by 2040-2045. It is of prime importance to use advanced techniques to improve the productivity of farms, despite non-favorable situations. The major weakness of the backpropagation algorithm is its slow convergence rate. Its main cause is the local minima and flat spot problem. In this research, the flat spot problem of the BP is studied and attempts are made to improve its convergence rate of the backpropagation algorithm by modifying the weighted sum entity input given to the hidden and output layer neurons. Here the two variants of backpropagation are proposed which majorly modify the weighted sum term of the backpropagation algorithm.

The noise injection methods in BP are studied and found that BP converges faster with the noise injection method. Hence in the weighted sum entity, the white Gaussian noise is added and its performance is measured in comparison with the standard BP. This contribution is named noisy net backpropagation (NNBP), as in many pieces of literature the weighted sum entity is aliased as a net. But this approach does not work well when used in the rice crop disease diagnosis system, as a weighted sum is added to a random number. This approach is termed a magnified weighted sum backpropagation (MNBP). MNBP is applied on various standard benchmark problems such as parity (2-bit, 3-bit, 5-bit), encoder, 5-bit counting problem, regression and character recognition, and standard real-time dataset of UCI machine depositories such as Iris, wine, and MNIST. It outperforms all the above-mentioned problems and standard datasets as compared to the standard error backpropagation. The comparative analysis of the BP, NNBP, and MNBP is also done by applying it to a common rice disease dataset. NNBP outperforms BP and NNBP in rice crop disease detection applications.

1.11.2 Problem Statement

In current farming technology, it is essential to improve pest infestations with highly effective approaches that cause the least amount of environmental harm. Computer-aided diagnostics approaches have been prominent for controlling agricultural pests and diseases in the past few decades when paired with crop images [32-35]. An automatic rice disease diagnosis system might give information for rice disease prevention and management, allow time for disease treatment, minimize pesticide residues, and enhance the quality and quantities of agricultural products. Research on effective algorithms for feature extraction as well as classification of rice illness is essential to building such a system [36-38]. There is presently no comprehensive database for the categorization of rice diseases. To address this gap, we created a hybrid dataset for rice illness that was used to train and evaluate a CNN-based system. Apart from dataset issues, the existing CNN models suffer from computational complexity and overfitting issues on high-dimensional datasets [39-42].

This can be solved by applying the dimension reduction technique but the traditional techniques result in lossy reduction which reduces the performance of the model. Therefore, keeping insight into these issues, this research work has designed a hybrid model over a large dataset to improvise the performance according to objectives given in the below section. Considering the magnitude of the root rot disease problems, especially in the hills of Uttarakhand, a detailed study of the plant growth promoting character and control measures with the suitable bio-control agent of this disease has been undertaken in the thesis. These studies would not only impart detailed information about the bio-agent effect on PGP but also help provide suitable eco-friendly control measures against this disease for better production of the crop.The rice crop disease detection system consists of many research sub-domains. Many researchers have contributed to this domain. A brief review of this concept is given in [41]. Every automatic disease detection system consists of the following steps.

Image acquisition: In this step, the infected disease image is captured using computer vision. Computer vision is the ability of the computer to see, which consists of a digital camera, ADC, and DIP system.

Image preprocessing: In Image preprocessing, the quality of the image is enhanced and the noise is removed from the captured image by using various image processing techniques.

Image segmentation: In the rice crop disease diagnosis system, the spots that appear on images are important. Hence the whole image is segmented into parts and only infected parts of the image are extracted using the segmentation algorithm.

Feature extraction: From the infected image, its features are extracted using different feature extraction techniques. Color, texture, and shapes are the features that are used in many rice crop disease diagnosis systems.

Classifier: There are plenty of different methodologies-based ML models such as naive bayes, SVM, BP, and CNN which are used as a classifier in the rice crop disease diagnosis system.

The detailed literature survey of each step of the rice crop disease diagnosis system is covered further. Here the research is carried out in the classifier domain. The attribute selection algorithm is chosen for research and its problem of slow convergence is studied and analyzed. Below the final problem statement is mentioned.

Problem statement: Develop an effective and intelligent rice crop disease diagnosis system using new variants of the backpropagation algorithm. The plan for the research conduction is given below.

- Analysis of existing backpropagation and proposing a new variant of the backpropagation algorithm.
- Implementation of rice crop disease diagnosis system using the proposed variants of the backpropagation.

1.11.3 Pattern recognition and attribute selection

From a broader perspective, pattern recognition is the science that deals with the characterization and recognition of patterns. A pattern is a set of characteristics inherent in a sample and recognition is the ability to classify it. A set of measurements is obtained from the real-world pattern recognition problem. Attributes together characterizing and discriminating the patterns are then extracted from the measurement set, to train a classifier for identifying the class of an unknown pattern.

1.11.3.1 Types of attribute selection

Based on the subset search strategy, state-of-the-art techniques of attribute selection are broadly classified into three categories.

- Exponential selection technique
- Sequential selection technique
- Randomized selection technique

The exponential selection technique evaluates all possible combinations of attribute subsets exhaustively to select the best one. Branch and bound and beam search are a few techniques that fall under exponential algorithms.The sequential selection technique adds or removes attributes sequentially to find the best attribute subset. Sequential forward search (SFS), sequential backward search (SBS), and sequential floating forward search (SFFS) are the three main sequential search algorithms for identifying an optimal attribute subset in the problem space. If the increasing count of attribute values of train inputs, the possible combination of attributes also increases exponentially, thus making it impractical to choose the best attribute subset exhaustively or sequentially. Randomized search techniques incorporate metaheuristic optimization algorithms such as evolutionary algorithms such as genetic algorithm (GA), naturally inspired algorithms such as ant colony optimization (ACO), human-based algorithms, and swarm-based algorithms such as particle swarm optimization (PSO) as search strategy to implement the process of attribute selection.

1.11.4 Dimensionality reduction

Generally, a comprehensive set of attributes is drawn from the measurement space to allow better characterization of a given pattern recognition problem. Nonetheless, it can also introduce redundant and irrelevant information resulting in a high dimensional attribute space making classification a complex task. Hence to battle the problem of the curse of dimensionality in real-world pattern recognition problems, attribute space is either reduced or transformed to a manageable size without discarding the vital information about the problem. The dimensionality reduction technique aims to devise a computationally feasible attribute space yielding better predictive performance than with the comprehensive set of attributes.

On the other hand, attribute selection, apart from dimensionality reduction, selects highly predictive attributes in the original form by rejecting attributes that do not contribute to the superior performance of the classifier. It provides a better understanding of attributes that characterize the problem space. Thus, attribute selection plays a vital role in reducing the dimensionality of attribute space that is fed as input to the final decision space. The modified pattern recognition process with attribute selection as its important constituent is shown in figure 1.11.

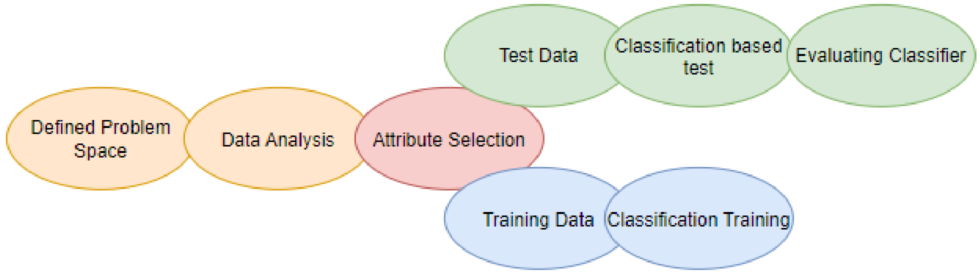


Figure 1.11 Process in modified pattern recognition

From the figure 1.11, it is observed that highly predictive attributes from the problem space are only transferred as input to the classifier. Such a classifier is computationally simple and results in superior predictive performance [43].

1.11.5 Applications of pattern recognition

Researchers are concerned with developing a well-defined pattern recognition process to build intelligent machines in the following fields that demand either or both dimensionality reduction and superior predictive performance.

- Medicine - Medical diagnosis and disease identification
- Natural resources - Study and estimation of Agriculture,
- Forestry and Geology
- Industry – Designing and monitoring, product testing, and quality control.
- Human-Machine communication - Speech recognition, face recognition, and computer vision.

1.12 Research Objectives

According to the world health organization's estimate approximately 7,50,000 people are facing various issues prolonged exposure and use of high doses of fungicide have led to the development of fungicide-resistant strains in several plant pathogens. Over 150 plant pathogens have been reported to have developed resistance. In Tamilnadu, 1742 cases of poisoning with 237 deaths were reported, and out of these, 500 cases with 168 deaths were proved to have been caused by the poison.

- Assessment of the magnitude of different diseases in rabi and kharif seasons in hill areas of Tamilnadu such as vegetable, cereal, and pulses.
- Determination of major disease prevalence and incidences of study.

- Assessment of individual and consortium effects of antagonistic seed treatments for selected diseases after survey through a suitable application of antagonistic such as seed treatment, seedling treatment/foliar spray/value addition of compost in a field experiment.
- Assessment of the impact of bio-agent seed treatment on germination of vegetable peas and paddy.
- Assessment of the impact of antagonistic seed treatment on percent disease incidence and severity.
- To examine the potential of plant growth promoting (PGP) strain consortium for enhancing consistency and level of antagonism against different pathogens and to evaluate the efficacy of plant growth promoting activity in field conditions.

The main objectives of the research work are.

- Study various classifications and their existing variants.
- Implement a fast rice crop disease diagnosis system using LIMO Classification.
- Propose new variants of the MOCLEAR algorithm and LIMO framework.
- Implement a rice crop disease diagnosis system using the proposed variants of the deep neural network (DNN) with the GAN network.
- Performance analysis of the proposed SIOC+LIMO framework used for rice crop ailment diagnosis system.

1.13 Motivation and Research Contributions

The retrieval of the disease area from the leaf image is discovered to be the driving step in the reviewed literature as well as summarized image processing and artificial intelligence techniques that have been used in disease diagnosis. For this, we have investigated and compared numerous segmentation methods and classification methods. Previous studies have attempted to address issues in numerous domains that directly or indirectly impact society by combining image processing and machine learning techniques. Small farms could quickly recognize the diseases and implement preventive control measures. However, it is time and money consuming when it comes to large farms. Therefore, it is crucial to look for a technology that can identify plant diseases automatically, accurately, quickly, and at a lower cost [44].

Therefore, the novelty of this research work is: Deep learning models such as VGG19, XceptionNet, ResNet50, UniversalNet, DenseNet, SqueezeNet, and CNN can produce generic features to classify if the rice plant is diseased or not but also classify which disease it is and provide more accurate results. But still, there are some limitations such as lower efficiency and computational complexity due to the presence of high dimensional data. This research focused on improving rice disease detection performance, effectiveness, and accessibility [45].

Therefore, this work is mainly focused on rice plant disease detection and classification using CNN models. For this, we have implemented three types of CNN architectures such as basic CNN, Vgg16, and ResNet50. The model is hybridized with an image dimension reduction technique using the stacked auto-encoder (SAE). Previous models were designed for improvement in classification accuracy. But their focus was not on the image dimension reduction technique. Moreover, they have considered fine-tuning parameters of classification models, but the impact of dimension reduction on learning parameters and classification accuracy was not mentioned. This work was also extended by considering a large dataset for the detection model (SAMResNet) and is quite efficient to handle the curse of dimensionality issues that exist in existing deep learning models. In SAMResNet we have reduced the trainable parameters by fine-tuning the pre-trained models for learning.

The major concern in the agriculture field is that the yield is proportional to the increasing population due to the limited natural resources. Nowadays precision agriculture is introduced which uses the recent advanced technology in the agriculture field to increase its yield. The automatic rice crop disease diagnosis system illustrated in figure 1.12 belongs to the precision agriculture stream which predicts the disease by analyzing the infected rice crop disease images using computer vision, IP, AI, and ML algorithms.

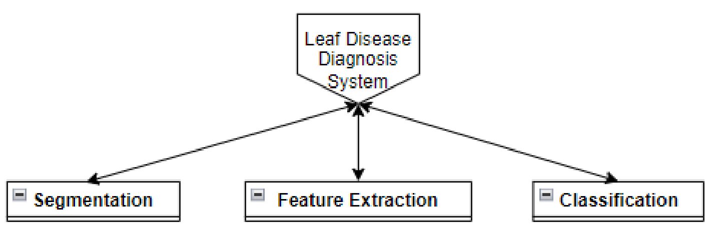


Figure 1.12 Research domains in paddy leaf disease diagnosis system.

1.14 Thesis Organization

In this chapter 1, the report presents the introduction of rice plant diseases, their life cycle, and their types. The chapter also presented the benefits of using image processing techniques to detect diseases properly. The chapter provides a brief description of image processing techniques and steps. Further, the chapter presented an overview of various machine-learning algorithms and artificial intelligence techniques for rice plant disease detection. Moreover, it mainly gives details about attribute selection and the backpropagation process. This chapter also presented the architecture of rice plant disease detection and image processing techniques. This chapter also presented the problem statement with related objectives and motivation of this research work.

The literature survey discussed in chapter 2, the report presents the researcher's contribution to plant disease or rice disease detection methods. The chapter 2 also presented a comparative literature review of different network models, machine learning algorithms, pattern selections, and optimizations with their key features. This chapter also provides various research and results on rice plant disease detection.

In this chapter 3 describes the material and methods adopted for this research work. This chapter describes the tools and functional techniques used for the implementation of this research work. Along with that details of data collection from different sources have already been explained in the initial part of this chapter. Further chapter calculates the performance on different parameters.

In this chapter 4 describes the role of deep learning in rice disease detection and the investigation of various pre-trained CNN models. Ensemble learning can learn patterns such as attribute selection from unlabeled and unstructured data whether it is from images or others. This chapter proposed SIOC-CAS with the HMDF model and investigates the various CNN pre-trained models. Used PSO, FPA optimizations support cognitive attribute selection with three proposed methodologies. In this chapter, the performance of the randomized pattern recognition is analyzed for the rice crop disease diagnosis system. Here the system is implemented using two feature extraction techniques such as texture feature extraction method and feature extraction using pre-trained deep learning model alexnet. The comparative analysis of these two feature extraction techniques on the same paddy leaf disease dataset is included in this chapter. SIOC-CAS models and their performance is evaluated on a rice disease detection dataset. Some best pre-trained models are evaluated for their performance evaluation.

The chapter 5 proposes a novel hybrid approach for rice plant disease detection. Laurent series with Intelligent Multidimensional Object Optimization (LIMO Optimization framework) based on generative adversarial network (GAN) framework provides great support to Agri-farmers who have to take remedy against rice plant diseases. Various types of diseases and their detection methodology were discussed. IMO and MOCLEAR algorithms improve rice plant disease detection and classification. In this chapter LIMO classification accuracy sensitivity and specificity are compared with various machine learning algorithms in different environments.

In this chapter 6 describes the results of all cases discussed in previous chapters. The obtained results of the developed model have been compared with the state-of-the-art literature design. This chapter provided a discussion about the rice plant disease detection system and its performance. All the outcomes of the results are discussed in this chapter.

The chapter 7 explains the conclusion observed in this research along with some future research directions also. Moreover, this chapter has future enhancements with various advanced research areas.

CHAPTER 2

LITERATURE REVIEW

In this chapter, a literature review of different techniques for the application of DL using DIP is presented. This chapter first of all reviewed the contribution of various ML algorithms and classification methods in agriculture. Then the chapter reviewed deep learning applications for various crop disease detection other than rice. In the last section, the chapter reviewed the contribution of researchers for rice crop disease detection methods using DIP, AI, and ML.

The literature survey is carried out in two parts as mentioned below:

- Literature survey on rice crop disease diagnosis system using machine learning.
- Literature survey on pattern recognition with neural network algorithms.

2.1 Introduction

To perform a fruitful literature review, first of all, figure 2.1 shows research papers related to deep learning were selected from different sources.

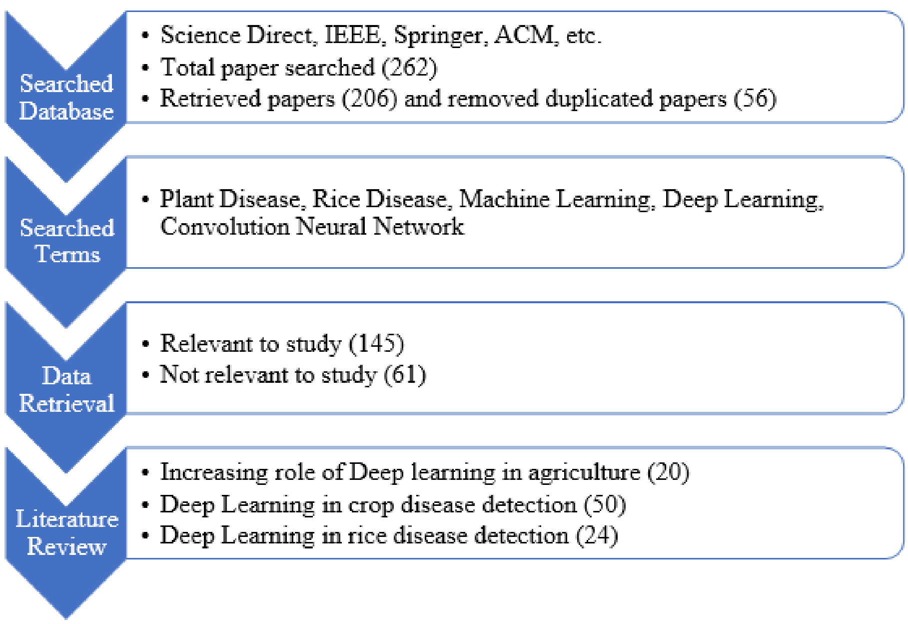


Figure 2.1 Summary of literature survey flow

2.2 Increasing Role of Deep Learning in Agriculture

Researchers looking to diagnose illnesses of grapevine crops using various machine vision and image processing techniques may find the dataset that was published valuable [20]. It contains images of normal and infected grapevine branches. The collection includes images of several grapevine leaf surfaces that were taken from vineyard harvests. The images depict two possible states: healthy and unhealthy, with the unhealthy condition corresponding to the esca infection. Only from July through September, when the environment is most conducive to the illness, can grapevine vines exhibit symptoms of esca disease. A grapevine illness specialist assisted in the manual acquisition of all photographs throughout this time using three distinct technologies such as two cell phones and a tablet.

A hybrid architecture based feature fusion and selection strategy was suggested to classify cucumber illnesses utilizing three fundamental phases [21]. In the first stage, the contrast of an image is improved. The steps two then involve feature extracting, fusing, and selecting. Lastly, a collection of classifiers is used to categorize the majority of discriminant traits. The suggested probabilistic distribution-based entropy (PDbE) technique is used in this study to first decrease the derived characteristics. The suggested approach, which can choose features larger than a threshold, is used to pick strong features just after the serial-based fusion stage. The offered approach is equivalent to many other current strategies given the attained accuracy of 0.93 for 900 images with six classes [22].

Optical scans of potato leaves investigated contemporary deep-learning algorithms for the automatic detection of late as well as early blight infections [23]. First, the PlantVillage dataset was used to train four DNN such as VGG16, VGG19, MobileNet, and ResNet50. This research displayed the fine-tuned VGG16 model's comprehensive structure together with accuracy rate and loss. Additionally, the suggested approach and the practices now in use have been contrasted. An EfficientNet design for classifying plant leaf diseases was suggested to use as the PlantVillage dataset, which had 55,448 photos divided into 39 classifications [24]. The testing dataset findings revealed that EfficientNet architectures have the greatest accuracy, sensitivity, and precision scores of 91.91%, 90.97%, and 91.42% respectively. The suggested approach has drawbacks in that it can't find illness under challenging conditions.

To enhance the effectiveness of TLB identification and severity analysis, presented a DL approach. To optimize the results of blurred, obstructed, and tiny bits of sick leaves, the TLB leaves are recognized using a deep network methodology called faster region-based CNN's networks [25]. In the sugarcane leaf diseases created a DL-NN design that predicts the kind of sugarcane disease by model training on images of diseased leaves. Rust spots, yellow leaf illness, and various types have been identified. The prediction rate of the proposed model is 96%. The technique's drawback is that it needs a large training dataset [26]. A novel citrus leaf illnesses database with annotation, was introduced [27]. ML techniques are used in the identification of plant illness, with just a few DNN being applied in application areas. The findings reveal that the scaled YOLOv4 P7 performs quickly and earlier disease predictions. Leaf illnesses compared to certain other current and effective detection algorithms. Using an ideal mobile network-based CNN established an autonomous method for identifying and classifying plant leaf disorders. With improved sensitivity, AUC, accuracy and F1 score of 0.98, 0.989, 0.987, and 0.980 respectively, and kappa of 0.985, the extensive experiments have shown encouraging outcomes for the OMNCNN model over previously proposed approaches. The suggestion has drawbacks, including poor detection effectiveness for big datasets [28]. The author considered a crop disease detection approach based on CNN trained on consumer-grade camera images. The classifying accuracy was 95.4% as a consequence. Those encouraging findings pave the way for the autonomous detection of plant diseases in the future. Recommendations for putting this strategy into practice in real-world situations are also presented. This technique's drawback is that it needs a large training set [29].

In the field of biological sciences, images are used to convey relevant data and knowledge. In the biological and agricultural fields, DIP and image analysis technologies are critical. The classification of plant leaves and the cultivation of healthy plants are critical aspects of agricultural automation. The disease primarily infects the color and conditions of the roots, flowers, fruits, leaves, and buds of a plant. More plants are affected by this disease which results in a low yield of the crop. The technological development which relates to the processing of images has grizzled the interest of the people and has created a specific way for improving crop yield. Agriculture has a huge impact on the economy of the country. The number of fertilizers and pesticides used to kill bacteria on the crops also impacts the growth of the crop. Some individuals find it to be hard to spot the disease of the plant with their eyes. And if the disease is detected, they cannot find the correct treatment. If the disease is detected on the crops at an initial stage, then this might reduce heavy damage to the crops. For this, they need to identify the additional technique for the classification of the plant by new research ways. DL and ML provide useful results to detect the plant leaf whether they are healthy or infected by a disease, with the help of image processing. Leaf's classification finding has a major role in the agriculture field where it indirectly affects the economy of the country. They need to provide proper care for identifying and analyzing the plant which has healthy leaves or bacterial leaves in the initial stage. The new technology of image processing has made it happen for classifying plants with minimum intervention from a human.

Crop diseases are one of the biggest concerns that currently affect the agricultural industry worldwide. Just as pandemics can be spread among humans, so can diseases be spread among plants. These diseases can be devastating and wipe out entire fields of crops in a short period. In the United States alone, crop diseases account for roughly $21 billion lost each year. The impact of plant diseases also extends far beyond those of economic losses alone. It is estimated that between 20% and 40% of crops are lost each year due to infection, depleting several communities worldwide of their food supplies [30]. Various strains of common plant diseases threaten famine in countries that lack the resources to defend against these outbreaks and could lead to forced migration. Plant pathology is a vital field of study for those in the agricultural and biological fields. Experts must be able to recognize symptoms of plant diseases early on and know how to defend against a wide variety of pathogens. Employing plant pathology experts can be expensive and difficult to come by for growers, however, and often by the time issues are self-identified with a harvest much is too late to salvage.

However, while several papers developing plant disease recognition models have been released in the past few years, many fail to employ good practices in their model development or use realistic [31]. This makes it extremely difficult to compare the effectiveness of one paper against another when deciding which models may be useful in the field. To address an effective detection issue by outlining a set of procedures to follow when developing a model for plant disease detection. Many research articles that implemented CNN image classifiers on some form of plant disease detection, and address various points of their methodologies. Using these works, additional survey papers, and generalized machine learning approaches, describe a set of best practices to use and critique previous literature on these grounds. Implemented our neural networks using these best practices and compared their levels of accuracy on strawberry images. By outlining best practices to use when developing a model for plant disease detection, to inspire future work that follows proposed guidelines. This will standardize much of the research in the field, promote the creation of more accurate models for real-world use, and allow for better model comparison between various research papers.

For diagnosing multi-crops leaf disease employed CNN approaches. Employing a DL algorithm to extract visual attributes, the ill and normal leaves were categorized. The precision, sensitivities, specificity, accuracy, recall, and F1 measure quality assessment variables were computed and tracked. The primary goal of research using the suggested approach is to continue to improve significantly. The suggested study attained an accuracy of 0.98 and 0.95 for grapes and tomatoes in the trial [32]. Introduced a CNN with an incorporated enhanced convolutional block attention module (CBAM) founded on Inception and residual architecture, to improve the diagnosis of leaf disease detection. Southern China's most widely grown crops include corn, potatoes, and tomatoes. For identifying the three illnesses of maize, potato, and tomato, the suggested model obtains an accuracy rate of 99.55% [33].

A tomato pest image was collected that included a variety of typical tomato pests. To classify different kinds of tomato pests, deep learning CNN models were used. To cut down on training time, transfer learning was implemented. With an accuracy of 94.95% after image augmenting, the VGG16 model performed better than the others models. The ResNet50 with discriminant modeling process obtained an accuracy of 97.12% classification results in the CNN+ML models. The identified and classified plant diseases from leaf images recorded at different resolutions. A big plant leaves an image collection from different nations that is used to train a dense CNN architecture. On unseen images with complicated backdrop circumstances, the observational results show a better result with an accuracy of 0.91%. The model's weakness is that it cannot handle a huge number of datasets [34].

The various analysis conducted a comprehensive assessment of the literature to determine the existing research of CNN in the identification and classification of plant infections, as well as to find patterns and gaps. A novel DL approach has been taught to reliably diagnose more than nine types of wheat illnesses. The suggested approach has a 97.88% testing accuracy. Moreover, it outperforms the additional two prominent DL models such as VGG16 and ResNet50 – in terms of accuracy by 7.01% and 15.92%, respectively. The suggested strategy outperforms other criteria [35-36]. An investigated DCNN transfer learning (TL) to train the model with transferred image values of plant leaf disease. The recommended method offers an accuracy rate of at least 91.83% on the public dataset, which is a major increase over earlier state-of-the-art. Experimentation portrays suggested technique is valid and that it may be used to identify plant diseases effectively. The suggested model has the drawback of not applying it to real-world scenarios [37]. A DL-based technique was used for tomato disease diagnosis that produces synthetic images of tomato plant leaves using conditional generative adversarial (C-GAN). On the publicly released PlantVillage database, the suggested model was thoroughly designed and evaluated. For tomato leaf image categorization into 5, 7, and 10 classes, the suggested technique obtained an accuracy of 99.51%, 98.65%, and 97.11% respectively. The efficiency of deep learning approaches used to plant pathology is influenced by the number and diversity of datasets [38].

2.3 CNN in Various Crop Disease Detection

Suggested that tea disorders be identified in hopes of preventing and handling tea disorders promptly. The preliminary findings demonstrate that a mash-up of classical learning approaches and deep knowledge methods may effectively diagnose tea diseases via different processes. For the selected tea sample, it received 90% acceptance rating, which is more than 30% more than the SVM [39-40]. The primary problem that is experienced while using this screening tool is the absence of visual data which can portray the essence of the numerous conditions and diseases existing in the implementation [41]. The intensity of this problem is lessened by ion-augmentation technologies, but it is unable to recreate much of the real distinction. Instead of focusing on the full leaf afterward, this study explored the usage of single-stranded lesions in this study. Further photos are not required to increase data migration because every location seems to have its unique features. It also allows for the diagnosis of many illnesses that affect the same leaf. On the other side, the allocation of possible symptoms must be done in such a manner that total automation is avoided. In comparison to the accuracy of the original image, the technique's accuracy is typically 12% higher. The treatments developed in this work not only affect the volume of image files and also raise the data dimensionality since evolutionary divergence for each image is not properly considered. This approach has significant drawbacks as well, but in the context of limited data, it produces trustworthy conclusions. The proposed CNN based model, which was approved by three different CNN sites. It streamlines the process of disease classification by combining the benefits of research with several data augmentations simultaneously [42].

An investigated how various problems may be solved by utilizing CNN modeling and image-sharing approaches [43]. The S-CNN model with the merged image doubled its performance in comparison to the F-CNN trained model with the whole image. The accuracy rating is 98.6%. As a result of this study, non-experts may now use computational algorithms to identify illness throughout time. For detecting tomato leaf virus infections offered two different techniques. To acquire significant work, the initial architecture takes a dozen classes. The pull operation is attached to the surface of the residual deep network in the second design. Botanical village data was used in the trial, which contained three illnesses which are natural blight, fire measures, and butterflies [44].

A technique for creating image acquisition plantvillage datasets was presented [45]. Android applications using java web services and DL are the two phases they apply in the framework. To make our android models and apps more user-friendly using JWS, CNN deployed CNN in the five, four and third layers. Various aspects investigated a technique based on in-depth research that may detect the leaf diseases of several blue plant species. A library of 1200 images of healthy and sick mangoes was used to make the identification. The suggested CNN model diagnoses leaf illnesses in blue plants with a 96.67% accuracy rate, demonstrating its capacity to be used in real-time [46].

For fruit research programs investigated image dispensation or ML methods. Fruit massaging is depending on the magnitude, form, coloring, form, body, or calyx of the fruit. Color is the most important aspect of an image. The colors in this article have been altered. In addition, this article discusses comparative research on machine learning approaches including rules-based systems, SVM, KNN, and ANN [47]. For plant disease tracking and categorization employed digital imaging technology. The green screen is recognized, and the original image is wrapped around it. To capture features, the impacted regions groups are transformed to such as hue, saturation, and intensity (HSI) colors [48]. Centimeters of spatial resolution are used to construct a matrix from the HSI model's elevation and saturation. Finally, classification is performed using neural networks. A different machine-learning approach for the detection of disease in fruits [49]. The proposed digital imaging technologies are to monitor agriculture applications with faster processing [50]. As infection in the leaf would harm the green part of a leaf, attempting to enhance the damaged section of the leaf by eliminating the green portion. This is because the infection will damage the green when it happens in the leaf. Their author chose to employ the median filter approach for this work [51-52].

Recommended building an autonomous system to identify mango plant illnesses. The article presents a rule-based framework and leaf image collection [53]. Calculating the trajectories and histograms of plant parts like leaves and fruit trees was one of the tasks presented using MRKT-based techniques [54]. These histograms, together with a set of markers that were used by ANN, lead to the development of better approaches for recognizing anthracnose, which manifests itself as black spots inside blue fruits and leaves. The findings show that the proposed model has a sensitivity of 98%. Investigated healthy leaf litter and introduced powdery mildew into the experiment. Using a hyperspectral microscope, leaflets were collected daily at intervals ranging from three to fourteen days following the inoculation. This microscope offers two different approaches to analyzing predetermined sets of situations. Using a C-GAN, from the plant growth and its symptoms predicting plant problems and finding remedy for it [55]. Examined the outward characteristics of a variety of plants. A significant number of farmers detect and identify plant illnesses via the use of the visual technique. It is impractical to employ on farms and giants because it needs continual monitoring and it has limited use. In addition to that, farmers do not get diseases that are localized [56-57]. It was the goal to compare deep learning systems and optimize CNN training.

The synthetic minority over-sampling technique (SMOTE) model and cross-entropy loss with DCNN from scratch were the focused [58]. To properly anticipate underrepresented classes, it was necessary to address the problem of high-class imbalance. To address problems associated with disease detection suggested many various ML techniques to facilitate the automated categorization of plant species [59]. DNN, often known as DNNs, is one kind of machine learning method that has been implemented across a variety of data sets. However, DNNs have often been used in silos contexts, and there have been no efforts made to reuse or transfer the information gained by using DNNs in a variety of contexts [60]. The field based field (FBF) model generates a picture of artificially straightened maize. They improved the FRP model's using the ReCNN model. The imaging-based knowledge-based systems technique for identifying sickness intensity was proposed to assess using 155 pictures, and the overall correct classification rate was 95.48%. Adapting the CNN size based on images and the area to be investigated can quickly improve R-CNN marker achievement [61]. This approach creates greater related-group results than modern methods.

The investigated whether or whether it is more useful to fine-tune a pre-trained model on a task involving the identification of plants as opposed to a test involving the recognition of generic objects. In particular, the author demonstrated via the use of visualization methods that the qualities taught vary depending on the approach that is taken, and that they do not necessarily concentrate on the portion of the body that is afflicted by the illness [62]. An ensemble model with CNN was suggested for plant disease detection. This model was constructed to identify the images of the leaves as either healthy or sick [63]. In this particular setting, CNNs are used because of their capacity to circumvent the technical challenges that are connected to the categorization issue of plant illnesses. However, CNNs suffer from a large range of hyperparameters with distinct designs, making it difficult to manually determine the hyperparameters that are best for CNNs. Both of these goals were achieved. Using techniques of random minority interpolation and randomized majority under-sampling, the issue of the unbalanced dataset utilized in this research has been handled. Additionally, several limits in terms of the amount and variety of samples have been overcome.

It is required to create a DCNN design for the automatic identification and categorization of biotic and abiotic paddy plant stresses using field photographs [64]. This work employed the pre-trained VGG-16 CNNs to automatically categorize photos of damaged rice crops obtained during the booting development stage. The held-out dataset reveals that deep learning has been applied to 30,000 field photographs of five rice different crops with twelve stress levels including healthy/normal, suggesting technological viability. The model was trained with visual data that was made accessible to the public and consisted of 87,848 photos collected on farmed land both in a laboratory setting and in real settings [65]. It has 25 plants that belong to 58 distinct plant classifications, as well as both diseased and healthy plant species.

The most current partitioning technique was described as an in-depth technique available. The use of images acquired in varying degrees of gray when exposed to light provided support for the decision to put the development method into action. The findings indicate it may lower the mistake and enhance the performance, which is in line with the concept of harvesting using robots [66]. The deep dive is the analysis D-CNN network ML method for rice crop disease detection, segmentation, classification and decision [67]. In the suggested approach, employ ImageNet and Inception-trained VGGNet. Instead of having to start and randomly set weights, utilize ImageNet's pre-trained networks. On the public dataset, the recommended strategy achieves 91.83% validation accuracy, an improvement above the prior state-of-the-art technique. Even with tough backgrounds, the recommended approach predicts rice plant class with 92% accuracy. Experiment findings show that the suggested technique is viable and that it may be used to detect plant diseases quickly. There are 500 rice images and 466 maize images, as well as over 1000 crop leaf images. The CNN architecture for plant leaf disease detection was implemented. CNN was trained using 54,305 images of damaged and healthy leaves from the PlantVillage dataset. The recommended design has 95.81% illness classification accuracy, and many observations were made with various CNN hyper-parameters. The study's results are comparable to previously published approaches [68].

In the transfer learning approach, employ ImageNet and Inception-trained VGGNet. Instead of starting from scratch and randomly setting weights, utilize imageNet's pre-trained networks. The proposed approach validates at 91.83% on the public dataset. The recommended approach yields 92% class prediction performance but under demanding conditions. The total images used are 500 rice and 466 maize images, as well as over 1000 crop leaf images and rice plant images. Experiment findings show that the suggested technique is viable and that it may be used to detect plant diseases quickly [69]. The suggested algorithm automatically detects capsicum illnesses and classifies them as normal or diseased, based on whether the capsicum or its leaf is infected with bacterial or fungal disease. The contaminated area of the capsicum plant is identified using the k-means cluster algorithm, after which texture, that is GLCM features, are retrieved for this infected area, and different bacterial or fungal capsicum infections may be categorized using these features using the support vector machine (SVM). Shimla and Solan are two districts in the state of Himachal Pradesh. To acquire a better image, all images were taken using a high-quality digital camera with a 16-megapixel resolution and saved in jpeg format [70]. Suggested solution is based on a combination of color balancing and Superpixel technology, to remove the impacts of uneven lighting, color-balanced images. Second, using the Superpixel method, compact areas are generated from the altered image. Finally, many classifiers are examined, and random forest (RF) is chosen as the disease classification algorithm. The suggested technique attained a cross-validation method prediction rate of 93.12% on a combined dataset, demonstrating that it is successful in classifying illnesses in the presence of a crowded backdrop [71-72]. The proposed utilizing the PlantVillage database, which has 18,162 images for nine sick classes and one seine class, to investigate a plant disease such as tomato [73]. DenseNet169 and InceptionV3 CNN architectures were used to detect and classify several tomato plant illnesses. The RMS propagation and adam optimizers, as well as transfer learning technologies with a batch size of 32 were used.

The method employs google colab for the training process and a PlantVillage dataset with 5740 images [74]. Data collecting involves training, verification, and testing. 70% of the information such as 4018 images was used for training, 20% for validating, and 10% for testing. After training on the PlantVillage dataset, detection performance was 81.28% and the SVM classifier was 91.9% accuracy. Studying an ideal plant disease detection model incorporates several plant diagnoses [75, 76]. ML based technique for detecting soybean illnesses utilizing pre-trained AlexNet and GoogleNet. Compared to traditional pattern recognition algorithms, this accuracy was much greater. The suggested model has the maximum effectiveness in terms of identifying soybean illnesses, according to the testing data. Using CNN presented an effective way for systematically classifying plant disease symptoms [77, 78].

EfficientNet DL architecture for plant leaf disease detection was suggested and evaluated to current framework systems [79]. Train models on PlantVillage. The classifiers are trained using 55,448 original and 61,486 improved images. EfficientNet as well as other DL models employed transfer learning. All design layers have been capable of learning in transfer learning. In the testing data, the EfficientNet B5 and B4 networks had the highest accuracy and precision scores. Rice plant disease identification and classification were conducted with various resolution images reported using a deep-learning-based technique. A big plant leaves an image collection from different nations that is used to train a dense CNN architecture. The proposed study considers six crops in 27 distinct categories under laboratory and on-field situations using the PlantVillage dataset, which contains 26,590 image leaf classifications. The suggested DL based system can accurately categorize different varieties of plant leaves, according to experimental data [80]. The vigna mungo was offered as an autonomous DL based viral infection detection approach for a leguminous plant that is widely farmed in India. After comprehensive testing of the method, all achieved high validation accuracy and gave testing various measure metrics of 91.234%, 96.429%, and 97.403% on diverse leaves images, respectively 433 images [81]. Suggest a deep learning strategy for improving tea leaf blight (TLB) identification and intensity analysis performance. The original photos are enhanced using the retinex algorithm, which reduces the impact of light fluctuation and shadow. To increase the diagnosis effectiveness of blurry, occluded, and tiny fragments of diseased leaves, a deep learning framework dubbed faster region based CNN is used to identify TLB leaves. To accomplish severity grading and enable illness seriousness analysis, the identified TLB leaves are fed into the trained VGG16 networks [82-84].

DL NN architecture uses photos of diseased leaves to forecast the sort of illness that will impact the sugarcane crop. A total of 1000 images were utilized in the prediction training method. The accuracy of the prediction model is 96%. As a user interface for this model, an android application is being created. Farmers may use this software to shoot photographs utilizing their phone camera or choose from a collection of photographs. The image is transmitted to the model in the server, which processes it and predicts the illness. This forecast will be presented on the farmer smartphone through the app, allowing them to take the required precautions to avoid damage [85]. The data fusion modality method CNN was used in the detection plant ailments categorization, as well as identifying developments and inadequacies. In this regard, they examine 121 articles over the previous 10 years that use various methods for illness identification, dataset features, crops, and pathogens under investigation. It is feasible to comprehend the new developments in the application of CNNs in the diagnosis of plant illnesses and to determine the shortcomings that require to be addressed by academic researchers based on the findings of the comprehensive study [86].

The visual geometry group (VGG) model, which is based on CNN, is utilized to enhance performance measurements. The dataset of crop leaves images is used to train and evaluate the model. The correctness, sensitivity, specificity, high accuracy, remembrance, and F1-score performance measuring variables were computed and examined. The primary goal of research using the mentioned approach is to continue to enhance functionality. The newly developed model is more accurate in classifying diseased leaves. The model was proposed for leaf disease detection, in which rice disease samples were also considered [87-90].

2.4 CNN in Rice Disease Detection and Classification

Research the topics to recognize or classify illnesses based on plant leaflets. The elimination of agricultural illnesses via the use of image-processing technology would save the lives of farmers while simultaneously protecting the fruits of their labor. During pre-processing, RGB images are converted to HSV to remove their source, and afterward binary images are formed based on hue and saturation to differentiate infected individuals from healthy ones [91]. The proportions and beginnings were created to research the distribution of the places that were impacted by the disaster [92]. Wheat production can be increased through the utilization of profound neural networks. The figures collected by the Institute indicate that farmers lose up to 37% of their harvest each year as a direct result of illnesses and pests. These techniques include image processing and ML. Establishing an independent conversion technique for the device that will be used in the color conversion system should be the first step after developing a color conversion system [93].

Grapes and wheat both showed symptoms of the plant disease that was discovered [94-95]. There were a total of 185 photos examined, 85 of which were grape leaves and 100 of which were huge leaves. In cases when the image resolution was left unchanged, the closest neighbor approach was implemented. Create a new image using a median filter that you set up. It was decided to utilize the K-means distribution. The introduced method was analyzed images to find various genetic diseases. The author created a smartphone application for diagnosing diseases using probabilistic neural networks and fuzzy entropy. The proposed pattern was discrimination model to detect rice disease detection. The model was improvised with a voting classification technique to classify four classes of rice disease [96]. The strategies of machine learning that identify rice illness were used to achieve better accuracy. In conducting this study, the following guiding principles were utilized such as class type, format, input, and correctness [97]. The test accuracy score may well be improved by increasing the amount and quality of datasets. It is anticipated that by using this method, rice plant disease control would be properly carried out, leading to increased yields using kaggle around 1600 images [98-99]. The author used CNN and other classifiers like SVM to achieve an accuracy of almost 95.83% on a defined rice plant disease dataset. In response to the success of CNNs in image classification developed deep learning-based methods for diagnosing illnesses and pests from images of rice plants [100]. Agriculture technology is exploring AI to improve its output. Experiments show that using 1536 Images, the target prediction rate is 93.3% while using a much less model size [101]. DNN were suggested to identify crop diseases for the rice images. A KNN approach that is already in use is used to do the categorization. DNN is offered as a method to boost accuracy [102].

Table 2.1 A literature review of the latest rice plant disease detection research papers and their functionality

Author's	Research Title	Year	Source	Disease	Techniques	Key Finding Limitations
Rutuja R. Patil, Sumit Kumar	Rice Data Fusion (RDF): Rice Plant Disease Detection with Multimodality Data Fusion.	2022	IEEE Access	Diagnoses Rice Disease	CNN MLP	Non adaptive data fusion. Achieved 95.31% accuracy.
Deming Zhai et.al	Rice Plant Disease Detection with Rectified Meta Learning.	2022	ACM Transactions on Multimedia Computing, Comn, &	Plant Disease	CNN And ML	Huge noise input. Achieved 96% classification accuracy.

			Apps (TOMM)			
Shruti Aggarwal et al	Rice Disease Detection using Artificial. Intelligence and ML Techniques to Improvise Agro-Business.	2022	Hindawi Scientific Programming	Rice plant leaf diseases	KNN SVM Fuzzy classifie r	Overall image dimensionality issues. Precision archived 92.06%.
Dengshun Li et al	A Recognition paddy Plant Diseases Detection with DCNN.	2020	MDPI - Sensors	Rice diseases and pests	Deep convolu tional neural network	Video streaming conversion difficulty. Part occlusion.
Asmiaty Sahur et al	Effect of Methanotroph Bacteria Isolated from Paddy Rice Plant on Growth and Yield Components of rice.	2022	Hindawi - International Journal of Agronomy.	Methanotr oph Bacteria	Rando mized complet e block design	Environmental variation. Yield improvement is 95.45.
Junde Chen et al	LWI Networks for Rice plant Disease Detection.	2022	IEEE Sensor Journal	Plant Disease	SVM Deep Learnin g	High computational cost.
Davindar et al	Plant Doc: A Dataset for Visual Plant	2020	ACM IKDD CoDS	Rice plant diseases	KNN SVM	Applicable with the minimal structured dataset.

	Disease Detection.				Fuzzy classifie r	Time complexity increases.
Muhammad Hussain et al	Object detection for rice disease detection with CNN.	2022	MDPI – Food Engineering and Technology	Rice leaf defeats	Deep learning Lightweight	Stringent quality inspection protocols. Overall surface coverage difficulties.
Wen-Liang Chen et al	IoT with Rice Blast Detection System.	2020	IEEE Internet of Things Journal	Rice blast	Artificial intelligence Precision farming	Transmission data loss percentage high. Low Accuracy calculated with raw images.
Ashok Kumar Saini et al	Citrus fruits diseases detection and classification - transfer learning	2021	ACM - International Conference on Data Science, ML, and AI.	Citrus Fruits Diseases	Machine learning Neural network.	High computational cost.
Ashutosh Kr Singh et al	Hybrid Feature-Based Disease Detection in Plant Leaf Using CNN, Bayesian optimized	2022	Wiley Hindawi - Journal of Food Quality	Plant diseases	Bayesian Optimizer Random forest classifier.	Achieved an Accuracy of 96.6%. Hyper parameters were selected for non-evolutionary methods.

	SVM, and RF Classifier					
Rabia et al	Mango Leaf Disease Detection with Vein Attribute Selection Method.	2021	MDPI – Applied Science	Mango leaves diseases	SVM Vain pattern	SMVDU, Katra, and J&K Dataset were used. Achieved a quite low accuracy of about 88.39%.

The CNN is used to identify and categorize diseases in images. Six significant rice diseases were looked upon. The author trained the model on a total of 6,330 images and assessed the outcomes. Mask R-CNN, Faster R-CNN, and RetinaNet all received accuracy scores of 75.92%, 70.96%, and 36.11%, respectively [103]. The highest accuracy achieved was 97% using CNN [104]. Analyze rice leaf disease images. The SVM system then classifies and forecasts the illness. This model's accuracy exceeds standard back propagation neural networks. This approach uses deep learning to improve agricultural disease diagnostics. With 8911 images, the CNN model is 96.8% accurate, AlexNet is 93.79% accurate, and VGG is 91.65% accurate. Data was obtained in the rural countryside using a canon 66D digital camera. The recommended approach detects rice leaf diseases with the greatest accuracy, proving its efficiency [105].

The study focuses on various rice ailments: leaf smut, stemborer, BPH, hispa, brown spot, and bacterial leaf blight, etc. To identify various rice diseases, a fully connected CNN was trained using 4000 sick leaves and 4000 healthy rice leaf images. The fully connected CNN recommended achieves 99.7% accuracy on the kaggle dataset. This accuracy exceeds earlier illness detection and categorization approaches [106]. The approach was used in a modified k-means algorithm to segment the rice plant picture. After segmentation, color, shape, and texture parameters extract features. Performance is compared using naive bayes, SVM, and PNN. The proposed technique outperformed the other three. Using five fold cross-validation, the final accuracies for 1000 bacterial leaf blight, brown spot, healthy leaves, and rice blast were 95.2%, 97.6%, 99.2%, and 98.4%, respectively. Images were taken using a canon eos 3000D [107]. The transfer learning is used to classify illness signals in rice plant images. The author used VGG and GoogleNet architecture for learning and achieved 92.24 and 91.28%, respectively [108]. A color-based method was used to classify rice plant diseases, total 14 color spaces, extracting four features from each channel for 172 features. The SVM classifier had 94.65% accuracy. The total 619 images were trained and tested using public datasets. This data came from a four-type agricultural field to developed rice-fusion for diagnose rice disease. Rice-Fusion has a 95.31% testing accuracy. Experiments show that the multimodal rice-fusion framework outperforms unimodal frameworks [109].

Leaf blast, brown spot, healthy, and hispa were utilized. This research combines deep convolutional neural networks (DCNNs) with SVMs, ANNs, and DCNNs with LSMs long short-term memory (LSTM). PSO, AFSO, and EAFSO are used to discover appropriate LSTM weights. The suggested strategy's results are compared to standard feature extraction approaches employing SVM, ANN, and LSTM. The suggested DCNN-LSTM (EAFSO) approach outperforms others. The suggested EAFSO technique combines DCNN-LSTM and DCNN-SVM, outperforming DCNN-ANN [110]. The utilized mobilenet-V2 pre-trained on ImageNet as the backbone network and incorporated the attention method to boost learning capacity for minute lesion information. Table 2.2 summarizes the summary of research papers for the detection of various disease detection methods with their limitations.

Table 2.2 Comparative literature review on overall plant disease detection system and its research functionality

Reference	Year	Disease Type	Technique	Key Findings/Limitations
[40]	2019	Tea Leaves	Support Vector Machine	● PlantVillage Dataset was used. ● 1500 Image samples were used. ● Achieved 90% accuracy. ● Increasing samples of image SVM results in overfitting issues and increases the complexity of training.
[41]	2019	Plant-Based	CNN	● The dataset was taken from https://www.digipathosrep.cnptia.Embrapa.br ● 1567 images were used.

				● 85% of accuracy was achieved. ● Low efficient model.
[42]	2019	Multi-Crop Plant Disease	CNN	● PlantVillage Dataset was used. ● 1000 images were used. ● 93% of accuracy was achieved. ● Can detect multiple diseases but with lower efficiency as for training very few samples were taken.
[43]	2019	Plant Disease Detection	CNN	● Plant Village Dataset was used. ● 600 images were used. ● 95% of accuracy was achieved. ● Very few samples for training. ● The pre-processing step has not been used this results in ambiguous detection results in real-time processing.
[44]	2019	Tomato	CNN	● Plant Village Dataset was used. ● 98% of accuracy was achieved. ● Restricted on input image size.
[45]	2018	Leaf Disease Detection	CNN, AND	● Plant Village Dataset was used. ● 84% of accuracy was achieved. ● Very low efficiency.
[46]	2018	Mango Leaf Diseases	CNN	● Plant Village Dataset was used. ● Achieved accuracy of 96.67%.
[58]	2019	Cassava Plant Diseases	CNN	● Only 600 images were used for detection efficiency evaluation. ● Optimal parameters were not used. ● The dataset was taken from a Kaggle source. ● Achieved accuracy of 93%. ● Not optimal learning parameters.

[59]	2019	Plant Species Classification	CNN	● Flavia, Swedish Leaf, UCI Leaf& Plant Village image datasets were used. ● Achieved an accuracy of 90.17%. ● Low efficiency. ● A very large number of learning parameters are used.
[60]	2019	Sugar Beet Leaf Spot Disease	CNN/ Faster R-CNN deep learning model	● Plant Village Dataset was used. ● Achieved an accuracy of 95.48% ● Was trained and tested with 155 images only.
[61]	2019	Maize Seedling Detection	CNN/ Faster R-CNN	● Plant Village Dataset was used. ● Achieved an accuracy of 97.71%. ● The computational complexity of the model was high.
[63]	2019	Plant Diseases Diagnosis	CNN	● PlantVillage dataset was used. ● Achieved an accuracy of 96.6%. ● Hyper-parameters were selected for non-evolutionary methods.
[64]	2019	Paddy Crop Stresses Classification	CNN	● The Kaggle dataset was used. ● Achieved an accuracy of 92.89%. ● Lower efficiency.
[52]	2019	Mango Leaves Anthracnose Disease Detection	Multilayer CNN	● Achieved a quite low accuracy of about 88.39%.
[71]	2020	Tomato Plant Disease Detection	RF	● The Plant Village dataset was used. ● Training and testing were performed on very few images. ● Achieved accuracy of about 93.12%.

[72]	2020	Corn Plant Disease Detection	DNN-CNN	● The Plant Village dataset was used. ● Achieved a quite low accuracy of about 88.46%
[74]	2022	Plant Leaf Diseases Detection	DL	● The Plant Village dataset was used. ● Achieved quite a low accuracy of about 81.28%
[76]	2021	plant diseases	CNN	● Public Dataset was used with 649 and 550 images only. ● Achieved an accuracy of about 98.75% and 96.25%.
[34]	2021	Vigna mungo plant Disease Detection	DL	● Leaf Image dataset with 433 Images as training and testing samples were used. ● The accuracy rate was fluctuating.
[83]	2022	Rice leaf disease detection	Lightweight attention network	● Experimented on coffee leaf dataset and apple leaf dataset. ● Achieved low accuracy of 92.59 %
[91]	2019	Paddy Leaf Diseases	Optimized DNN	● Tested only on 400 images. ● Achieved only low accuracy of about 86.42%
[92]	2018	Rice Blast Disease	SoftMax CNN	● Achieved accuracy of 95%.
[100]	2020	Rice Plant Disease Detection	CNN	● Datasets were taken from https://drive. google.com/open?id=1ew BesJcguriVTX8sRJseCD bXAF_T4akK ● 1536 Images and achieved 93.3%. ● The model was time-consuming.

[102]	2020	Rice Disease Detection	DNN	● Datasets were taken from Rice Leaf Disease. ● Training and testing were performed on 209 Images and achieved very low accuracy of 82%.
[103]	2020	Rice Plant Leaf Disease Detection	Deep Learning	● Datasets were taken from Plant Village. ● Tested on 6330 Images. ● The highest precision rate was approx. 75.92% which is quite low.
[104]	2020	Rice Leaf Disease Identification	SVM-CNN	● Datasets were manually collected. ● Tested on 5932 images. ● Achieved accuracy 97.96%. ● Results in overfitting issues.
[105]	2020	Rice Leaf Disease Identification	CNN	● Datasets were manually collected ● 8911 Images were used for training and testing and achieved accuracy between 91.65 % - 96.8%.
[107]	2022	Rice Plant Disease Identification	K-means segmentation algorithm	● Datasets were manually collected. ● Tested on 1000 and achieved 95.20% of accuracy.
[108]	2022	Rice Diseases Detection	CNN	● 12000 Images. ● 92.24% and 91.28% for VGG-16 and G-CNN respectively. ● Lower accuracy on such high data.
[109]	2021	Rice plant disease classification	ML	● Datasets were collected from the public. ● Tested on 619 Images and achieved an accuracy 94.65%.

[112]	2021	Rice plant disease Detection	Deep CNN	● Datasets were taken from rice leaf disease dataset. ● Tested on 1600 images and achieved an accuracy of 97.5%. ● Convergence toward the optimal solution is slow.
[110]	2022	Rice Disease Diagnosis	Multimodality Data Fusion	● Datasets were manually collected. ● Tested on 3200 Images and achieved an accuracy of 82.03 which is quite low.
[113]	2021	Rice leaf diseases	DNN	● Datasets were taken from kaggle. ● Tested on 5200 images and achieved an accuracy of 95.67%.
[114]	2021	Rice crop disease	Transfer learning model	● Datasets were taken from a public dataset. ● Achieved an accuracy of 98.48%. ● This model transfer learning was performed twice for model training which increases the complexity of the model.

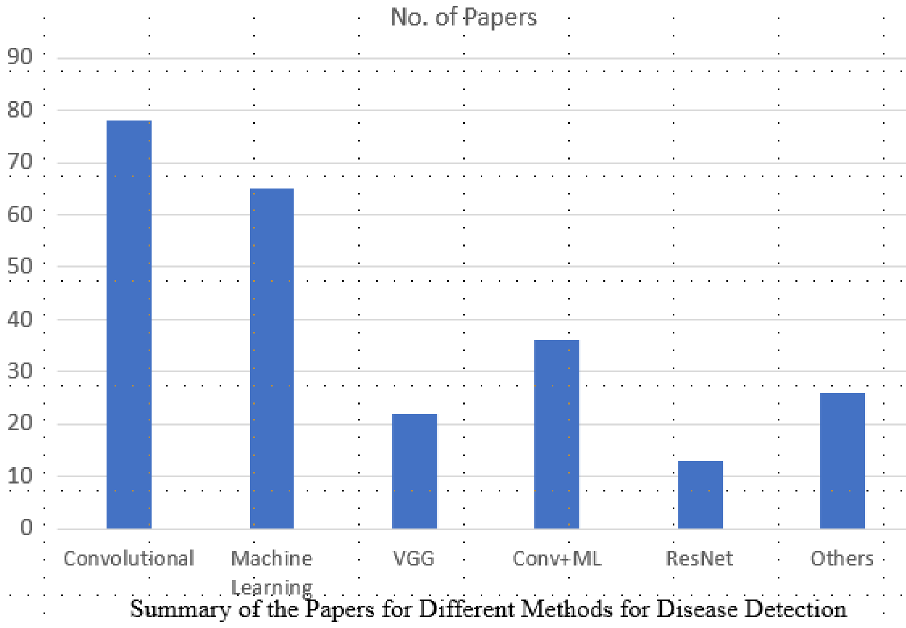


Figure 2.2 Summary of the papers for different methods for disease detection

After a detailed review of techniques and methods, figure 2.2 represents the research contributions of various techniques for plant disease detection models. From the above figure 2.2, it is clear that major research work is focused on CNN for disease detection. But still, very few contributions are presented to the application of pre-trained models. Similarly, figure 2.3 shows the contribution of ML methods in rice diseases detection and other plants disease detection comparison. From the figure, it is inferred that there are very few contributions to rice crop problem identification. This motivates the dissertation in the direction of rice crop disease identification using pre-trained models such as VGG, RESNET, etc.

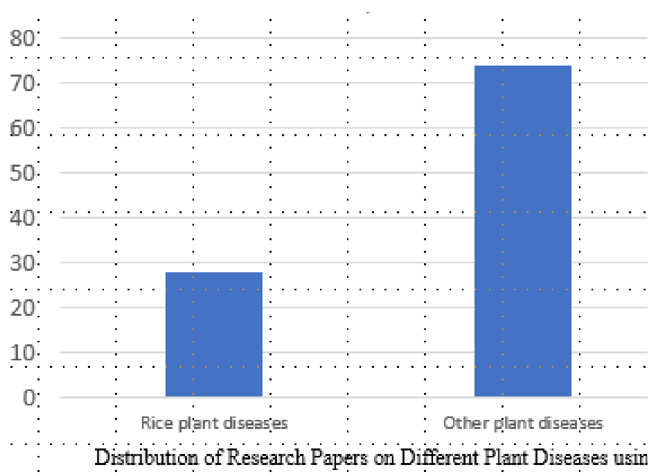


Figure 2.3 Distribution of research papers on different plant diseases using DL

2.5 Various Plant Diseases

Surveys were conducted during 2009 and 2010 for identification of major diseases in crops from June to September for paddy and October to December for vegetable peas, a peak period of the occurrence of diseases. Vegetable peas and paddy, both are the major off-season vegetable crops cultivated in Uttarakhand hills. Here, the farmers have small and scattered land holdings, therefore, intensive cultivation with reduced crop rotation and extensive monoculture resulted in a decline in biodiversity which has led to highly resistant pathogen load and increased crop losses. Root complex is observed as a very serious disease in vegetable peas and paddy, crops which are cultivated as a major vegetable in almost every district of Uttarakhand hills. Vegetable pea is an important crop cultivated in Uttarakhand hills. The total area of vegetable peas is 11187 ha, with 86937 MT pod yield obtained 7.7 MT/ha in 2011-12. This crop has good market potential in the local market. Vegetable Pea is adapted to grow in a cool moist climate with a temperature range of 5-22^{0}C. In India, this crop is grown in Uttarakhand, Himachal Pradesh, Jammu and Kashmir, Uttar Pradesh, Sikkim, and South Indian hills. In Uttarakhand hills, a pea is grown in the rainy and winter seasons. Pea straw is also used as nutrition fodder.

Another selected crop Paddy is also an important crop of Uttarakhand agriculture, by good quality of its ability to fix the atmospheric nitrogen in its root nodules thus, rejuvenating the soil health. Paddy offers perhaps the only and most economic means of solving the problems of protein-linked malnutrition. Besides, pulses are also suitable for different cropping systems, multiple cropping systems and rainfed cropping which help in the sustainability of specific cropping systems. The off-season vegetable growers have no strategy for disease management. Some of the progressive farmers use agrochemicals but others do not use any precautionary method for disease management because of the high treatment cost. Unfortunately, a large area of the crop gets damaged every year due to root rot disease causing heavy financial losses to the farmers. Root rot in paddy and pea mostly occurs in fields worldwide. The early symptoms that appear are narrow, longitudinal, and red to brown streaks on hypocotyl, and other portions of the root become progressively more streaked and generally necrotic. Sometimes small, pinched radish brown lesions on hypocotyls and roots are typical symptoms of early disease progression. As these lesions increase in size and become more sunken, they become cancerous, and the red color may predominate until the cankers are old. Root rot is a major soil-borne disease of both crops and is often considered to be the major limiting factor in the production of these crops. The most common pathogens causing this disease are Rhizoctonia solani and Fusarium oxysporum.

Root rot caused by rhizoctonia solani is one of the most serious pathogens of pea and paddy in Uttarakhand hills, causing heavy damage to crop production. The root rot disease causes damage to the root system of the seedling at both pre-and post-emergence stages. Sometimes the affected plants die prematurely without producing the pods thus, reduces the potential yield [115].

Rapid and reliable assessment of vegetation health is important for the management of plant protection. The health of vegetation is impacted by biotic stressors, such as fungal pathogens and insects, and abiotic stressors, such as drought and temperature. In crops, diseases often decrease the yield and therefore cause a loss in profit. In natural settings, the die-off of forest species impacts the ecosystem functioning, including watershed hydrology. Traditionally, trained experts manually check for signs of infection. Manual assessments can be laborious because of the vast number of plants, as well as inaccurate due to the subjectivity of the assessor. Remote sensing, and specifically hyperspectral imaging (HSI), offers the potential for rapid and reliable assessment of vegetation and thus can facilitate large-scale disease detection. HSI has been used since the 1980s, and in the last 20 years, it has gained ground in plant pathology. A more typical multispectral camera measures the reflected light in four to ten broad bands and an RGB camera only in three bands shown in figure 2.4. Hyperspectral cameras continue to find their greatest application as specialized research tools and are not yet broadly employed for practical applications.

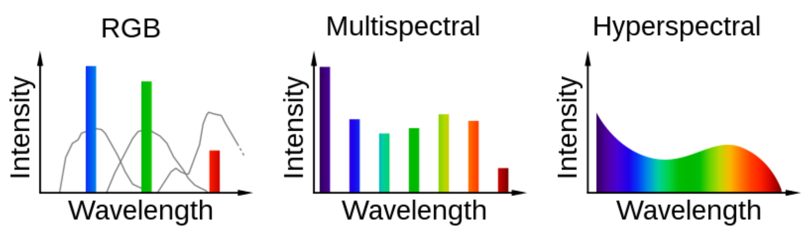


Figure 2.4 Schematic of RGB (red, green, blue), multispectral, and hyperspectral data.

The RGB image is measured in three colors, the multispectral image is typically measured in four to ten colors, and the hyperspectral is typically measured in 100+ bands. A pixel in a hyperspectral image contains a near-continuous reflectance signal. In the rapidly developing field of hyperspectral imaging for phenotyping, little standardization of workflows has been adopted. To analyze the vast data, complex machine learning tools, that need expert input, are used. HSI in real-world applications acquires knowledge from diverse fields such as computer science, remote sensing, plant pathology, forestry, and others. Because of this complex setting, standardized workflows could open HSI to a much broader set of applications. However, many unknowns still need to be investigated to lay the foundation for this endeavor of standardization. We have separated these unknowns into four categories: time, space, machine learning, and plant susceptibility. There is little research on the HSI of vegetation with differences over time and in spatial scale. HSI to identify useful wavelengths to integrate with a multispectral camera, there is a need for high-resolution target detection in controlled lab settings. The translations from proximal scans to further-range field scans remain challenging. Scaling of hyperspectral signatures seems to be possible, but more research is needed to prove the viability of this possibility.

2.6 PSO for Attribute Selection

Many researchers have exploited PSO and SVM in implementing wrapper-based attribute selection methodology. Authors have also proposed variants in particle swarms to improve the efficiency in selecting highly predictive attributes. The aim is to select an optimal attribute subset using PSO and NN. This methodology requires the user to specify the size of the attribute subset for selection which is rather difficult without prior knowledge about the attribute space and is problem specific. The PSO and SVM to build a wrapper-based attribute selection methodology for identifying keystroke dynamic systems. In a designed attribute selection techniques using a particle swarm optimizer and SVM with the one-versus-all method to classify multi-class problems. A wrapper-based attribute selection methodology was developed using PSO and SVM. The proposed methodology traverses the problem space to select an optimal attribute subset while simultaneously optimizing the parameters of SVM. Unlike applied continuous valued PSO for optimizing parameters in SVM and PSO for optimizing attribute subset search [116]. In an adaptive selection strategy to select an attribute subset based on both its likelihood and its influence on other attributes already added to the subset. This strategy shows superior performance to search and scatter search algorithms. The authors also state that the efficiency of the adaptive selection strategy and the quality of the solution can be improved by relaxing the restriction on the size of attributes considered for adding to the subset.

A multi-swarm PSO methodology simultaneously optimizes parameters in SVM and the search for an attribute subset with high predictive accuracy. However, this methodology is computationally expensive as it employs more particles in the swarm for search with complex communication between different sub-swarms. In the process of wrapper-based attribute selection using PSO, the position of each particle is a binary string in Hamming space requiring no boundary conditions. Hence, the problem of divergence does not occur; however, the problem of pre-mature convergence occurs in two cases.

- When Swarmbest in itself is locked in a sub-optimal solution in the problem space - Swarmbest stagnation
- New areas are untried when velocity reaches maximum or minimum stagnates as it corresponds to bit position left unchanged at one and zero respectively-velocity stagnation.

2.6.1 Swarmbest stagnation

Each particle of the population, in search of an optimal solution, adjusts its position depending on its personal best and the best solution chosen on social interaction with neighbors. In conventional PSO, Swarmbest is updated only on achieving a better solution than the previous Swarmbest. If the Swarmbest in itself gets stagnated in local optima, the particle's search area is restricted around the Swarmbest in the problem space. Particles don't explore the entire problem space and limit their search around the local optima. Such a solution provides an attribute subset that may not yield superior predictive results during classification. In addition, PSO tends to converge rapidly during the initial search and retards the convergence rate quite often. As a consequence, particles get trapped in local optimal.

To avoid particle traps in local optimal, Swarmbest is monitored in each run. On stagnation, if Swarmbest remains unchanged for a prefixed number of runs, Swarmbest is reset. Authors have employed operators such as selection, local search, mutation, perturbation in PSO to avoid premature convergence. These operators are either introduced in every run or after a preset number of runs. These methods have improved the particle’s searching ability in finding an optimal solution. Nevertheless, it can be further enhanced when these methods are applied only when the particles stagnate in local optima rather than in every run or after a fixed number of runs. It is proposed a wrapper-based attribute selection technique with PSO and SVM that resets the Swarmbest when it stagnates in local optima for a pre-fixed number of runs. On stagnation, the authors reset Swarmbest such that FoM is zero (no attributes selected).

2.6.2 Velocity stagnation

When the velocity of the particle reaches the maximum, velocity stagnation occurs and the particles stagnate in a sub-optimal point in the problem space resulting in premature convergence. Also, when Swarmbest is trapped in local optima, particles at different points in the problem space converge to local optima with large velocity as the farther the target, the larger the step size required to reach it. Velocity may increase to maximum and particles need to be diverted from the local optimum which is not possible with a small change in velocity. Thus, velocity stagnation becomes predominant on Swarmbest stagnation.

2.7 FPA for Attribute Selection

Very few researchers have employed FPA as the search strategy in wrapper-based attribute selection based proposed FPA and its application to attribute selection. Authors have employed optimum-path forest (OPF) to evaluate the selected subset. Wrapper-based attribute selection technique using FPA and OPF is compared with PSO, Harmony search, and Firefly algorithm. Considering the recognition rate, all the techniques were found similar to each other in their performance and thus authors claim FPA as a potential search technique in the process of attribute selection [117]. A hybrid FPA with a rough set for attribute selection. In the initial stage, the proposed methodology uses mutual information among attributes as FoM, and in the later stages FoM is made more classifier such as rough set dependent. Results are compared with search algorithms such as PSO and GA.

2.7.1 Population initialization

Researchers have proposed several initialization techniques that improve the likelihood of finding a globally optimum solution. Population initialization improves the quality of the solution vector and reduces the computational complexity. Extensive research on population initialization has helped authors in categorizing the techniques based on randomness, compositionality, and randomness.

2.7.2 Initialization strategy in PSO and FPA for attribute selection

Only very few researchers have designed initialization strategies specifically for the combinatorial attribute selection problem. Small initialization, initialize the population with a small number of attributes around 10% of the total number of attributes. Large initialization, initialize particles with a large number of attributes more than half the total number of attributes. However, experimental results show that small initialization that selected a small number of attributes resulted in predictive performance not being as good as large initialization that selected a large number of attributes. Designed mixed initialization by considering the merits and demerits of small and large initialization strategies. In most cases, attributes selected by mixed initialization are larger than small initialization but smaller than large initialization and traditional random initialization techniques. Other researchers have mostly employed traditional random initialization techniques to generate the initial population for the evolutionary process employed in the process of attribute selection [118].

2.8 Stochastic and Deterministic Techniques

Prior information about the search space in any optimization problem is necessary to generate the initial population. Generally, such information about the problem does not exist and researchers employ pseudo random number generator (PRNG) for population initialization. PRNG produces uniformly distributed samples and is simple to implement. However, PRNG fails to produce uniformly distributed points in high dimensional problem space owing to a constraint on population size. On increasing population size, it lessens non-uniformity, but at the cost of computational complexity. A chaotic number generator (CNG) generates a stochastic initial population owing to its characteristics such as randomness, repeatability, and periodicity. But a proper and the best map needs to be chosen to generate a chaotic initial population. Also, it is sensitive to parameter settings.

Unlike PRNG and CNG which produce the initial population based on seeding, defined population generation methodology, despite the initial condition. Deterministic techniques lack randomness and unpredictability but maintain uniformity in producing the initial population considering the entire problem space. It doesn't involve a random element to generate low discrepancy sequences and contradiction in discrepancy measures makes the method less popular among the population initializers. It is a space-filling algorithm that evenly scatters points in problem space & largely depends on parameter setting [119].

2.9 Compositional and Non-Compositional Techniques

The opposition learning algorithm is a widely used multi-step technique to generate the initial population. As a first step, it generates an original population using an initialization technique and in the next step, it generates another population opposite population using a simple heuristic. Finally, a union of the original and opposite population is formed and a subset of the union based on the fitness function is selected as the initial population. Centroidal voronoi tessellation is a multi-step technique that uses metrics to generate an improved initial population rather than using a fitness function. Although multi-step techniques achieve good results, they suffer from high computation costs and largely depend on the quality of the original solution.

2.10 Generic and Application Specific Techniques

Initialization techniques can be either generic or application specific. Techniques discussed so far are generic and can be applied to any optimization problem. These techniques are effective when no prior knowledge about the problem space is available. Domain-specific knowledge is the basis for constructing the initial problem space. Though potentially designed for a specific problem, it lacks generality and requires an expert in the corresponding domain to investigate the technique. Researchers have applied various initialization techniques in PSO to improve its evolutionary search process. Techniques such as the nonlinear simplex method, CVT, space transformation search method, orthogonal array initialization, and opposition based techniques are exploited by researchers for generating initial population in PSO. From the literature survey, it is understood that only a random initialization strategy is employed in FPA for generating the initial population.

2.11 Multidimensional Object Optimization

In a corresponding image analysis pipeline and a combined rice panicle phenotyping system have been developed. The study compared five approaches to counting spikelets and showed that the fast RCNN model attained higher speed and accuracy 0.99. Also, fast RCNN is employed for japonica and indica classification and attained an accuracy of 91%. The presented model provides an advantage for rice breeding and functional genetics. Introduced a DCNN architecture for automated classification and recognition of different abiotic and biotic paddy crop stresses through the field image. Study is conducted by fine tuning the ResNet50 architecture with hyper parameters namely the number of epochs, learning rate, and mini-batch size. The presented method attains a maximum classification performance of 99.26% for fine-tuning ResNet50 [120].

The primary goal is to detect and analyze the disease affecting the plant using ML and image processing methods. The symptom of plant diseases is segmented and identified by color and edge-based image processing approaches. Related feature from the segmented leave portions is extracted and the disease types are categorized by multiple-class SVM. In the plant diseases affected by distinct pathogens like viruses, bacteria, and fungi are examined for periodical and early diagnosis of plant disease. To aim in the design of the grapes plant disease recognition method. The process takes an individual plant leaf as input and segmentation is implemented afterward the removal of the background. Then, the segmented leaf images are examined by a high-pass filter for detecting the infected portion of the leaf. The segmented leaf texture is recovered by fractal-based texture features. Fractal based feature is invariant and thus provides a better texture module. The texture of each disease would be changed. Then, the extracted texture patterns are categorized by multi-class SVM. It is mainly focused on diseases frequently seen in grape's plants that as black rot and downy mildew.

2.12 Research Gaps

New technologies are now playing a huge role in helping farmers elevate their rice production. Farmers are looking for a faster and more reliable solution, a solution where they can easily take feasible action with their diseased rice crops. Most of the current plant leaf disease detection techniques are either classifying if the plant is diseased or identifying the diseased area of the leaf. Recently, most of the research works are presented on ML algorithms for plant disease detection. A challenge here is to use various ML techniques, ensemble learning models exist which can produce generic features for models such as VGG19, XceptionNet, U-Net, ResNet50, DenseNet, SqueezeNet, and CNN to not only classify if the rice plant is diseased or not but to also classify which disease it is and provide more accurate results. But still, there are some limitations.

The author proposed a fine-tuned CNN model on the rice plant disease dataset. The authors created their dataset and evaluated the performance. The data provided is not big enough to allow the usage of a DNN which would provide a more accurate classification [121]. The dataset is considered very small, and images of different rice diseases look very similar, DNN cannot be trained from scratch since the data is very small, and shallow feature extraction produced around 80%. But they didn't provide any comparative analysis of the publicly available datasets. The performance was limited for their proposed dataset. In terms of accuracy, the highest accuracy of approximately 97% was achieved by the VGG16 model but the trainable parameter was approximately 138 million. This results in the high computational cost of the model that is needed to be reduced.

PlantVillage dataset was used. The author designed a hybrid model using the CNN model with an RF classification model. But the limitation of this model is its minimal accuracy rate. The designed hybrid model achieved approximately. 96% of accuracy. In this hybrid model, the voting strategy was used from five different models of Generative adversarial network, LeNet, U-Net, AlexNet, MobileNet, ShuffleNet, and EffNet. The combination of these five models requires 64 million trainable parameters which make its computational complexity very high [122]. The Inception CNN model was proposed and this model achieved approx [123]. This model has 95% of testing accuracy but the limitation is that this model is designed for four types of rice plant diseases only taken from [124].

A hybrid framework system for SAE and CNN models such as basic CNN, VGG16, and ResNet50 to detect rice plant diseases has not been proposed in any state-of-the-art systems present in the literature.

CHAPTER 3
MATERIALS AND METHODS

This chapter is dedicated to designing materials and methodology for rice plant disease detection using a machine learning approach. The methodology is designed to hybridize with the integration of the feature engineering approach with deep learning models.

3.1 Introduction

Throughout this work, a methodology is provided for classifying various rice diseases using imaging technology with deep learning tools as shown in figure 3.1.

Figure 3.1 Disease detection system

3.2 Data Descriptions

Rice diseases as well as pests can be found in various portions of the plant. Temperature, moisture, rain, rice plant wide range, season, nourishment, and other variables all influence their incidence. In this research work, three dataset sources are collected and used for performance evaluation. The details of a dataset are discussed below.

3.2.1 Dataset details

The dataset is taken from the following three sources.

- The first dataset contains 2869 images of rice diseases & it was collected from paddy fields of the Indian Rice Research Institute (IRRI) [125].
- The second dataset contains 14291 diseased rice leaf images of four varieties & it was captured from different rice fields in India from PlantVillage [126-127].
- The third dataset contains 1488 images of rice diseases & it was collected from Kaggle [128].

The gathering of the image was conducted in a variety of conditions of weather, including cold season, hot season, as well as cloudy weather, in sequence to obtain many images that were as relevant as conceivable. The identities of the classifications, as well as the number of images gathered for every categorization, are listed in table 3.1. Some samples of rice diseases are illustrated in figure 3.2.

Table 3.1 Dataset description

Ref Number	Dataset	Number of images
Ref1 [125]	IRRI dataset	2869
Ref2 [126-127]	Plant Village Rice Disease Dataset	14291
Ref3 128]	Kaggle Rice Disease Dataset	2385

Figure 3.2 Various types of rice plant diseases

3.2.2 Data acquisition

In this section, details about all used datasets are given.

3.2.2.1 IRRI dataset

Rice diseases as well as pests can be found in various portions of the plant. Temperature, moisture, rain, rice plant wide range, season, nourishment, and other variables all influence their incidence. The dataset is taken from Indian Rice Research Institute paddy fields (IRRI) [125]. The gathering of images was conducted in a variety of conditions of weather, including cold season, hot season, as well as cloudy weather, in sequence to obtain many images that were as relevant as possible. There are a complete nine classifications in this research work. The identities of the classifications, as well as the number of images gathered for every categorization, are listed in table 3.2.

Table 3.2 Indian Rice Research Institute dataset description [100]

Disease Class List	Number of images
BPH	225
BLB	465
Neck Blast	290

Disease Class List	Number of images
Stemborer	217
Hispa	112
Sheath Blight and Sheath Rot	227
Brown Spot	478
Healthy	437
False Smut	418
Total	**2869**

3.2.2.2 Kaggle rice disease dataset

This dataset was collected from a publically available resource, kaggle including 2385 on-field images of three types of rice leaf diseases and healthy samples [128]. The details of rice disease classes are presented in table 3.3.

Table 3.3 Kaggle rice disease description [115]

Diseases	No.of images
Stemborer	202
Brown Spot	478
Bacterial Blight	495
Leaf Blast	487
Tungro	223
Healthy	500
Total	**2385**

Moreover, a countable number of raw images were directly taken through research scholars for real-time execution. These images are captured with various resolutions, various conditions, various environments, and various stages of rice crops.

3.3 Methodology

In this section, different methodologies and algorithms adopted for this research work are presented in detail. The flowchart of the proposed methodology is presented in figure 3.3. All steps are discussed in detail in the subsections.

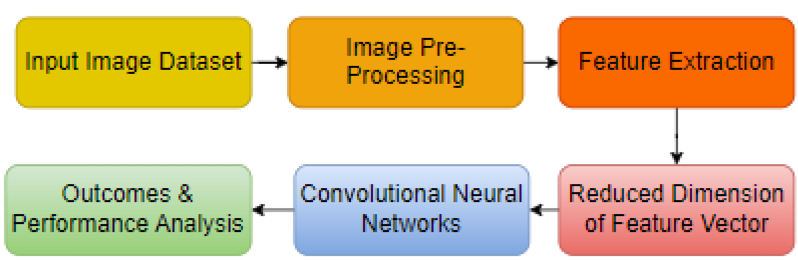


Figure 3.3 Overall image processing flow chart

3.4 Model Developments and Methods

First, the image acquisition process collected all field and crop-oriented images and posted them to pre-processing. Then, the image pre-processing method helps segment creations to design pixel masking of every object which is present in the image. It means an image is converted into a suitable image. In simple explanation pre-processing supports image smoothening, transformation, and image reduction to improve image quality for better object detection. Pre-processing is playing an important role in image processing; it increases the probability of detecting objects accurately. Image enhancement calculating noise ratio level, dimensional value, HSV, smoothening, grey scale value, size, brightness values, sharpening images. Moreover, at the time of pre-processing data reduction supports reducing unnecessary noise and information from images. After that, image restoration reduces noise and blurs orient things from an input image. This process increases the appearance of the rice crop image with better quality, which increases the accuracy of the rice plant disease detection and prevention process shown in figure 3.4.

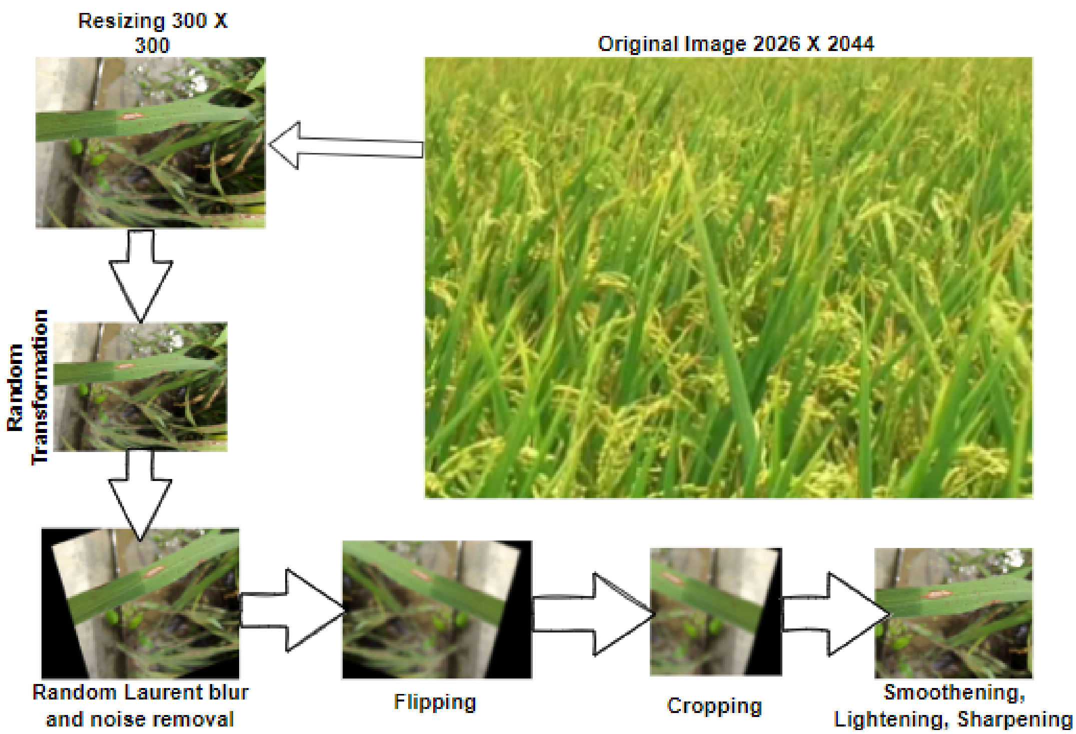


Figure 3.4 Data acquisition process (avoiding overfitting and underfitting)

3.5 Features Extraction and Segmentation

After the image enhancement and restoration process then segmentation, this is supported by offering high dimensional segmented information. Segmentation is partitioning which makes it easier for the image reader to understand the object boundaries. For the rice plant disease detection model, a cognitive network is provided to detect the boundaries range of disease information collected from each segment of crop images. Segmentation is partitioning the images to enable the user to detect the object and extract the important feature which decides the types of disease and disease stages using a cognitive network. Moreover, the cognitive expert provides support to decide on information extracted from each pixel. Divide and conquer method provisions to cognitive experts, operational images convert it into multi segments such as partitioning and extract the best fit in it. Image classification of every pixel and its values is based on cognitive decision-making with a pre-determined group. A cognitive residual convolution network (CRCNet) is contained in the expert's direction of encoder and decoder with feature learning and feature vectors. Also, the cognitive network creates intellectual segments to improve the accuracy of a disease detection system. Consider the intellectual segments created from input crop images.

$$IS = \{IS_1, IS_2, \ldots, IS_e, \ldots, IS_f\} \tag{3.1}$$

where f denotes overall segments surviving on input rice plant image, IS_e represents e^{th} segment of the input.

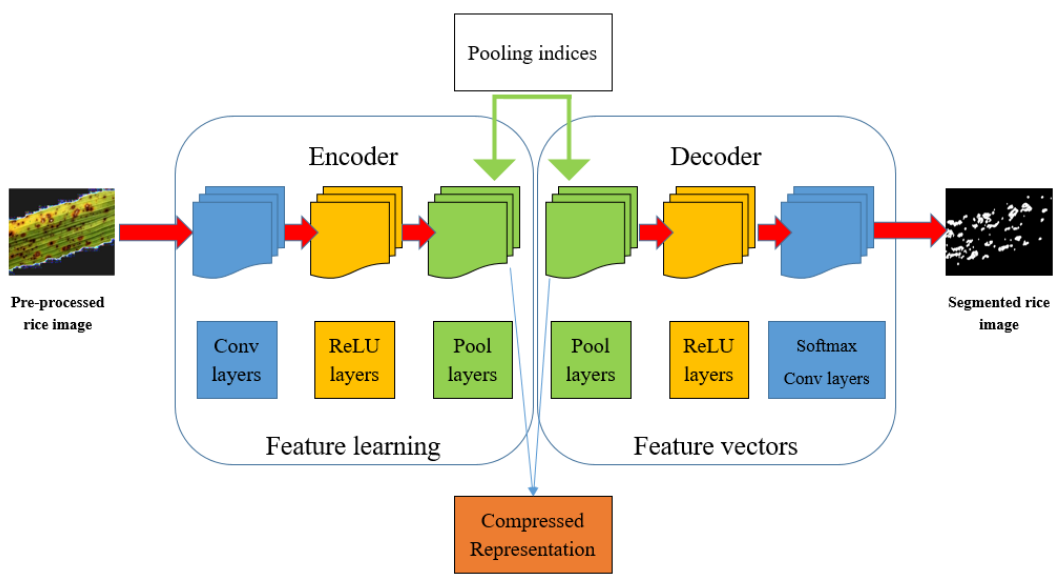


Figure 3.5 CRCNet flow with feature machine learning and feature vectors for image segmentation

Figure 3.5 shows the flow of the cognitive network with feature learning and feature for image segmentation. Feature learning encoder includes convolutional layers, ReLU layers, and pool layers-based segmentation avoiding the curse of dimensionality issues. ML starts its process of training sets creation with weights values (w). The feature learning process consists of corresponding feature vector rates for the reconstruction of the segmented image. Figure 3.6 shows, convolutional layer converts the images (300X300) with RGB values (300X300X3) and then continuously does convolution processes with filtering, slicing, pooling, and padding. After pooling values send to feature vectors with softmax class values to create the probabilities of multi classes for image pixels such as for both types of pixels. Next, an element-based CNN activation function of ReLU is adapted with sub-sampled such as linear and non-linear based activation. ReLU is used to confirm the efficiency and implication of the ReLU layer when maintaining a set of vast networks and their weights. In pool layers, no need to train bias and neuron weight values this process decreases the train data complexity.

The max pooling and padding process starts the convolution with different slicing(s) and integrates regional filters. Then the reverse process of feature learning is supported to sub-sample procedural feature maps with trained map values and trained through maximum pooling. Feature mapping methods are convolved with comfortable feature vectors and filter to increase the map's denseNet. At last, creates final feature vector values which expose the final vector class values to the probability vector. Softmax classifier is a self-governing classifier used to segregate image pixels. The final result created by softmax classes is SC, which represents the number of softmax classes. An estimated segmentation boundary with SC and it provides a high probability vector rate for every pixel. The CRCNet supports segmentation slicing and padding for generating segments. CRCNet is effective as it takes filters, slice rate, max pooling, feature maps, and vectors used for better accuracy.

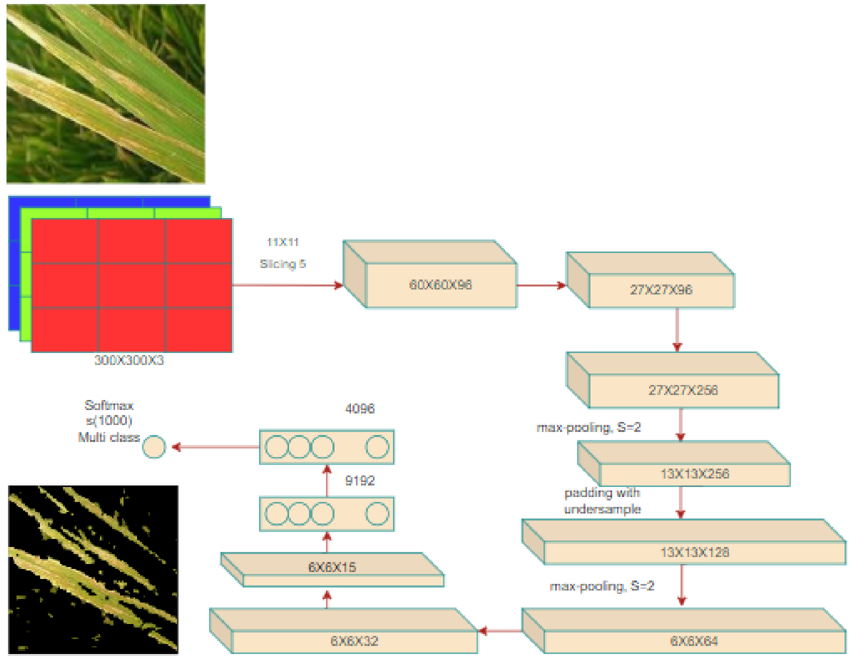


Figure 3.6 Feature learning with MOCLEAR segmentation algorithm

3.6 Image Feature Extractions

One receives the segments of a particular image and then has to start to acquire features from the segmented image. The extracted features are categorized into various types which helps to classify crop disease identification accurately and quickly. Here we were calculating various types of features and their extracted values. They are entropy, inverse difference moment, angular sec moment, LGP (local gradient pattern), and variance. For rice plant disease detection mainly collecting GLCM, statistical and texture features to make results betterment and improve detection accuracy [135]. Feature extraction method calculated the above-mentioned feature values from an input image and its description following eight steps:

- Mean: Mean values calculated with convolutional-based numbers pixels. The mean value (T_1) calculated from the sum of segments present in input is divided by the total values of segment values.

$$Mean_{seg} = \frac{1}{\left|d(S_f)\right|} \times \sum_{f=1}^{\left|d(S_f)\right|} d(S_f) \tag{3.2}$$

  where f represents the total segment in an input image, $d(S_f)$ denotes single-segment pixel rates, $\left|d(S_f)\right|$ is the cumulative value of pixels present in the segment.

- Variance: It is calculating pixel spreading rate with mean values $Mean_{seg}$.

$$Variance_{seg} = \frac{\sum_{f=1}^{IS} * \left| S_f - \frac{1}{|d(S_f)|} \times \sum_{f=1}^{|d(S_f)|} d(S_f) \right|}{(IS)} \quad (3.3)$$

- Standard deviation: SD supports descriptive statistical measures; an advanced variance is SD which estimates random deviation expressed as an amplitude rate.

$$Standard\ Deviation_{seg} = \sqrt{\frac{\sum_{f=1}^{IS} * \left| S_f - \frac{1}{|d(S_f)|} \times \sum_{f=1}^{|d(S_f)|} d(S_f) \right|}{(IS)}} \quad (3.4)$$

- Skewness: from variance collected amount of variability and through skewness provide direction of V. Skewness is the lack of symmetry for frequency distribution such as the affected level of the plant leaf and its boundary sharp using a numerical value.
- Kurtosis: Kurtosis is measuring outliers and lacks outlier values. Kurtosis provides sharpened values of probability and its deviation like peak, saturation, and low.
- Entropy: $Entropy_{prob}$measures the count of bits needed for encoding image data conditional entropy. It provides the value of complexity in a certain portion of an image. It measures that is operated to find certain and uncertain information about an image. These value castoffs between mutual data in every entropy. Entropy calculates the variation between adjacent pixels or clusters such as segments. Additionally, it is defined as a corresponding complexity level based on trained values. Entropy image segments are computerized with analytical quantities to get image and pixel information. Also, entropy calculates an actual probability of hit rate and miss rate. Later, entropy is appraised to attain the probability of intensity level.

$$Entropy_{prob} =- \ Q \log log \ (\ Q) \quad (3.5)$$

Where Q denotes the probability of encoding image information and discrete value.

- Local Gradient Pattern:

LGP acclimates the filter value and curve representation of gradient patterns in a group of pixels [42] in current research. Local gradient pattern calculations,

$$T_7^{q,l} = LGP_{\delta,\alpha}^{u^l,v^l} = \sum_{n=0}^{\beta-1} \eta \left[\bar{\mu}_n - \bar{\mu}^* \right] \times 2^n \quad (3.6)$$

$$\eta(s) = \{0 \quad ; \quad s < 0 \ 1 \quad ; \quad otherwise \quad (3.7)$$

where μ^* and $\bar{\mu}_n$ denotes mean gradient and center pixel point $\left(\lambda^*\right)$ situated at $\left(u^l, v^l\right)$ it's neighboring pixels (λ) shows as,

$$\bar{\mu}_n = \left|\lambda - \lambda^*\right| \tag{3.8}$$

Where λ^* dynamic threshold fixing with cognitive values for code generation. The range of color values was plotted and it's expressed as [3×γ] 'γ' represents bins count.

- GLCM statistical texture feature:

  GLCM defines statistical 2nd order values and texture features are developed by distinctive greyscale input images using location, position, and spatial information. Each pixel rate calculates co-occurrence combination frequencies for improving the quality of feature extraction. GLCM defines combination frequencies of all possible occurrences that are calculated with image grey-level value. GLCM is used to maintain the pixel position and explicit pixel location stored through the two-dimensional (2D) array.

$$GLCM_{feq} = P(h_1, h_2, ..., h_t) = \frac{A_m(h_1, h_2, ..., h_t)}{\sum_{z_1}^{N_h-1} \sum_{z_2}^{N_h-1} A_m(h_1, h_2, ..., h_t)} \tag{3.9}$$

Where N_h represents the total grey levels in an input image, it is calculated with t^{th} order GLCM statistical feature rate. z_1, z_2 denote the first and second position and location of gray level values. A_m describe the angular moment and inverse difference moment of every pixel value. $P(h_1, h_2, ..., h_t)$ represents a population of explicit pixel positions.

3.7 Implementation of Image Decoding (feature vectors)

In this image decoder section attained each group of explicit pixels or segments is mean, variance, standard deviation, skewness, kurtosis, entropy, local gradient pattern, and grey level co occurrence matrix. In equation 3.9, a set of important features are exposed from the input image they are GLCM, statistical, and texture image features. Equation 3.10 image decoding is the processing of received feature values from the image encoder which provides proper information to take a final decision to predict crop diseases. Unique attributes and vital attributes (features) are having max prediction rate to increase the accuracy of the optimization algorithm and cognitive learning algorithm by doing image enhancement, restoration, and image classification. This extraction process receives values from each segment present in an input image.

$$Img\_dec = \left\{Mean_{seg},\ Variace_{seg},\ SD_{seg},\ skew_{shar}p,\ kurtosis_{peak},\ Entropy_{prob},\ LGP_{centre\_value},\ GLCM_{feq}\right\} \tag{3.10}$$

Image decoder Img_dec inserts values into GAN, which classifies with statistical texture values, which softmax classes are trained. The rice crop is segregated by the classifier as a level of infection occurs on the rice plant in the corresponding input image.

3.8 Pattern Recognition

Patterns the language of nature is how a man interprets the world. Human beings recognize objects around them and act concerning them. Pattern recognition by machines involves designing an intelligent system to perform certain tasks with skills comparable to that of humans. From a broader perspective, pattern recognition is the science that deals with the characterization and recognition of patterns. A pattern is a set of characteristics inherent in a sample and recognition is the ability to classify it. The process of pattern recognition is shown in figure 3.7. A set of measurements raw data is obtained from the real-world pattern recognition problem. Attributes, together characterizing and discriminating the patterns are then extracted from the measurement set, to train a classifier for identifying the unfamiliar pattern class.

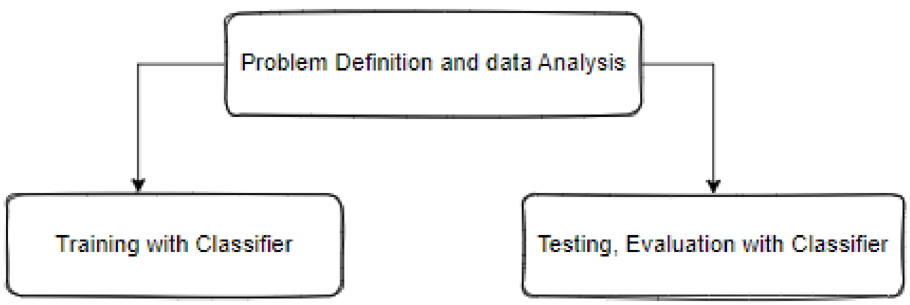


Figure 3.7 Process in pattern recognition

3.8.1 Dimensionality reduction

Generally, a comprehensive set of attributes is drawn from the measurement space to allow better characterization of a given pattern recognition problem. Nonetheless, it can also introduce redundant and irrelevant information resulting in a high dimensional attribute space making classification a complex task. Hence to battle the problem of the curse of dimensionality in real-world pattern recognition problems, attribute space is either reduced or transformed to a manageable size without discarding the vital information about the problem. The dimensionality reduction technique aims to devise a computationally feasible attribute space yielding better predictive performance than with the comprehensive set of attributes. The mathematical measure for building such an attribute space is generally classified into two types of processes such as,

- Calculating attribute selection and attribute problem space

$$x \rightarrow x'$$

$$|x| = D; |x'| = D'; D > D' \quad (3.11)$$

- Attribute extraction in the transformed space

$$x \rightarrow y' = f(x) \quad (3.12)$$

$$|x| = D; |y'| = D'; D > D' \quad (3.13)$$

Attribute selection, from equation 3.12 reduces the dimensionality of attribute space by selecting only highly predictive attributes. It is the process of selecting a subset of attributes from $X\ (X_1,\ X_2, ..., X_D)$ where D represents dimension problem space. Attribute extraction, from equation 3.13 utilizes all the attributes in the measurement space and maps a dimensional vector from high to low. Attribute selection and extraction are a process of choosing the best attributes for classification and best prediction form y' = f(x) by which a sample, $X\ (X_1,\ X_2\ ..., X_D)$ where D represents dimension problem space values converted into$y'(y'_1, y'_2, y'_3, ..., y'_{D'})$ where D' average dimensional attribute space.

Attribute extraction transforms the entire attribute space to achieve dimensionality reduction. It may result in a loss of originality of attributes, obtained from the measurement space. In addition, it does not portray the attributes responsible for the superior performance of the classifier, thus failing to map the physical significance of attributes with the pattern recognition problem. On the other hand, attribute selection, apart from dimensionality reduction, selects highly predictive attributes in the original form by rejecting attributes that do not contribute to the superior performance of the classifier. It provides a better understanding of attributes that characterize the problem space. Thus, attribute selection plays a vital role in reducing the dimensionality of attribute space that is fed as input to the final decision space. The modified pattern recognition process with attribute selection as its important constituent is shown in figure 3.8.

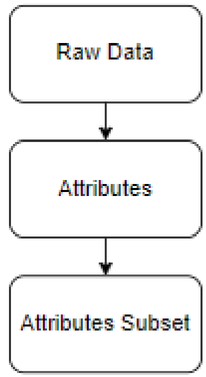


Figure 3.8 Process in modified pattern recognition

From the figure 3.8, it is observed that highly predictive attributes from the problem space are only transferred as input to the classifier. Such a classifier is computationally simple and results in superior predictive performance.

3.8.2 Attribute subset selection

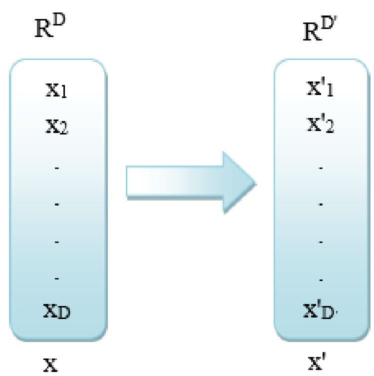


Figure 3.9 Attribute selection (Presence or absence of an attribute in the subset)

Attribute selection is a combinatorial problem such as presence or absence of an attribute in the subset and demands a d-dimensional hamming space with the 2D combination of attributes as elements in the problem space. The attribute selection process iteratively looks for an optimal attribute subset from the problem space by evaluating its figure of merit (FoM) till a stopping criterion is reached. Thus, the basis of attribute selection is subset search and subset evaluation.

3.8.2.1 Need for attribute selection

Attribute selection is a vital task in any pattern recognition problem. It is observed as a dimensionality reduction technique. Attribute selection satisfies some or all of the following needs:

- To enhance the predictive performance of classifiers employed in the final stage of the pattern recognition problem.

- To enable the development of a faster and computationally simple classifier.
- To better understand the nature of highly predictive attributes involved in describing the pattern recognition problem.

3.8.2.2 Significance of attribute selection

Attribute selection is a non-deterministic polynomial (NP) hard problem. Nevertheless, it accelerates the processing rate and favors a less computationally intensive training model for classification. Also, the predictive accuracy, comprehensibility and generalization ability of a classifier tend to improve on incorporating attribute selection in the process of pattern recognition. Attribute selection techniques depending on subset evaluation are classified into two categories:

- Filter-based technique
- Wrapper-based technique

The filter method ranks every attribute to select or unselect it for the final subset using a simple criterion.

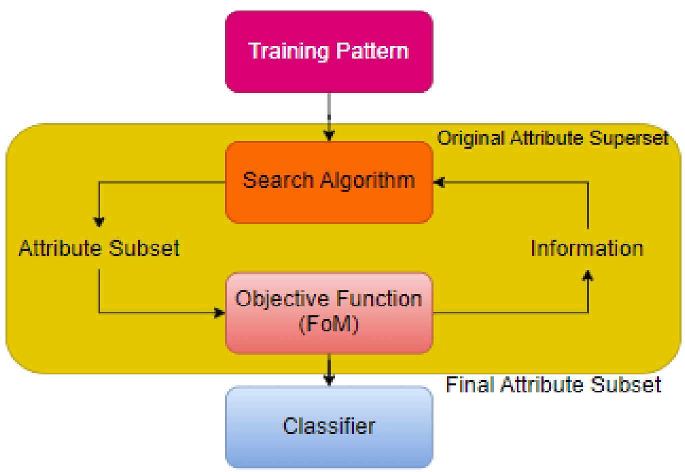


Figure 3.10 Filter-based attribute selection

Filter-based and wrapper-based attribute selection is explained in figure 3.10 and figure 3.11. Both the attribute selection techniques evaluate the FoM of the attribute subset selected using a search strategy. From figure 3.10 and figure 3.11, it is clear that filter and wrapper-based approaches differ only in the way attribute subsets are evaluated. The filter approach employs measures like mutual information and correlation for evaluating the attribute subset without using any learning algorithm like a classifier.

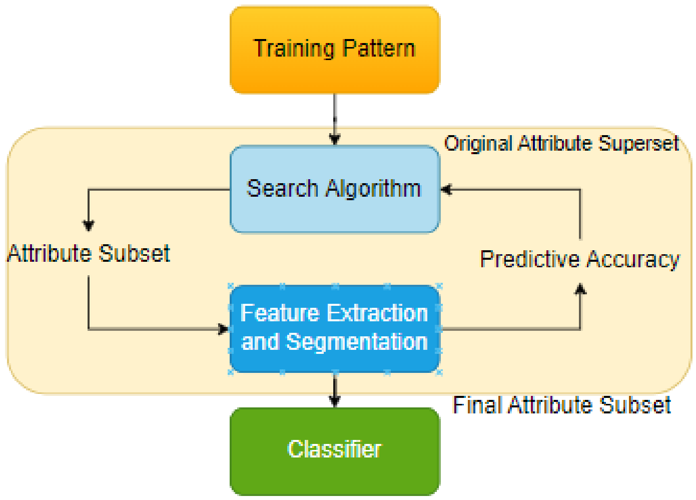


Figure 3.11 Wrapper-based attribute selection

Filter-based methods involve non-iterative computation of attribute subsets and are more general. However, a user has to fix the attribute subset size owing to monotonic criterion functions in the filter approach. Alternatively, the wrapper-based approach exhibits superior predictive performance compared to a filter-based method. Nevertheless, there are a few disadvantages to using wrappers.

- Wrappers are slow in execution
- Lack of generality in wrappers.

However, these issues can be resolved by employing global optimization methods for search and k-fold cross-validation respectively. Thus, the basis of research lies in developing a wrapper-based attribute subset selection methodology using a randomized global search algorithm.

3.9 Wrapper-based Attribute Selection

The wrapper-based methodology employs various learning algorithms to calculate the predictive performance of an attribute subset selected using a search algorithm. First random attribute subset selection meanwhile conditional iteration checks the criteria. If means based on learning model attributes gets wrap and FoM then search the next attribute subset. The general procedure for wrapper-based attribute selection involves four important steps:

- Initialization
- Subset generation
- Subset evaluation

3.9.1 Initialization

The problem of attribute selection demands a D-dimensional such as attribute count. All possible combinations of D attributes result in 2D attribute subsets as possible vectors in the Hamming space. Initially, an attribute subset is chosen randomly from the problem space to begin the process of attribute selection.

3.9.2 Subset generation

Subset generation demands a search technique for selecting an attribute subset from the problem space. Randomized algorithms such as PSO, ACO, cuckoo search, tabu search, FPA, and bat algorithm are exploited by researchers as a search strategy for attribute subset selection.

3.9.3 Subset evaluation

The success of the subset search is measured by evaluating its FoM. It is calculated as a tradeoff between the two essentials in search of a potential solution.

- Goodness of the attribute subset
- Size of the attribute subset

3.10 Figure of Merit

To measure the goodness of the attribute subset generated during the process of search, FoM to be optimized is evaluated. Hence, a classifier is employed to measure the predictive accuracy of the selected attribute subset and in turn it's FoM.

A classifier partitions the attribute problem space into small regions. An unknown pattern can then be classified by identifying the region in which it lies. The goodness of the attribute selected is measured in terms of predictive accuracy (P), learning model, and k-fold cross-validation to classify an unknown pattern using equation (3.13).

$$P = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.12}$$

Where $TP - True\ Positive,\ TN - True\ Negative,$

$$Solution = FP - False\ Positive\ FN - False\ Negative \tag{3.13}$$

k-fold-based cross-validation method was employed to achieve better generalization for a classifier. The entire dataset is divided into k subgroups so that each subgroup tests the classifier trained with (k-1) subgroups of instances from the dataset. Finally, the predictive accuracy of the classifier from k cross-validations is averaged. FoM is then calculated as a linear combination of predictive accuracy (PA) and count of selected attributes ratio differ from original attributes in the dataset using equation 3.14.

$$FoM,\ f = a(p) + (1 - a)(1 - \frac{N_{selection}}{N_{total}}) \tag{3.14}$$

Where a is the weighing factor that decides the trade-off between the two objective functions, $N_{selection}$ is the attribute subset size and N_{total} is the attribute space of the problem size [129].

3.10.1 Classifier

Wrapper-based attribute selection induces a learning algorithm to judge the attribute subset selected. The predictive performance of the classifier with a set of instances characterizing the attributes is measured to evaluate the merit of the selected attribute subset. SVM classifiers exhibit good generalized performance and rank top against 16 classifiers using 21 different data sets as calculated. Hence, SVM is chosen as the learning model to pattern a wrapper-based attribute selection methodology. SVM maps inputs in problem space R^D to high dimensional space R^K (K >D) such that it is linearly separable. A linear discriminant function for a two-class problem in R^D attribute space is a hyperplane given by equation 3.15 and equation 3.16.

$$g(x) = m^T x + m_0 \tag{3.15}$$

$$g(x) = \sum_{i=1}^{N_s} \lambda_i y_i K\left(x_i, x\right) + m_0 \geq 0 \tag{3.16}$$

m = [m_1, m_2,…, m_D] is the weight vector to be computed to define the decision hyperplane and m_0 is the threshold. N_s is the number of support vectors being the Lagrange multiplier and $K\left(x_i x\right)$ is the kernel function. For each$\left(x_i\right)$, a corresponding class indicator is y_i (+1 for class1 and -1 for class2). Assign x in $class_1$ if

$$g(x) = \sum_{i=1}^{N_s} \lambda_i y_i K\left(x_i, x\right) + m_0 \geq 0 \tag{3.17}$$

Assign x in $class_2$ if

$$g(x) = \sum_{i=1}^{N_s} \lambda_i y_i K\left(x_i, x\right) + m_0 < 0 \tag{3.18}$$

Radial basis function (RBF) as in equation 3.19 is the widely used kernel owing to its ability to analyze higher dimensional data and requires only standard deviation of gaussian distribution, σ as the parameter to be specified.

$$K\left(x_i, x\right) = exp\left(-\frac{\|x_i - x\|^2}{\sigma^2}\right) \tag{3.19}$$

3.11 Termination

The process of subset search and evaluation iterates repeatedly by traversing the problem space till a termination criterion is met. Pattern recognition and attribute selection finalizing stage based on the following factors.

- When the search completes
- On reaching a pre-defined maximum number of runs
- On reaching minimum classification error rate.

An important motivation for research in this field lies in designing intelligent machines to mimic the tasks that human beings do, thus, generating future computing systems. However, the ease with which human beings recognize and characterize patterns often leads to a misconception that this capability is easy to automate. Since the ultimate goal of pattern recognition using intelligent systems is to distinguish and recognize different types of patterns, it is logical to study the basis for this discrimination ability. Different pattern classes are composed of different attributes or attributes with different numerical values capable of characterizing and distinguishing the patterns and an intuitive solution lies in identifying them. With no prior knowledge about the problem, the larger number of attributes allows a better description of the problem. Nevertheless, a larger number of attributes doesn't always guarantee a higher classification rate. Also, the relevancy of an attribute may diminish when combined with other attributes such as epistasis or attribute interaction, thus decreasing the overall predictive performance during classification.

A weakly relevant attribute, by concept, can improve the predictive performance if it combines with other attributes. Similarly, an individually relevant attribute can reduce the predictive performance of the classifier on working with other attributes. Hence, selecting highly predictive attributes from a comprehensive set of attributes, and characterizing the pattern recognition problem is challenging. On the other hand, selecting only highly predictive attributes enables the design of a simple classifier with improved predictive performance. Attribute selection is an indispensable pre-processing technique for the final decision phase. In some cases, mathematical tools aid in attribute selection, while, in other cases, a simulation may help in identifying appropriate attributes. Wrapper-based attribute selection, compared to the filter approach of attribute selection, anticipates the advantages of adding or removing a particular attribute to the subset in terms of predictive performance. The wrapper approach demands a learning algorithm and a search algorithm to identify an optimal attribute subset from the problem space. The SVM is employed as the learning algorithm. It is a methodology for learning functions that represent two-class pattern recognition problems. The SVM provides adequate generalization to unknown patterns and requires only a few parameters to implement. In the attribute selection problem, search space increases exponentially with an increase in the number of an attribute 2D, for D attributes in D-dimensional attribute space making exhaustive analysis of every possible attribute subset impractical. Hence, an evolutionary search is more favored than an exhaustive evaluation of all different combinations of attributes for subset selection. PSO developed is a stochastic search algorithm with swarm intelligence that derives biological inspiration from the swarming behavior of birds, bees, and fishes. PSO is an effective and efficient randomized global search technique. Also, PSO is easier to implement and has only a few parameters for tuning. These advantages mark PSO as a more desirable random search strategy in implementing the wrapper-based attribute selection technique.

FPA is simple and requires very few parameters to implement. Also, it exhibits good exploring capability in finding a potential solution. But not many studies have been made on FPA as a subset identifier in the problem of attribute selection. As discussed in section 3.2, attribute selection demands a discrete problem space, as the elements of the problem space are attribute subsets specifying the status of each attribute in the subset. Researchers have approached this problem of pre-mature convergence by discarding the current best solution if it remains stagnated for a pre-fixed number of runs. An innate solution to this problem in literature is the re-initialization of the best solution to a random position in the problem space. Re-initialization mechanisms, on stagnation, have been used in PSO and FPA by various authors for improved performance in search of an optimal solution. But, none of the methods in the literature is well-defined and is only a random jump to a different position in the problem space. A well-defined heuristic-based re-initialization strategy can further direct the search toward the objective of finding a subset with a reduced number of attributes with high predictive performance. This research background lays the inspiration to develop re-initialization strategies with alternate random re-initialization. PSO and FPA employed for subset search in the process of attribute selection are population-based stochastic search algorithms. Any population-based evolutionary algorithm begins the process of search with population initialization. Good population initialization in an evolutionary search algorithm increases the likelihood of identifying potential solutions from the problem space. Also, population size greatly influences the computational time required for completing the search process. Most of the researchers have employed random initialization. Moreover, population size is fixed based only on simulation. An important motivation for designing a population initialization strategy is to maintain diversity in an initial population such that the initial solution vectors represent every dimension of the problem space and to decide the population size.

Attribute selection gives the best solution to resolving the pattern recognition problem. Still, the pattern recognition problem is a challenging task for researchers considering the large problem space and interaction between the attributes. The overall objective of the thesis is to design a wrapper-based attribute selection methodology, to select an attribute subset with reduced attributes capable of achieving better predictive performance than with all the attributes defining the pattern recognition problem. In specific, this research aims to improve the capability of the biologically inspired subset search algorithms (PSO and FPA) to identify an optimal solution from the problem space by re-directing the individuals in the population on encountering local optima in the problem space.

To achieve this objective, the following goals are established to guide this research in attribute selection. Design, simulate, and analyses a new wrapper-based attribute selection methodology using PSO with Greedy reset and localized random mistakes to restructure the best solution identified by the candidates of the population when it remains unchanged for a pre-fixed number of runs and on velocity stagnation respectively. The proposed algorithm is anticipated to select an attribute subset with high predictive performance and a reduced number of attributes. Investigate whether FPA with its global walk capability, avoid individuals in the population from getting stuck in local optima for best solutions. Develop, simulate, and analyses a new wrapper-based attribute selection technique using FPA with Greedy crossover on pre-mature convergence to identify an attribute subset with reduced size and superior predictive performance.

Develop a new wrapper-based attribute selection methodology using a flower pollination-inspired PSO algorithm to address the problem of partial optimization. The proposed attribute selection methodology is expected to choose better-performed attributes. It is also expected to show superior predictive performance than the existing PSO-based attribute selection algorithm. Devise a new population initialization strategy labeled One to all random initialization to generate the initial population for PSO and FPA for diversified exploration in all the dimensions of problem space.

Apply the three proposed wrapper-based attribute selection methodologies along with a population initialization strategy in the process of solving a pattern recognition problem in various types of disease in various environmental conditions.

CHAPTER 4

SWARM INTELLIGENCE OPTIMAL CLASSIFICATION THROUGH COGNITIVE ATTRIBUTE SELECTION

4.1 Introduction

Cognitive-based attribute selection methodology employs an evolutionary search algorithm to identify an optimal attribute subset from the problem space. Evolutionary search algorithms are population-based metaheuristic techniques. The outbreak of diseases in rice leaves affects agricultural productivity. It is a threat to food safety and security. For rice leaf diseases, the distinction between brown spots and the narrow brown spot is challenging as the latter resembles the former except that the disease spot is long, narrow, and distinctly more linear. The problem of disease distinction is critical as the treatment mechanism for crop protection varies for brown spot and narrow brown spot diseases. Machine vision proves to be an alternate and promising solution in disease identification to the conventional practice of visual inspection by agricultural experts and farmers. Wrapper-based attribute selection employs a search technique to select an attribute subset on traversing the problem space, which optimizes the criterion, FoM. FoM is evaluated using a classifier with selected attributes as input. A SIOC-based attribute selection methodology to select an optimal attribute subset that achieves improved predictive performance with minimal attributes than that with the complete set of attributes.

Population Initialization: Evolutionary search algorithms draw inspiration from the process of biological evolution through mutation, crossover, selection, and reproduction. These operators demand a population of individuals for the process of evolution. Thus, population-based evolutionary algorithms, as a first step, generate a population of initial solution vectors that are iteratively improved during the search across the problem space. On the other hand, a bad initial population may forbid the search technique from finding the potential solution in the problem space. Thus, population initialization is an important process of evolutionary search and it is necessary to develop a strategy to generate the initial solution vectors of the population [131].

4.2 Literature Survey

Researchers have proposed several initialization techniques that improve the likelihood of finding a globally optimum solution. Population initialization improves the quality of the solution for the problem vector and reduces the computational complexity. Extensive research on population initialization has helped authors in categorizing the techniques based on randomness, compositionality, and randomness.

a) Stochastic and deterministic techniques

Prior information about the search space in any optimization problem is necessary to generate the initial population. Generally, such information about the problem does not exist and researchers employ pseudo random number generator (PRNG) for population initialization. PRNG produces uniformly distributed samples and is simple to implement. However, PRNG fails to produce uniformly distributed points in high dimensional problem space owing to a constraint on population size. On increasing population size, it lessens non-uniformity, but at the cost of computational complexity. A chaotic number generator (CNG) generates a stochastic initial population owing to its characteristics such as randomness, repeatability, and ergodicity. But a proper and the best map needs to be chosen to generate a chaotic initial population. Also, it is sensitive to parameter settings. Unlike PRNG and CNG (stochastic techniques) which produce the initial population. Deterministic techniques lack randomness and unpredictability but maintain uniformity in producing the initial population considering the entire problem space.

There are few deterministic techniques that generate the initial population in evolutionary algorithms. QRS doesn't involve a random element to generate low discrepancy sequences and contradiction in discrepancy measures makes the QRS method less popular among the population initializers. Uniform experimental design (UED) is supported for evenly scattering the points in the problem space and largely depends on parameter setting [132].

b) Generic and application-specific techniques

Initialization techniques can be either generic or application specific. Techniques discussed so far are generic and can be applied to any optimization problem. These techniques are effective when no prior knowledge about the problem space is available. Application-specific techniques are applied to specific real-world problems. Domain-specific knowledge is the basis for constructing the initial problem space. Though potentially designed for a specific problem, it lacks generality and requires an expert in the corresponding domain to investigate the technique. Researchers have applied various initialization techniques in PSO to improve its evolutionary search process. Techniques such as the nonlinear simplex method, space transformation search method, orthogonal array initialization, and opposition based techniques are exploited by researchers for generating the initial population in PSO. From the literature survey, it is understood that only a random initialization strategy is employed in FPO for generating the initial population.

4.3 Pattern Recognition Problem

Rice yield across the world as stated by food and agricultural organization (FAO) is approximately 3.6 thousand tons per hectare in 2018. In the past few decades, rice yield per capita has been progressive across most of the rice-cultivated regions. Yet, one in every nine people in the world population is undernourished failing to meet the least dietary energy requirement (FAO 2021). Fundamental challenges such as unpromising climatic changes, shrinking natural resources such as water, fertile agricultural land, and expensive labor impede the growth of rice production. New improved varieties of rice by genetic modifications compelled farmers to use improved cultural practices to increase rice production; however, it also amplified the susceptibility of the rice crop to diseases. Plant disease is an abnormal physiological functioning of plants caused by pathogens such as fungi, bacteria, and viruses. About 40 percent of crops are destroyed worldwide by diseases. As most of the disease symptoms are visible on the leaf, the region of analysis in identifying the disease type is the leaf, rather than the whole plant. Rice leaf diseases that decrease yield and productivity are (i) rice leaf blast, (ii) brown spot, (iii) narrow/wide brown spot rice disease, (iv) bacterial leaf/root/collar blight disease, (v) bacterial leaf streak, (vi) tungro viral disease, and (vii) leaf scald.

4.3.1 Initialization Strategy in PSO and FPO for Attribute Selection

Only very few researchers have designed initialization strategies specifically for the combinatorial attribute selection problem. In a proposed three initialization strategies for PSO in identifying optimal feature models and compared them with the influence of proposed techniques against traditional random population initialization. Small initialization, initialize the population with a small number of attributes around 10% of the total number of attributes. Large initialization, initialize particles with large attributes counts such as more than half the total attributes count. However, experimental results show that small initialization that selected a small number of attributes resulted in predictive performance not being as good as large initialization that selected a large number of attributes. Hence designed mixed initialization by considering the merits and demerits of small and large initialization strategies. In most cases, attributes selected by mixed initialization are larger than small initialization but smaller than large initialization and traditional random initialization technique. Other researchers have mostly employed traditional random initialization techniques to generate an initial population for the evolutionary process employed in the process of attribute selection.

Cognitive-based attribute selection employs a search technique to select an attribute subset on traversing the problem space, which optimizes the criterion, FoM. FoM is evaluated using a classifier with selected attributes as input. A PSO-based attribute selection methodology to select an optimal attribute subset that achieves improved predictive performance with minimal attributes than that with the complete set of attributes. Nevertheless, particle swarm is a partial optimizer Greedy reset with localized random mistakes is devised, to divert the trapped particles from discovering untried areas in the problem space.

4.4 PSO: Particle Swarm Optimization

PSO works with solutions discovered by particles in the swarm on traversing. Velocity directs the flight of the particle in the problem space in search of an optimal solution. The particles then move from one position to another in the problem space according to their current velocity derived in algorithm 1. In the course of a search, the velocity of the particle needs to be updated such that the particles move toward the target.

The velocity of each particle is updated in every run on collectively considering the past velocity of the particle and the two best values defined below:

- Optimal position confronted by each particle, Ownbest
- Best position confronted by the swarm that is the optimal solution achieved with any particle in the swarm, Swarmbest.

Algorithm 1: PSO best finding

begin

Define problem space and ***FoM***

Generate a swarm of ***N*** *particles with initial velocity and position*

while *(maximum* ***runs*** *is not met)* ***do for*** *each particle*

Evaluate ***FoM***

Record ***Ownbest****and* ***Swarmbest***

Update the velocity and position of the particle

end for end while

end

Attribute selection is a combinatorial problem that involves the inclusion or exclusion of attributes to identify an optimal subset of highly predictive attributes for superior classification performance. Thus, the problem of attribute selection demands a large Hamming space with all combinations of attributes as solution vectors on it. A PSO is more appropriate to explore and exploit the discrete problem space than a PSO that best combs a continuous problem space. The Pseudocode of cognitive-based attribute selection using PSO is given in algorithm 2. The position vector of N particles in the swarm with D dimensions each represents the status (presence or absence) of attributes in the subset. The initial position vector of N particles is selected randomly from the problem space i.e., each particle is assigned any one of the possible combinations of the attribute subset. Each particle is represented with index $i = 1, 2, 3..., N$

The position of the particle in swarm is denoted as $(X_1^{run}, X_2^{run}, ..., X_N^{run})$.

$X_i^{run} = (x_{i1}^{run}, x_{i2}^{run}, ..., X_{ij}^{run}, ..., X_{iD}^{run})$ represents the position vector of a particle, i with dimension j, j=2, …, D in the current run.

$$X_1^{run} = 1 \{ presence\ of\ attribute\ j\ in\ the\ subset\ 0\ absence\ of\ attribute\ j\ in\ the\ subset \} \quad (4.1)$$

Algorithm 2: Wrapper-based Attribute Selection using PSO

begin

Define a D-dimensional Hamming space, R^D

Represent all possible combinations of D attributes as elements of R^D

Randomly place N particles in the problem space, R^D

Initialize the velocity of each particle using

$$v_{ij}^{run} = v_{min} + \left(v_{max} - v_{min}\right) * rand(0, 1)$$

while *(maximum runs is not met) do*

for *each particle*

Evaluate FoM using $FoM,\ f = \alpha(P) + (1 - \alpha)(1 - \frac{N_{selected}}{N_{total}})$

Calculate Ownbest and Swarmbest

for *each dimension of a particle*

Update velocity using

$$v_{ij}^{run+1} = \omega v_{ij}^{run} + C_1 r_{1j}^{run}\left[ownbest_{ij}^{run} - x_{ij}^{run}\right] + C_2 r_{2j}^{run}\left[ownbest_{ij}^{run} - x_{ij}^{run}\right]$$

Update position using $x_{ij}^{run+1} = \left(1\ u_{ij}^{run} < S_{ij}^{run}\ 0\ u_{ij}^{run} \geq S_{ij}^{run}\right)$

end for

end for

end while

Return Swarmbest as the optimal attribute subset

end

The velocity of the particle decides the change in direction of the particle's flight toward the potential solution. It represents the step size that each particle has to take to comb the problem space in search of an optimal solution. Particle velocity is calculated with equation 4.2.

$$v_{ij}^{run} = v_{min} + \left(v_{max} - v_{min}\right) * rand(0, 1) \tag{4.2}$$

$v_1^{run}, v_2^{run}, ... v_N^{run}$ denotes velocity of each particle in the swarm.

$v_i^{run} = v_{i1}^{run}, v_{i2}^{run}, ...v_{ij}^{run}, ..., v_{iD}^{run}$ marks velocity of the i^th^ particle at j^{th} dimension in the current run. v_{max} and v_{min} represent minimum and maximum allowable velocity and N is the particles count in the swarm with D dimensions each. rand operator generates a random number ∈ (0,1).

On initializing position and velocity, particles span through the search space to adjust their velocity and position under their knowledge (Ownbest) and that of the swarm (Swarmbest) through social interaction. Each particle in the swarm demands neighbors for social interaction. The neighborhood of each particle for social interaction is the entire swarm and it is represented by a star topology with a fully connected network. Each particle interacts with every other particle in the swarm, as a result, star topology aids in faster convergence.

In the process of search, the step size of each particle required for traversing the problem space toward the target is decided in every run. It is an indication of how far the individual data is from the target. If the particle is farther from the target, the velocity of the particle flight is increased to reach the target. It is derived by considering the memory of flight direction in the immediate past, a memory of the position that was best so far for the particle, and that of the entire swarm. Velocity updating in each dimension of the particle is calculated using equation (4.3)

$$ownbest_{ij}^{run} = \left\{1\ presence\ of\ attribute\ j\ in\ the\ ownbest_i^{run}\ 0\ absence\ of\ attribute\ j\ in\ the\ ownbest_j^{run}\right\} \tag{4.3}$$

Swarmbest is the best solution found so far during searches by all swarms of particles. It is represented as $swarmbest = (swarmbest_1, swarmbest_2, ..., swarmbest_j, ..., swarmbest_D)$

$$swarmbest_{ij}^{run} = \left\{ 1\ presence\ of\ attribute\ j\ in\ the\ swarmbest_{i}^{run}\ 0\ absence\ of\ attribute\ j\ in\ the\ swarmbest_{j}^{run} \right\} \tag{4.4}$$

Velocity update uses acceleration coefficients, C_1 and C_2 to measure the confidence that a particle has in itself and its neighbors in the swarm respectively. To balance both exploration and exploitation, C_1 is kept equal to C_2 such that the particles move towards the average of Swarmbest and its Ownbest**.** Both the coefficients along with r_1 and r_2 maintain stochasticity during the velocity updating of particles. Velocity update is inferred as the change in the probability of finding the particle in one state or another. Hence velocity is mapped to an interval (0,1) by using a monotonically increasing sigmoid function as in equation 4.5.

$$S_{ij}^{run} = \frac{1}{1+e^{-v_{ij}^{run+1}}} \tag{4.5}$$

Where S_{ij}^{run} denoted normalized velocity of the i^th^ particle at j^{th} dimension in the current run on the velocity, v_{ij}^{run+1}.

The swarm position in each dimension is corrected with equation (4.6)

$$x_{ij}^{run+1} = \left(1\ u_{ij}^{run} < S_{ij}^{run}\ 0\ u_{ij}^{run} \geq S_{ij}^{run} \right) \tag{4.6}$$

Where u_{ij}^{run} is a random number interval (0,1) x_{ij}^{run+1} indicates the position where inclusion or exclusion of attribute of the i^th^ particle at $the\ j^{th}$ dimension in the next run.

Finally, Swarmbest and Ownbest solution vectors are restructured by comparing with FoM of current and past potential solutions obtained during the search such that the particles are directed towards an optimum point in the problem space. This process is iteratively repeated till it reaches a pre-fixed maximum number of runs. On termination, Swarmbest is the optimal solution obtained on executing PSO as a search strategy [133].

4.4.1 PSO with greedy reset and localized random mistakes

To handle the problem of pre-mature convergence in PSO due to Swarmbest and velocity stagnation, PSO with Greedy reset and localized random mistakes is proposed. It will be more meaningful to reset the Swarmbest based on a heuristic approach rather than resetting the Swarmbest with zero attributes or resetting all the attributes except at one random position. An attribute selection technique aims to find a subset with fewer attributes of high predictive capability. Of the two objectives such as attribute subset size and predictive accuracy of attribute selection, it is difficult to anticipate the predictive capability of a set of attributes without prior knowledge about the problem. It is problem specific depending on how well attributes are extracted from the problem space. Hence, it is difficult to derive a heuristic approach based on the predictive performance of attribute space to reset the Swarmbest. On the other hand, attribute size, D, of a problem is known apriori and the dimension of the selected attribute subset is D' with D'< D. On Swarmbest stagnation, this heuristic is adopted to design the proposed methodology to greedily look for the subset with reduced attributes. On pre-mature convergence i.e., if Swarmbest position doesn't change for a prefixed number of runs, Swarmbest is reset. It is reset such that the dimension of Swarmbest, k, is reduced to k (k'<k). The Cardinality of Swarmbest after reset on pre-mature convergence is one less than the Swarmbest before reset as in equation 4.7. Pseudo-code of greedy Swarmbest reset and Ownbest diversion is given in algorithm 3.

$$k' = k - 1 \tag{4.7}$$

Algorithm 3: Greedy Reset and Ownbest Diversion Algorithm

Algorithm Parameters

$\boldsymbol{swarm_{stag}}$, Maximum allowable interval for ***Swarmbest*_**stagnation

N, Swarm size= Number of particles

D, Dimension of the particle=Number of attributes

begin

    if (***Swarmbest***constant for $\boldsymbol{swarm_{stag}}$times)

        k= length (***Swarmbest***)

        Reset ***Swarmbest***

        Randomly choose (***k***-1) positions in the range [1 ***D***]

        Set the chosen positions to one in the ***Swarmbest***

    */*Ownbest diversion*/*

        for *i*= 1 to ***N***

            checki=generate range [1 ***D***]

            Randomly choose ***to*** $\boldsymbol{check_i}$ positions in ***Ownbesti***

            Flip only those positions chosen in the ***Ownbesti***

        end for

    end if

end

The proposed methodology resets Swarmbest except at (k-1) positions where k is the attribute size of Swarmbest. Positions in Swarmbest for greedy reset are randomly chosen, thus holding up the stochasticity of the PSO algorithm. On reset, particles in the swarm that are bribed towards Swarmbest are diverted to untried areas by flipping randomly chosen positions in Ownbest.

4.4.2 PSO with velocity clamping and localized random mistakes

After velocity updating in PSO, if the updated velocity is greater than the maximum velocity, then the updated velocity has to be clamped to the maximum velocity. After velocity updating, if v_{ij}^{run+1} is greater (lesser) thanv_{max} ($-v_{max}$), it is clamped to v_{max} ($-v_{max}$). The process of velocity clamping is explained in equation 4.8.

$$v_{ij}^{run+1} = \left(min(v_{max}, v_{ij}^{run+1}\right), \ - v_{max}) \tag{4.8}$$

where $v_{ij}^{run+1}$is the updated velocity of i[th] particle position in the j[th] vector dimensional. $\pm v_{max}$ is the allowable step size to comb the problem space.

On velocity stagnation, the position vector of particles remains unchanged and no further exploration takes place in further runs. Hence, localized random mistakes are introduced such that particles are diverted to unexplored areas in the problem space. The originality of solution vectors obtained on social interaction by particles in the swarm is lost if mutated in every run. Thus, localized random mistakes are introduced in the position vector of particles only on velocity stagnation. Pseudocode for velocity clamping and localized random mistakes is given in algorithm 4. The number of velocity clamping per particle is calculated as clamp. From the solution vector, clamp positions are chosen randomly and flipped, thereby introducing localized random mistakes only during velocity stagnation [134].

Algorithm 4: Calculates the Number of velocity clamping per particle and localized random mistakes

Algorithm Parameters

v_{max} *maximum allowable step size to comb the problem space*

begin

    for *each particle clamp=0*

        for *each dimension of the particle*

Update velocity, vijrun+1 using

$$v_{ij}^{run+1} = \omega v_{ij}^{run} + C_1 r_{1j}^{run}\left[ownbest_{ij}^{run} - x_{ij}^{run}\right] + C_2 r_{2j}^{run}\left[ownbest_{ij}^{run} - x_{ij}^{run}\right]$$

if $v_{ij}^{run+1} = v_{max}$

clamp =clamp+1

Velocity clamping as in $v_{ij}^{run+1} = \left(min(v_{max}, v_{ij}^{run+1}\right), \; - v_{max})$

end if

Update position, x_{ij}^{run+1} *using* $x_{ij}^{run+1} = \left(1\ u_{ij}^{run} < S_{ij}^{run}\ 0\ u_{ij}^{run} \geq S_{ij}^{run}\right)$

end for

Choose the 'clamp' number of positions in x_{ij}^{run+1} *randomly Flip only those positions chosen in* x_{ij}^{run+1}

end for

end

Velocity clamping and localized random mistakes

Algorithm parameters for PSO with greedy reset and localized random mistakes are given and the overall pseudo-code is explained in algorithm 5.

N, Swarm size=Number of particles

D, Dimension =Number of attributes

C1 and C2, Acceleration constants

v_{max}, Maximum allowable step size to comb the problem space

$swarm_{best}$, Maximum allowable interval for $swarm_{best}$ stagnation

run_{max}, the Maximum number of runs

Algorithm 5 Parameters for PSO with Greedy reset and localized random mistakes

begin

Define a combinatorial problem space, $\mathbf{R^D}$

Let all possible combinations of attributes represent elements in $\mathbf{R^D}$

Initialize algorithm parameters as in algorithm 5

Random initialization n of particles in the swarm, x_{ij}^{run}

Initialize$v_{ij}^{run+1}$of each particle using (2) and evaluate FoM, fi

Assign the fitness of each particle and its position as $\mathbf{f}_{\mathbf{ownbest, I}}$ and $\mathbf{Ownbest}_{\mathbf{i}}$

Assign the best fit among particles and its position as $\mathbf{f}_{\mathbf{Swarmbest}}$ and **Swarmbest**

while ($\mathbf{run}_{\mathbf{max}}$<$\mathbf{run}_{\mathbf{i}}$) **do** for i=1 to n number of run

for j=1 to **D**

Randomly choose r_{i1}^{run}, r_{i2}^{run}, u_{ij}^{run} and update velocity v_{ij}^{run+1}

if$|v_{ij}^{run+1}| \geq v_{max}$

Velocity clamping using equation (4.8)

endif

Update position x_{ij}^{run+1}using equation (4.6)

end for

Localized random mistakes to x_{ij}^{run+1}

Evaluate **FoM**, $\mathbf{f}_{\mathbf{i}}^{\mathbf{run+1}}$

if (firun+1 ≥ fownbest,i)

fownbest,i= firun+1 ; Ownbesti= xij

end if

if (firun+1 ≥ fSwarmbest)

fSwarmbest= firun+1;Swarmbest= xijrun+1

end if **end for**

if Swarmbestconstant for $\mathbf{swarm}_{\mathbf{stag}}$ **runs**

Greedy Swarmbest reset and **Ownbest Diversion end if**

end while

end

The proposed re-initialization mechanism such as greedy reset is significant in handling the problem of pre-mature convergence in PSO on making a meaningful reset with dimensionality reduction as the heuristic, unlike a reset with zero attributes, a reset except at one random position, perturbation-a jump to a random point with random velocity. SIOC-CAS with Greedy reset incorporates domain-specific knowledge such as reduced attribute size to reset the locked swarm's best solution. Introducing greedy reset with localized random mistakes on pre-mature convergence, the stochastic behavior of swarm optimization is not lost. The main advantage of SIOC-CAS methodology with Greedy reset and localized random mistakes is that it does not require additional parameters such as $\boldsymbol{r_{mut}}$ to introduce mutation and $\boldsymbol{d_r}$ to displace Ownbest.

4.5 Experimentation and Results

PSO with Greedy reset and localized random mistakes given in algorithm 5 is developed with a simulation tool with the platform of Intel i3 core processor with a high-resolution graphical card. To investigate the success of the proposed algorithm for attribute subset selection, datasets from the UCI repository as tabulated in table 4.1.

Table 4.1 Dataset description with many attributes and instances

$D_{\#}$	Data	Attributes	Instances
D_1	PID	8	768
D_2	WBCO	9	699
D_3	Australian	14	690
D_4	Hepatitis	19	155
D_5	Credit	24	1000
D_6	WBCD	30	569
D_7	WBC	32	198
D_8	Ionosphere	34	351
D_9	Spectf	44	267
D10	Sonar	60	208
D11	HillValley	100	606
D12	Musk	166	476

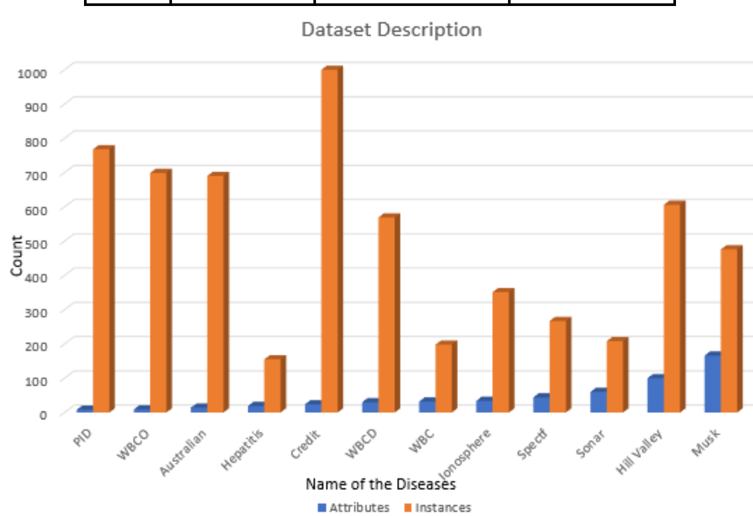


Figure 4.1 Dataset description with many attributes and instances

4.6 Heterogeneous Multimodalities Data Fusion

Attributes of different dynamic ranges in a dataset are normalized so that attributes with larger values do not influence the attributes with smaller values, thus, making the attributes lie in a specified range. It also reduces difficulties during numerical calculation. Heterogeneous multimodalities data fusion (HMDF) of data given in equation 4.9 is used in this work to linearly scale in the range (0,1).

$$z' = \frac{z - min_n}{max_n - min_n},\ |z| = n \tag{4.9}$$

z' is the normalized value of attribute z in the classification problem such as dataset. Each attribute in a given classification problem has 'n' instances. min_n and max_n represent the minimum and maximum values of an attribute with 'n' instances each in the dataset.

4.6.1 Data encoding

The presence and absence of attributes in the subset are represented by 1's and 0's to address the combinatorial optimization problem of attribute selection. Binary encoding of attribute space, $x = \{x1,\ x2,\ x3,\ x4,\ x5,\ x6\} \in R^5$ with cardinality, $|\boldsymbol{x}|$=5, refer figure 4.2.

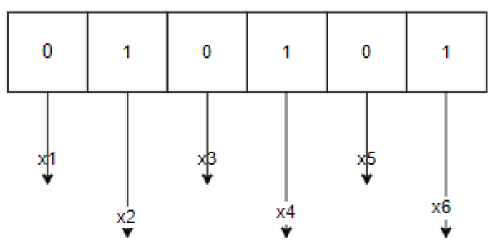


Figure 4.2 Binary encoding subset with five Attributes

As observed from figure 4.2, attributes x_1, x_3, and x_5 are unselected and x_2, x_4, and x_6 are selected for further processing. Attribute subset $x' = \{x_2, x_4, x_6\}$ with cardinality, $|x'| = 3$.

4.6.2 Evaluation of PSO with greedy reset as the search technique in SIOC-CAS-based attribute selection

SIOC-CAS algorithms for an effective attribute selection process are evaluated with defined datasets refer table 4.1. Maximum *runs* count, run_{max} is set to 100. Both acceleration constants $\boldsymbol{C_1}$ and $\boldsymbol{C_2}$ are fixed to 2, and allowable *Swarmbest* stagnation before greedy reset, $swarm_{stag}$ is set to 3. Inertia weight, m is set to 1. Swarm size, $\boldsymbol{N}$ is set to 60 empirical studies from most PSO implementations that use swarm size ∈ [20 60] and follows random initialization. [$-v_{max}$, v_{max}] is set to [-6, 6] such that sigmoid transformation in equation 4.6 yields a probability range [0.9975, 0.0025]. Thus, velocity $\boldsymbol{v_{ij}^{run}}$ is the probability of the particle's position, x_{ij}^{run} remaining at 0 or 1. SIOC-CAS with a radial basis kernel is used for subset evaluation with 10-fold cross-validation. Predictive accuracy and the corresponding subset size of the musk dataset are shown in figure 4.3(a) and 4.3(b) respectively for a single iteration with 100 *runs*.

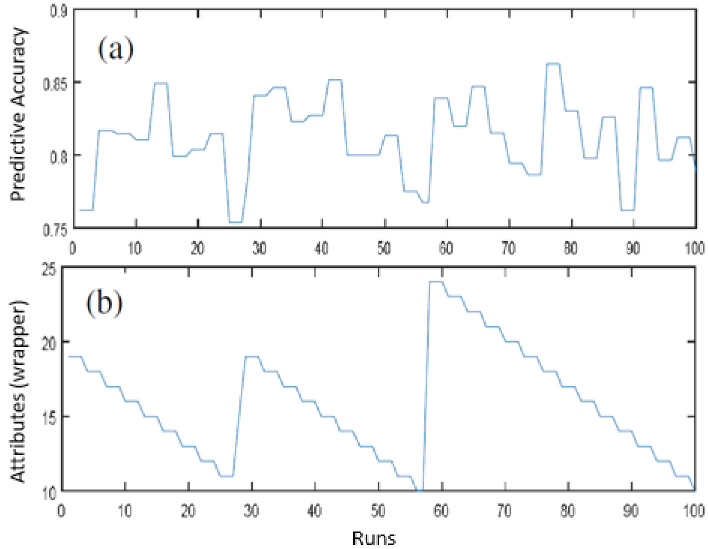


Figure 4.3 Predictive accuracy (a) and size of attribute subset (b) that are realized in 100 runs for musk dataset on applying PSO with greedy reset and localized random mistakes

The following observations are made from figure 4.3.

As evident from figure 4.3 (a), on pre-mature convergence, Swarmbest is reset to another point (new Swarmbest) in the problem space. From figure 4.3 (b), it is observed that the new Swarmbest selected has one attribute less than the past Swarmbest, resulting in Greedy reset and search. Figure 4.3 (b), in the course of the search, greedily looks for subsets with lesser attributes than the previous Swarmbest in the problem space such as from run 1-27 with the proposed FPO algorithm. At the same time, Greedy reset and search do not compromise on predictive accuracy as evident at run 28 because FoM calculated using the figure of merit values is modeled with more importance laid to accuracy than attribute subset size.

4.6.3 Accuracy gain and dimensionality reduction by SIOC-CAS approach using PSO with greedy reset

The predictive accuracy of the SIOC-CAS classifier without dimensionality reduction such as all the attributes as input is calculated to compare the predictive accuracy of the classifier after dimensionality reduction. Table 4.2 portrays the results of average ± standard deviation. The predictive accuracy of selected attributes, from the process of attribute selection, is compared with the predictive accuracy of the SIOC-CAS classifier with all the attributes as input to calculate accuracy gain (AG) in employing wrapper-based attribute selection technique. It is calculated as the improvement in predictive accuracy with selected attributes than with all the attributes in the dataset.

$$Accuracy\ gain\ (\%) = \frac{p_{sel} - p}{p_{sel}} X100 \tag{4.10}$$

where p_{sel} and P represent the predictive accuracy of the SIOC-CAS classifier with selected attributes and all the attributes as input. Similarly, dimensionality reduction (DR) is calculated as the reduction in the dimensionality of attribute space by employing a cognitive-based attribute selection technique.

$$DR = \frac{\#A - \#A_{sel}}{\#A} X100 \tag{4.11}$$

where $\#A_{sel}$ and #A represents the total selected attributes and total attributes in an original dataset. Of the twelve datasets, six datasets such as Australian, WBCD, Spectf, Sonar, Hill Valley, and Musk show an accuracy gain of more than 5% and only three datasets such as Hepatitis, WBC, and Spectf show a dimensionality reduction less than 75%. There is a commendable reduction in the dimensionality of attribute space for decision-making in the final classification phase.

Table 4.2 Accuracy Gain (AG) and dimensionality reduction (DR) by HMDF approach using PSO greedy reset

Data	Without Attribute Selection (Average ± Standard Deviation)		Proposed (Average ± Standard Deviation)		AG(%)	DR(%)
	P(%)	#A	P(%)	#A		
PID	72.25±0.72	8	74.03±0.3	1.5±0.51	2.4	81.25
WBCO	95.39±0.64	9	95.95±0.32	2.2±0.57	0.58	75.55
Australian	82.55±0.94	14	87.53±0.71	3.2±0.3	5.68	77.14
Hepatitis	74.64±0.84	19	74.75±0.72	7.8±0.6	0.14	58.94
Credit	69.88±0.81	24	70.50±0.5	5.5±0.2	0.87	77.08
WBCD	90.06±0.32	30	94.81±0.81	2.8±0.7	5.01	90.67
WBC	76.69±0.71	32	76.85±0.91	8.5±0.5	0.20	73.43
Ionosphere	87.57±0.98	34	92.14±0.46	7.2±0.92	4.95	78.82
Spectf	78.95±0.94	44	83.27±0.72	11.5±1.39	5.18	73.86
Sonar	56.60±0.74	60	86.77±0.92	10.3±1.2	34.77	82.83

HillValley	46.76±0.62	100	52.06±0.72	1.6±0.83	10.18	98.40
Musk	56.99±0.32	166	86.92±0.37	18.1±3.54	34.43	89.06

Datasets with more than 50 attributes such as spectf, hill valley, and musk achieve an accuracy gain of more than 10% and a dimensionality reduction of more than 82%. Thus, the proposed attribute selection methodology favors large dimensional datasets with superior performance in terms of both accuracy gain and dimensionality reduction. Although datasets such as WBCO, Hepatitis, Credit, and WBC show very improvement in accuracy, there is a remarkable reduction in the dimensionality of attribute space.

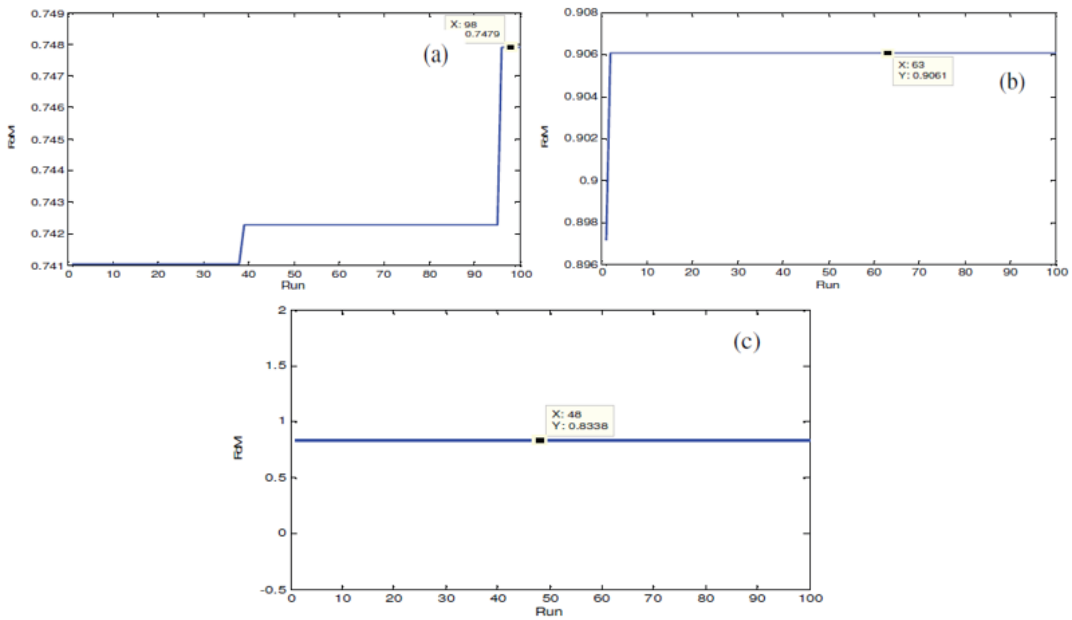


Figure 4.4 FOM of (a) PID (b) WBCD (c) Musk that is realized in 100 runs on applying BFPA as Search Strategy in wrapper-based attribute selection

FoM of PID, WDBC, and Musk datasets for a single iteration of 100 runs are shown in figure 4.4 (a), figure 4.4(b), and 4.4(c). It is observed that pollenbest remains unaltered in all the runs in the case of WDBC and Musk. Pollenbest is altered only thrice in the case of the PID dataset. Similar results are observed in all other datasets. As evident from figure 4.3, Pollenbest is trapped in local optima in the problem space, after which there is no exploration leading to stagnation of pollens in the population.

4.7 Performance of Various Classifiers with Selected Attributes by PSO with Greedy Reset

The performance of the SIOC-CAS algorithm compared with literature on PSO for attribute subset selection. Results of existing and proposed work using PSO are tabulated in table 4.3.

The predictive capability of six traditional classifiers with selected attributes as input is measured to validate the efficiency of the proposed attribute selection technique as a pre-processing tool for classification. The attribute subset selected by the proposed algorithm is input to six traditional classifiers such as SIOC-CAS (RBF, σ = 1) with c = 1, KNN with k=1, NB, SVM, DT and ensemble learning with adaboost algorithm. Tenfold cross-validation is employed and the results are averaged over 100 iterations. Table 4.4 shows the final consolidated values.

Table 4.3 D1 to D6 predictive performance of six classifiers with attributes selected by PSO with greedy reset

Data#	SIOC-CAS (%)	KNN (%)	SVM (%)	NB (%)	DT(%)	Ensemble(%)
D_1	74.2	68.2	73.0	74.1	69.1	73.8
D_2	96.2	95.1	95.4	95.8	95.4	96.1
D_3	87.2	82.1	86.1	86.5	83.2	86.9
D_4	75.0	72.3	74.8	71.0	75.2	75.4
D_5	71.1	69.1	70.8	70.2	69.4	70.1
D_6	94.7	92.9	94.8	93.2	94.4	94.2

Table 4.4 D7 to D12 predictive performance of six classifiers with attributes selected by PSO with greedy reset

Data #	SIOC-CAS (%)	KNN (%)	SVM (%)	NB (%)	DT(%)	Ensemble(%)
D_7	77.2	75.0	75.2	75.4	71.2	75.8
D_8	93.2	92.2	89.2	89.1	92.4	92.5
D_9	83.1	80.1	79.2	80.9	81.2	83.3
D10	86.7	84.2	83.8	84.3	84.8	85.1
D11	52.1	51.5	51.0	49.1	51.6	52.1
D12	86.1	82.7	83.2	81.8	84.3	85.0

All six classifiers achieve good predictive accuracy with selected attributes as input. SIOC-CAS and Ensemble learning classifiers show better results than other classifiers owing to their better generalization and predictive ability. Selected attributes, when fed as input to traditional classifiers such as KNN, NB, DT, DA, and Ensemble yield results on par with that achieved by the attribute selection technique employing the SIOC-CAS classifier. This proves the generality of the proposed algorithm as a pre-processing tool before any learning algorithm.

4.8 Random Population Initialization

A One to all random initialization is devised to generate an initial population for the search algorithm such that each candidate of the population represents a random solution in each dimension of the problem space. The proposed methodology draws motivation from forward and backward selection techniques for attribute selection. Forward selection begins with an empty set of zero attributes in it, while backward selection begins with all attributes included in the subset. The basic difference between the two techniques lies in initialization, that is the initial subset. The initial subset of the forward selection technique has no attributes and the initial subset of the backward selection technique has all the attributes in it. One to all random initialization is designed such that the initial population is comprised of solution vectors uniformly distributed in all the dimensions of the problem space. In this regard, the initial population constitutes solution vectors with attribute sizes ranging from one, two all attributes. In the proposed strategy cardinality of initial candidate solution vectors in the population ranges from 1,2, …, D where D is the problem dimensional space that is attributed to the problem size of the problem. Individual attributes are selected randomly thus the solution vectors of the initial population are generated such that it covers the entire problem space.

4.8.1 Population size

One to all random initialization not just describes how initial solution vectors are chosen but also decides the population size. Population size is an important parameter that largely influences the performance of any population-based search algorithm. A small population size may not be sufficient to explore the entire problem space. On the other hand, a large population increases the computational budget. Hence identifying an optimal population size is in itself a challenging optimization problem task. Hence, in One all random initialization, population size, N is fixed to D, that is the attribute size of the dataset D.

4.8.2 One to all random population initialization

Pseudocode of One to all random initialization. To generate the initial population of an evolutionary algorithm, population size needs to be fixed. In the case of datasets with a large number of attributes, computational complexity increases when the population size is fixed to the attribute size. Hence datasets with attributes less than N_{max} is the maximum population size fixed based on empirical results from implementations using respective evolutionary algorithms and One-to-all random initialization is employed. Pseudocode of One to all random initialization is given in the algorithm. For datasets with attributes greater than N_{max} random initialization is applied. Initial populations with solution vectors, X_i, i=1, 2, and N are generated stochastically by choosing i attributes in each X_i. Also here checking condition N_{max} is greater than or equal to dimensionality D with N iterations.

4.9 Significance and Merits of Proposed Methodology

The proposed initialization algorithm is novel in maintaining population diversity among initial solution vectors. The diversity of the initial population decides how well the initial solution vectors are distributed uniformly across the search space. The proposed methodology is evenly distributed in all the dimensions with sufficient randomness, thus upholding the diversity in the population of the initial solution vector. One to all random initialization technique for generating the initial population is simple, unlike other advanced initialization techniques as discussed in the literature. The proposed methodology generates uniformly distributed initial solution vectors. Population size is adaptive to the problem and does not require tuning unlike other random initialization techniques especially when there is no knowledge about the nature of attributes. Unlike other initialization techniques, it is not user-defined or fixed on tuning but set to an attribute size that is specific to the problem. The proposed methodology for population initialization requires less time to compute. One-to-all initialization strategy is an evolutionary search algorithm in the process of attribute selection. Moreover, this initialization strategy can be applied to any combinatorial problem space such as a knapsack problem. But initialization techniques do not generate uniformly distributed initial solution vectors covering the entire problem space. The proposed methodology maintains uniformity in initializing the individuals of the population by randomly selecting solution vectors in every dimension of the problem space, thus upholding population diversity.

4.10 Simulation and Results

Performance of the proposed attribute selection algorithm Greedy reset, Greedy crossover, and Greedy leap along with one to all random initialization is investigated to analyze the influence of One to all initialization strategy. Performance in terms of accuracy gain, dimensionality reduction, and computational times are evaluated. On employing One to all random initialization, population size is fixed to attribute size. On the other hand, random initialization uses a fixed population size. In the case of PSO and FPO, population size is fixed at

60 and 24 respectively such as empirical studies from PSO and FPO implementations. In the case of PSO with Greedy reset and Greedy leap, an attribute size greater than 60 such as hill and musk is employed with random population initialization. All other datasets from D1 to D10 compare traditional random initialization and One-to-all random initialization. In the case of FPO, a Greedy crossover attribute size greater than 24 in D6 to D12 is employed with random population initialization. Datasets from D1 to D5 compare traditional random initialization and random initialization technique. The performance of proposed attribute selection algorithms with One to all random population initialization is analyzed in terms of accuracy gain and dimensionality reduction. Its performance is compared with the random population initialization technique based on results consolidated in table 4.2 such as Greedy reset, table 4.3 such as Greedy crossover, and table 4.4 such as Greedy leap. Table 4.2 cognitive-based attribute selection algorithm using PSO with Greedy reset and One-to-all random population initialization achieves superior results in terms of predictive accuracy compared to attribute selection using PSO with Greedy reset and traditional random population initialization. Dimensionality reduction in attribute space of all the datasets is better with One-to-all random initialization than the traditional population initialization strategy except WBCO and WBC, nevertheless, the accuracy gain of these two datasets is commendable compared to the traditional initialization technique.

Table 4.5 Random population initialization Vs one to all random initialization PSO with greedy reset in terms of predictive accuracy (P), attribute subset size (#A), accuracy gain (AG), and dimensionality reduction (DR)

D	PSO with Greedy Reset and Random Initialization				PSO with Greedy Reset and One-to-All Random Initialization			
	P(%)	#A	AG	DR	P(%)	#A	AG	DR
D_1	74.03±0.3	1.5±0.51	2.4	81.25	74.13±0.4	1±0	2.5	87.5
D_2	95.95±0.32	2.2±0.57	0.58	75.55	96.46±0.43	2.3±0.47	1.10	74.44
D_3	87.53±0.71	3.2±0.3	5.68	77.14	87.63±0.53	2.6±0.61	5.79	81.42
D_4	74.75±0.72	7.8±0.6	0.14	58.94	75.96±0.81	3.6±0.5	1.73	81.05
D_5	70.50±0.5	5.5±0.2	0.87	77.08	74.48±0.57	4.5±0.57	6.17	81.25
D_6	94.81±0.81	2.8±0.7	5.01	90.67	95.37±0.82	2.2±0.5	5.56	92.5
D_7	76.85±0.91	8.5±0.5	0.20	73.43	78.81±0.41	10.2±0.42	2.69	58.12
D_8	92.14±0.46	7.2±0.92	4.95	78.82	95.62±0.92	7.1±0.82	8.41	79.11
D_9	83.27±0.72	11.5±1.39	5.18	73.86	84.38±0.01	11±1.58	6.43	75.0
D10	86.77±0.92	10.3±1.2	34.77	82.83	87.66±0.8	9.8±1.26	35.43	83.53

Table 4.6 Random population initialization Vs One-to-all random initialization in case of FPO with greedy crossover in terms of predictive accuracy (P), size of attribute subset (#A), accuracy gain (AG), and dimensionality reduction (DR)

$D_{\#}$	Random				Overall			
	P(%)	#A	AG	DR	P(%)	#A	AG	DR
D_1	77.13±0.29	2±0	6.32	75	76.42±0.42	1.2±0.44	5.45	85
D_2	96.05±0.52	2.1±0.42	0.68	76.67	96.84±0.25	2.2±0.42	1.49	75.5
D_3	88.26±0.71	2.5±0.52	6.46	82.4	88.35±0.38	2.3±0.48	6.56	83.57
D_4	76.19±0.3	4.2±0.72	2.03	77.89	81.02±0.43	7.2±0.44	7.87	62.10
D_5	72.27±0.27	7.4±0.2	3.30	69.16	74.72±0.72	7±0.81	6.47	70.83

From table 4.6 initialization performs better either in terms of accuracy gain or in terms of dimensionality reduction when compared to FPO with Greedy crossover and random initialization. The proposed methodology, in the case of the PID (D1) dataset, performs better in terms of dimensionality reduction while WBCO (D2) and Hepatitis (D4) datasets perform better in terms of accuracy gain. On the other hand, the Australian (D3) and Credit (D5) dataset shows superior performance in terms of both objectives on applying One to all random initialization.

Table 4.7 Random population initialization Vs one to all random initialization of PSO with greedy leap in terms of predictive accuracy (P%), attribute subset size (#A), accuracy gain (AG) & dimensionality reduction (DR)

$D_{\#}$	PSO with Greedy Leap and Random Initialization				PSO with Greedy Leap and One-to-All Random Initialization			
	P(%)	#A	AG	DR	P(%)	#A	AG	DR
D_1	76.45±0.51	1.5±0.5	5.49	81.25	78.07±0.27	2±0	7.45	75
D_2	96.53±0.2	2±0	1.18	77.77	97.10±0.34	2±0	1.76	77.77
D_3	88.24±0.36	2±0	6.44	85.71	89.09±0.3	2±0	7.34	85.71
D_4	85.80±0.71	8.2±0.59	13.0	56.57	86.44±0.71	7.1±0.3	13.65	62.31
D_5	75.06±0.7	6.2±0.5	6.9	74.16	76.23±0.85	3.1±0.52	8.33	87.08
D_6	98.28±0.41	3.2±0.2	8.36	89.33	98.31±0.24	3±0	8.39	90
D_7	83.99±0.21	10.1±0.23	8.69	68.43	83.86±0.72	9.6±0.54	8.54	70
D_8	96.16±0.44	7±0.3	8.93	79.41	97.08±0.83	5.8±0.92	9.79	82.94
D_9	89.74±0.25	14.2±0.71	12.02	67.72	88.23±0.98	9.8±1.4	10.51	77.72
D10	94.78±0.28	12.3±0.52	39.22	79.46	94.42±0.92	8.8±2.4	40.05	85.33

From table 4.7, all random initialization achieves superior results in terms of both dimensionality reduction and accuracy gain for all the datasets except PID (D1), WBC (D7), Spectf (D9), and Sonar (D10). One to all random initialization, in the case of WBC (D7), Spectf (D9), and Sonar (D10) achieves superior performance in terms of dimensionality reduction but fails in terms of accuracy gain, nevertheless, the difference in accuracy gain is only marginal compared to the random initialization.

4.11 Performance Analysis in terms of Computational Time

Generally, factors that influence computational time are population size (N), number of attributes (D), and instances (n), specific to the dataset. The difference in computational time between traditional random initialization and One-to-all random initialization technique, concerning Greedy reset, for all twelve datasets, ranges from 1261s (Hepatitis, D4) to 8918s (Credit, D5). Similarly, the difference in computational time for the Greedy crossover ranges from 1077s (Hepatitis, D4) to 4436s (PID, D1) and the Greedy leap takes a range from 1477s (Spectf, D9) to 9460s (Credit, D5).

Table 4.8 Random population initialization Vs one to all random initialization (Computation Time (sec))

$D_{\#}$	PSO with Greedy Reset		FPO with Greedy Crossover		PSO with Greedy Leap	
	Random	One To All Random	Random	One To All Random	Random	One To All Random
D_1	9671.2	1822.2	6649.5	2213.2	9446.5	1755.5
D_2	3125.2	959.4	2464.9	1029.4	2986.2	802.5
D_3	6420.3	2310.2	5629.9	3119.8	6027.3	2747.3
D_4	2659.5	880.2	1999.3	922.3	2399.1	910.2
D_5	17210.2	8292.1	11519.6	9828.3	17005.2	8095.5
D_6	7830.4	3180.6	2717.6		7262.9.5	2490.6
D_7	3715.2	1861.87	1798.2		35115.2	1727.4
D_8	4890.5	2911.5	2741.2		4492.5	2658.2
D_9	5162.5	3901.8	3762.5		5128.2	3651.55
D10	5380.7	3969.2	3068.66		5001.2	3312.8
D11	9460.6		9029.5		9239.8	
D12	10767.1		9812.2		10682.5	

Comparing the three Cognitive-based algorithms Greedy leap with One-to-all random initialization requires less computational time than Greedy crossover and Greedy reset. On the other hand, Greedy reset and Greedy crossover is a two-step process such as the stagnated solution is restructured followed by particle or pollen diversion to untried areas in the problem space. Thus, on generating initial solution vectors using the One-to-all random population initialization technique, the computational time required for attribute selection is considerably saved.

4.11.1 Comparison of one to all random initialization with literature on population initialization techniques for attribute selection

One to all random population initialization strategy is compared with population initialization strategies using datasets from the UCI repository for validating the proposed algorithms, applied only six datasets representing the two-class problem and table 4.9 shows the all-random initialization results.

Table 4.9 One to all random initialization Vs Literature on population initialization techniques for attribute selection

Data	Literature		Proposed	
	Technique	P (%)/#A	Re-initialization technique	P (%)/#A
Credit	Mixed	69.27/13.68	Leap	76.23/3.1
	Large	69.13/16.82	Crossover	72.27/7.4
	Small	67.35/2.34	Reset	74.48/4.5
WBCD	Mixed	93.78/8.12	Leap	98.31/3
	Large	93.02/19.06	Crossover	96.34/2.6
	Small	93.05/3.04	Reset	95.37/2.2
Ionosphere	Mixed	87.45/3.26	Leap	97.08/5.8
	Large	86.5/18.38	Crossover	96.13/7.9
	Small	88.06/3.36	Reset	95.62/7.1
Sonar	Mixed	77.4/11.3	Leap	94.42/8.8
	Large	78.03/31.7	Crossover	88.03/9
	Small	76.83/6.186	Reset	87.66/9.8
HillValley	Mixed	57.68/13.16	Leap	54.18/2
	Large	57.78/60.56	Crossover	53.36/1.4
	Small	56.98/6.74	Reset	52.06/1.6
Musk	Mixed	84.62/83.54	Leap	90.01/20.2
	Large	85.51/107.14	Crossover	88.26/16.4
	Small	81.57/16.66	Reset	86.92/21.1

In the case of the hill valley dataset, all the initialization techniques briefed in the literature outperform the proposed One-to-all random population initialization technique but at the cost of a large attribute subset size. Hence proposed initialization technique performs better compared to the literature on population initialization techniques for attribute selection.

4.11.2 Proposed attribute selection techniques with one to all random initialization vs literature on evolutionary search techniques

The proposed methodology such as Greedy reset, Greedy crossover, and Greedy leap with One to all random initialization is compared with attribute selection techniques using various other evolutionary algorithms as search strategy and table 4.10 shows the overall literature and proposed initialization values. It is observed that in the case of all the datasets, the attribute selection algorithm using PSO with Greedy leap outperforms literature on attribute selection techniques using metaheuristic search algorithms based on the biological behavior of ant, bee, bat, cuckoo, differential evolution (DE), firefly, GA and wolf. However, the weighted bee colony algorithm proposed, shows superior performance in terms of predictive accuracy. The difference in predictive accuracy between the proposed and weighted bee colony algorithms is small and the number of attributes is unspecified in the latter. So, a solid comparison cannot be accomplished. There is no literature on spectf such as D9 using other metaheuristics algorithms as a search strategy.

4.11.3 Performance of various classifiers with selected attributes by proposed attribute selection techniques with one to all random initialization

One-to-all random population initialization, the predictive performance of six traditional classifiers with selected attributes as input, is measured. The results tabulated in table 9 provide datasets D1-D10 in case of Greedy reset and Greedy leap D1-D5 in case of Greedy crossover. From table 4.10, it is observed that in the case of all the datasets, all six classifiers with selected attributes from corresponding attribute selection algorithms show consistent performance in terms of predictive accuracy. This proves the generality of the proposed attribute selection algorithms. The cognitive approach for attribute selection employs an evolutionary search algorithm to select highly predictive attributes from the problem space. Evolutionary search algorithms are population-based and demand an initial population of solution vectors to initiate the search process.

Table 4.10 Predictive performance of six classifiers with attributes selected by the proposed attribute selection algorithms with one to all random initialization

Data Label	Methodology	SIOC-CAS (%)	KNN(%)	SVM (%)	NB(%)	DT(%)	Ensemble(%)
D_1	Reset	74.1	68.2	72.8	74.5	69.2	72.9
	Crossover	77.5	72.1	76.5	77.1	72.3	77.2
	Leap	78.2	74.1	77.5	75.3	74.2	77.2
D_2	Reset	95.9	94.0	94.0	95.8	95.1	96.1
	Crossover	96.5	95.2	94.9	96.2	96.3	95.9
	Leap	97.3	94.2	95.3	97.1	95.1	97.0
D_3	Reset	87.2	80.1	85.2	87.1	84.8	87.6
	Crossover	87.9	85.2	86.5	87.5	86.1	87.2
	Leap	88.9	84.2	86.9	87.2	87.5	87.5
D_4	Reset	75.8	74.3	76.2	74.8	76..5	75.5
	Crossover	81.2	78.2	77.1	75.9	77.1	80.5
	Leap	85.8	86.5	83.5	85.2	84.4	85.1
D_5	Reset	73.8	70.2	71.2	74.2	70.9	74.1
	Crossover	74.8	71.1	72.1	71.7	73.0	73.1
	Leap	76.2	74.2	75.6	72.3	73.2	75.8
D_6	Reset	95.8	94.7	93.2	94.2	95.0	95.4
	Leap	98.4	97.8	97.2	96.8	97.2	97.3
D_7	Reset	77.8	72.1	71.2	70.5	73.4	78.1
	Leap	82.9	80.1	79.2	89.2	79.8	82.1
D_8	Reset	95.2	93.8	94.1	95.8	93.2	95.8
	Leap	97.1	96.2	95.5	96.2	95.2	96.8
D_9	Reset	84.5	81.2	82.1	83.2	82.4	84.2
	Leap	87.2	84.2	85.9	86.9	87.4	88.4
D10	Reset	87.8	85.7	86.2	86.8	84.9	86.9
	Leap	94.1	92.1	93.2	93.9	92.5	93.4

The proposed methodology generates an initial population such that candidates represent solution vectors from each dimension of the problem space, thus, upholding population diversity. Also, the proposed population initialization strategy decides the population size. Proposed cognitive-based feature selection algorithms along with a One-to-all random population initialization technique simulated and analyzed AG, DR, and time complexity. Results are compared with the proposed attribute selection algorithm along with the traditional random initialization technique. In most instances, attribute selection algorithms with One to all random initialization techniques outperform attribute selection algorithms with traditional random initialization techniques. The literature on population initialization techniques. The proposed technique is superior in performance in terms of either predictive accuracy or attribute subset size or both. Finally, attributes selected by the proposed techniques are fed as input to six traditional classifiers and their predictive performance is analyzed to prove the generalization capability of the algorithm.

4.12 Disease Spot in Rice Leaf

Brown spot disease reduces the quantity and quality of grain in the plant. Severely infected rice fields can have as high as a 45% loss in yield & on average 5% yield loss in lowland rice production.

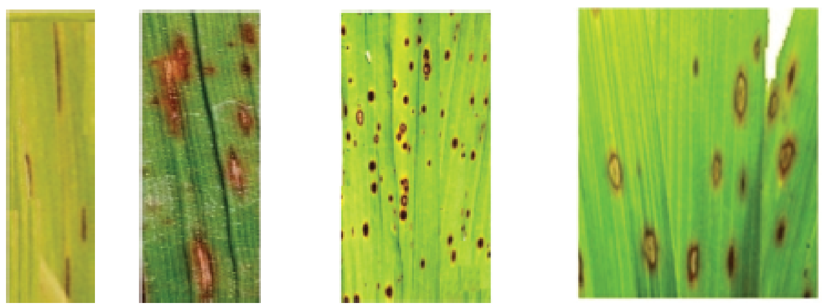


Figure 4.5 Rice leaf infected by Narrow Brown Spot (a), (b) and Brown Spot Disease (c), (d)

Table 4.11 Characteristics of disease during different stages of development

Disease	Stage	Descriptor	Characteristics
Brown spot	Vegetative	Color	Dark brown to reddish brown
		Shape	Small, circular too value (1/8 inch in diameter)
	Flowering	Color	Dark brown margin with reddish brown to grey center
		Shape	Circular to oval
Narrow brown spot	Vegetative	Color	Cinnamon brown
		Shape	Long (1/10 to 1/12 inch) and narrow (1/32 inch).
	Flowering	Color	Cinnamon brown
		Shape	Large (3/2 to 3 inch long) and narrow

Narrow brown spot disease has no major economic effect in terms of yield loss, but declines the market value of the grain and reduces the milling recovery. Disease spots expand across veins and leaves may be killed under unfavorable conditions. Foliar fungicides are not economical for controlling the brown spot disease on commercial rice grain varieties. Narrow brown spot disease is controlled effectively by spraying fungicides. Brown spot disease is subdued by improving the fertility of the soil. Since the corrective measure differs from the brown spot and narrow brown spot diseases, the identification and evaluation of these two diseases are significant. But from figure 4.5 and table 4.11, it is observed that a narrow brown spot resembles a brown spot except that the lesions are long, narrow, and more linear. Also, the color of the brown spot varies from a narrow brown spot in the flowering stage. Hence, to classify the disease type, color, and shape-based features are extracted from segmented disease spots. Damage caused by diseases in rice leaves becomes more severe as plant approach maturity. Hence accurate disease detection, at an early stage, is necessary for both protection against crop failure and economic justification of treatment. Since uninterrupted monitoring over the largely cultivated region by agricultural experts and farmers is challenging, automatic disease diagnosis incorporating image processing techniques and machine learning proves to be an alternate solution and is observed as a pattern recognition problem.

4.12.1 Disease spot recognition in rice leaf using machine vision

The flow diagram of the disease identification process. Machine vision extracts descriptors such as attributes from segmented disease spots to categorize the disease type using a classifier shown on figure 4.6. Disease spots in rice leaves are segmented using a k-means algorithm to extract characteristic and distinctive attributes based on shape and color such as labeled as rice leaf disease spot dataset, RLDS. A comprehensive set of attributes allows better characterization of segmented disease spots. As a preprocessing step to classification, attribute subset selection is employed to reduce the dimensionality of the RLDS dataset, which is input to the classifier for final decision-making. Thus, proposed cognitive-based feature selection methodologies along with One-to-all random initialization are applied to choose an attribute subset such as from an extensive set of attributes describing the disease spot for effectively classifying various rice diseases in leaf. Rice leaf diseases stated in the Section, much research using machine vision to date focuses mainly on detecting and classifying leaf blast and brown spot diseases.

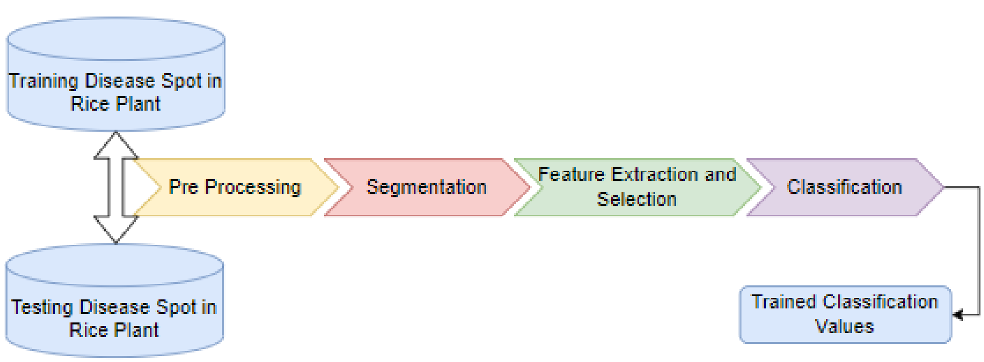


Figure 4.6 Process of disease identification in rice leaf using machine vision

4.12.2 Segmentation by k-means clustering

Color-based thresholding on color components of RGB color space, and indices describing mostly green color pixels are exploited by researchers to segment the diseased region with threshold obtained mostly from Otsu's algorithm. Two color-based traits of healthy leaf pixels in comparison with diseased pixels are observed.

Trait 1: In a healthy leaf pixel, the green color component value is larger than the red and blue such as RGB color space component value.

Trait 2: a* color constituent of L*a*b* color interplanetary is a balance between green and red color. Hence a healthy leaf pixel is greener than red. These color traits are the basis for discerning healthy and diseased leaf pixels using k-means clustering. It is a least-square partitioning methodology to cluster samples into k regions. In k-means clustering, the Euclidean distance between the center of the cluster and each pixel of the rice leaf image is calculated to label the pixel depending on its closeness to the cluster center. On assigning pixel labels, the center of clusters is re-calculated. The process is repeated till convergence.

4.13 RLDS Dataset

From the segmented disease spots, features describing shape and color are to be extracted for classifying the type of disease. Shape attributes provides from the boundary of disease spots following shape attributes are measured:

- Shape factor measure of how similar the lesion is to a circle.
- Extent measures how rectangular the lesion is. It is calculated as the area divided by the minimum bounding box area of the lesion.
- Eccentricity, calculated from the major and minor axis length of the lesion, measures how much the lesion deviates from being circular.
- The aspect ratio calculated as length to width ratio measures the elongation of the lesion.
- Convex area corresponds to the area of the convex hull which is the smallest convex polygon that contains the lesion.
- Solidity measures the overall concavity of the lesion means calculated ratio of convex hull area of the lesion.
- Roundness describes the lesion's resemblance to a circle. It is calculated using the area equivalent diameter and diameter of the smallest circle enclosing all the points of the lesion.

● Closeness, the ratio calculation of the different coordinate values such as major and minor axis measures how long and narrow the lesion is.

Color attributes provide the color of the disease spot as an important attribute for disease recognition. The mean (μ) average and standard deviation (σ) (set difference) of color component values of pixels in disease spots are measured as the color attributes. Color components of RGB, L*a*b*, HSV, YCbCr, and grey color space are considered, resulting in 26 attributes based on color to characterize the lesions.

4.14 Experimental design and results

Diseased rice leaf images for analysis and identifying the type of the disease are acquired from agricultural fields and official websites of the indian rice research institute (IRRI), louisiana rice crop information, rice knowledge management portal (RKMP), indian council for agricultural research (ICAR). The images from the field are collected at paddy breeding center, Tamilnadu agricultural university, Coimbatore, Tamilnadu, and krishi vignyan kendra, Gobichettipalayam, Erode, Tamilnadu, India under natural sunlight using a mobile phone with 16.0 mega pixel camera. 418 brown spots and 495 narrow brown spot diseased leaves are considered for analysis. In all cases, a JPEG image format with 72 bits is used.

4.14.1 Segmentation of disease spots

k-means clustering such as with a difference in green and red, difference in green and blue, and a* color component of pixels as input groups pixels as either healthy or diseased. The figure outlines the disease spots segmented from the infected rice leaf images shown in figure 4.7 using k-means.

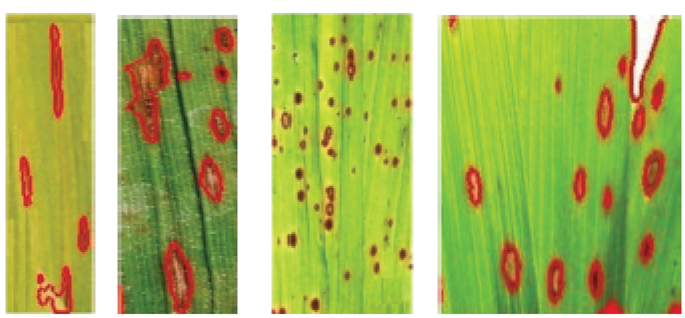


Figure 4.7 Segmented brown spot diseased by k-means clustering

Through the k-means cluster, segmented disease points, boundaries, and color attributes are extracted. From the comprehensive superset of attributes such as the RLDS dataset, a subset is chosen and fed as input to the classifier for disease recognition.

4.14.2 Attribute selection in RLDS dataset

Attribute subset selection is performed to select highly predictive attributes from extracted RLDS dataset. Proposed attribute selection techniques such as PSO with Greedy reset, FPO with Greedy crossover, and PSO with Greedy leap along with the One-all-random population initialization technique are applied to the RLDS dataset.

Table 4.12 Accuracy gain (AG) and dimensionality reduction (DR) by proposed HMDF attribute selection on RLDS dataset

Attribute Selection Technique	Without Attribute Selection (Average ± Standard Deviation)		Proposed (Average ± Standard Deviation)		AG(%)	DR(%)
	P(%)	#A	P(%)	#A		
PSO with Greedy reset	87.1±0.013	35±0	91.09±0.88	2.6±0.84	4.38	92.57
FPO with Greedy crossover			93.28±0.57	3.4±0.72	6.62	90.28
PSO with Greedy leap			95.72±0.82	2.2±0.52	9.00	93.71

From table 4.12 and figure 4.7, it is observed that PSO with Greedy leap and One-to-all random population initialization (9/93.71) shows superior performance (interims of AG/DR) compared to FPO with Greedy crossover (6.62/90.28) and PSO with Greedy reset (4.38/92.57) in terms of both accuracy gain and DR.

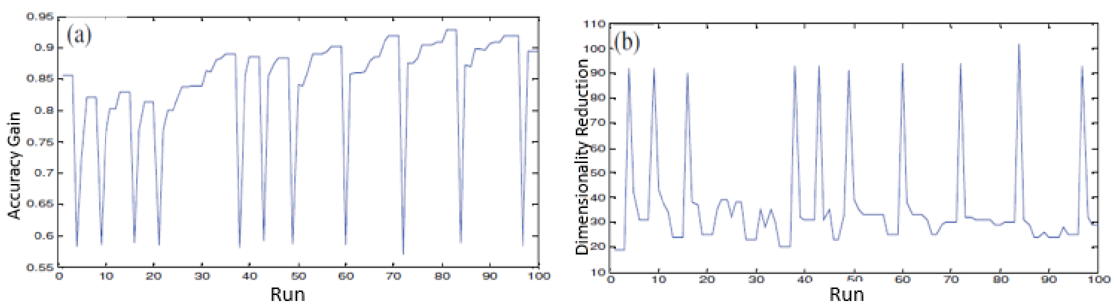


Figure 4.8 Accuracy gain (a) and Dimensionality reduction (b) that are realized in 100 runs of the musk dataset.

Table 4.13 Random population initialization Vs One to all random initialization computation time on RLDS Dataset.

Attribute Selection Technique	Traditional Random Initialization Time(s)	One to All Random Population Initialization Time(s)
PSO with Greedy reset	4251.5	2721.3
PSO with Greedy crossover	4118.3	2893.2
PSO with Greedy leap	4025.4	2357.2

Shape factor and standard deviation of CB color attribute are selected by attribute selection technique using PSO with Greedy leap and One-to-all random population initialization from RLDS dataset. PSO with Greedy leap requires 2357s to complete the process of attribute selection and outperforms Greedy reset and Greedy crossover techniques in terms of computational time.

4.14.3 Classification of disease spots

Selected attributes by proposed methodologies are feed as input to 6 traditional classifiers such as SIOC-CAS, KNN, SVM, NB, DT, and Ensemble learning to classify the type of disease spot in rice leaf and their results are analyzed and in table 4.14 tabulated all results of initialization technique.

Table 4.14 Six various types of classifiers and their predictive accuracy with attributes selected from RLDS dataset by proposed algorithms

Dataset	Reinitialization technique	SIOC-CAS (%)	KNN(%)	SVM(%)	NB(%)	DT(%)	Ensemble(%)
RLDS	Greedy reset	91.1	89.3	87.5	86.2	86.4	90.2
	Greedy crossover	91.5	91.2	90.5	91.0	90.3	90.5
	Greedy leap	92.0	91.5	90.2	89.3	91.1	90.8

From table 4.14 and figure 4.9, it is noticed that all six traditional classifiers yield in-par results with selected attributes as input, demonstrating the generality of the proposed attribute selection methodologies. Thus, disease spot in rice leaves are identified with a recognition rate of 95.3% such as the SIOC-CAS classifier with attributes selected using PSO with Greedy leap and One-to-all random population initialization technique. A machine vision system with nature-inspired attribute selection methodologies for classifying various rice plant diseases in rice leaves is developed. Disease spots are segmented such as k-means clustering to extract shape and color attributes such as RLDS dataset. Cognitive-based attribute selection methodologies using PSO with Greedy reset, FPO with Greedy crossover, and PSO with Greedy leap are employed to select significant attributes from the RLDS dataset for classification. Disease spots in rice leaf observed as a pattern recognition problem are solved and disease identification brown spot or narrow brown spot with classification accuracy of 92.0% is achieved.

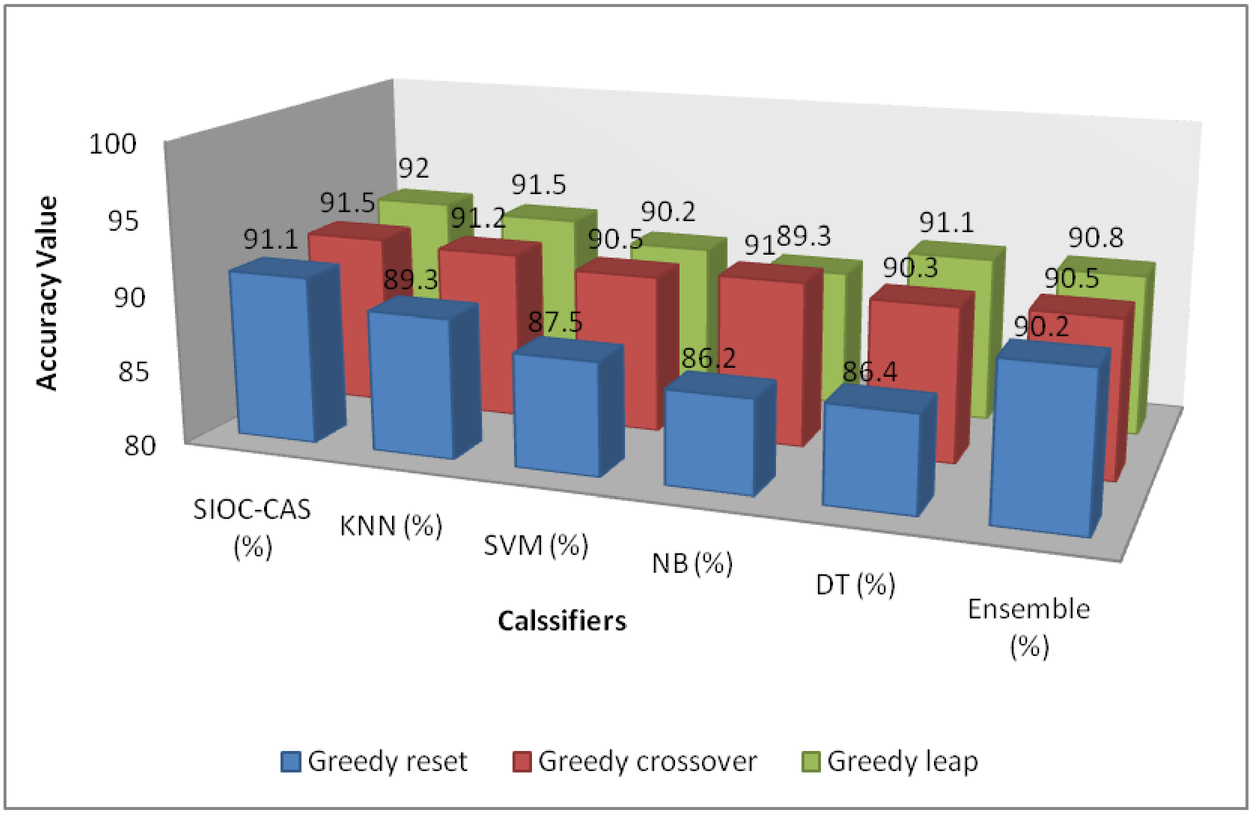


Figure 4.9 Predictive accuracy of six classifiers with attributes selected from RLDS dataset

Intelligent system design involves machine learning and pattern recognition, to mimic the decision-making capability of human beings to solve real-life problems. Pattern recognition classifies patterns based on a set of attributes extracted from them. The image classification method is an important process of pattern recognition and identification to classify each pattern into different classes. The attribute space of a pattern recognition problem is an indispensable factor influencing the performance of classification or learning algorithms. In the real world scenario, a large number of attributes is collected to better represent the data; however, not all the extracted attributes are useful. Irrelevant and redundant attributes significantly increase the computation cost of the classification process causing one of the major obstacles known as the curse of dimensionality.

Attribute selection is a data pre-processing technique and has become an important constituent of the pattern recognition process. The cognitive approach for attribute selection employs a subset search strategy and a learning algorithm like a classifier to assess the quality of each subset of attributes. This research engages the SIOC-CAS classifier with RBF kernel as the learning algorithm and randomized metaheuristic search algorithms as the search technique to identify highly predictive attributes in the problem space. PSO is an efficient global search technique and has been exhaustively employed by researchers as the search strategy in attribute selection. Also, FPO with its global walk capability traverses the problem space in search of a subset containing highly predictive attributes.

However, PSO and FPO are partial optimizers, resulting in pre-mature convergence to a sub-optimal solution and not traversing further and henceforth, leaving behind large areas unexplored in the problem space. Pieces of the literature suggest re-initialization of the stagnated best solution to a random point in the problem space if it remains unaltered for a prefixed number of runs. A well-defined heuristic-based re-initialization strategy is designed to overcome the problem of pre-mature convergence in PSO and FPO, as an alternative to random re-initialization. In addition, a population initialization strategy is devised to generate initial candidates to commence the search in diversified directions, as an alternative to random population initialization. Three cognitive-based attribute selection algorithms using PSO with Greedy reset, FPO with Greedy crossover, and PSO with Greedy leap are proposed to address the problem of partial optimization during the search process. One to all random initialization strategy is employed in all three attribute selection techniques to generate the initial population of the search algorithms, thereby improving the likelihood of identifying potential solutions from the problem space. An improved solution to the pattern recognition problem for disease spot identification in rice leaves is devised. One of the two main objectives in attribute selection that can be easily controlled is a count of attributes with minimal value attributes favored by the process of attribute selection. This heuristic is deployed in designing Greedy reset for restructuring Swarmbest solution on stagnation. The proposed SIOC-CAS attribute selection algorithm is tested on twelve benchmark datasets from the UCI repository.

AG and dimensionality reduction are then calculated by comparing the performance of the SIOC-CAS classifier with original and selected attributes as input. On attribute selection, the gain in accuracy ranges from 0.14% to 34.77%. HMDF dimensionality reduction of attribute space ranges from 58.94% to 98.4%. Attribute selection using FPO is devised and examined on its ability to escape local optima with its global walk characteristics. From simulation results, it is observed that pollenbest remains either unaltered throughout the search process or changes very minimally in the course of the search. There is no explicit crossover in FPO and this evolutionary heuristic is applied in developing the Greedy crossover mechanism to re-initialize pollenbest, on stagnation for a prefixed number of runs. Evolutionary jump derived from levy's flight is the underlying idea in designing cognitive-based attribute selection using PSO with Greedy leap strategy. Both the proposed algorithms, FPO with Greedy crossover and PSO with Greedy leap corroborated and evaluated with twelve rice plant disease datasets from the UCI repository and plantvillage. Similarly, PSO with Greedy leap achieves accuracy gain and dimensionality reduction in the range of 1.18% to 39.22% and 56.57% to 98% respectively. Also, it is observed that proposed methodologies outperform the literature on corresponding search algorithms for attribute selection, in most of the benchmark datasets representing classification problems. Finally, the classification accuracy of selected attributes was tested and verified using six traditional classifiers namely SIOC-CAS, KNN, NB, DT, DA, and Ensemble learning. It is noticed that all the classifiers yield similar results in terms of predictive accuracy, thus, substantiating the generality of the proposed algorithm as a pre-processing tool for classification.

Population initialization is the first step in any metaheuristic search process and can significantly influence the performance of the attribute selection algorithm. The evolutionary process of any population-based metaheuristic search technique is led by the best solution of the population. Hence, a good starting point means that a good leader appears earlier in the process of search, thus benefiting the entire evolutionary search process in finding potential solutions from problem space. A One to all random initialization technique is devised to generate initial solution vectors of the population representing each dimension of the attribute space. The performance of the attribute selection algorithms with One all random population initialization is evaluated in terms of accuracy gain, dimensionality reduction, and computational time. PSO with Greedy leap and one to all random initialization achieve an accuracy gain in the range of 1.76% to 40.05% and a dimensionality reduction in the range of 62.31% to 90% and out performs both PSO with Greedy reset and FPO with the Greedy crossover. Also, PSO with Greedy leap requires less computational time compared to PSO with Greedy reset and FPO with Greedy crossover.

All three proposed attribute selection algorithms with One to all random initialization are compared with other randomized metaheuristic algorithms employed as search strategies in the process of attribute selection. The predictive performance of proposed algorithms is validated with six classifiers and the results find that the proposed wrappers are general to different classification algorithms. All the proposed re-initialization algorithms and population initialization techniques require only a few simple steps for implementation. None of the proposed methodologies requires additional parameters for implementation. Also, proposed cognitive models are independent of the dimension of attribute space and can be applied to any dataset representing classification problems. Cognitive models with One to all random initialization require less computational time than that required by cognitive models with traditional population initialization techniques. All the proposed algorithms are generic and are not application specific.

- Heuristic-based re-initialization techniques can be applied to any optimization algorithm suffering from pre-mature convergence and a step further, to any application employing the optimization algorithm.
- The population initialization technique can be applied to any population-based optimization algorithm and any application with a discrete problem space.

All the proposed attribute selection methodologies with One to all random population initialization are applied as a pre-processing technique to classification in the process of solving agriculture-based pattern recognition problems. Rice leaf disease identification is solved using DIP and ML techniques. It is observed that on applying the proposed pre-processing technique, an accuracy gains of 9% and dimensionality reduction of 93.71% is achieved with a computational time of 2357s over random population initialization that takes 4025s for computation such as PSO with Greedy leap. SIOC-CAS classifier with selected attributes as input identifies brown spot and narrow spot disease with an accuracy of 92.0%.

CHAPTER 5

LAURENT SERIES WITH INTELLIGENT MULTIDIMENSIONAL OBJECT OPTIMIZATION (LIMO)

5.1 Introduction

Nowadays communication technology and image processing methods try to achieve effective outcomes using various advanced computational methodologies and techniques. Artificial intelligence, machine automation (MA), and the IoT provide the greatest support to the agricultural sector in plenty of ways to improve the quality of the cultivation process.

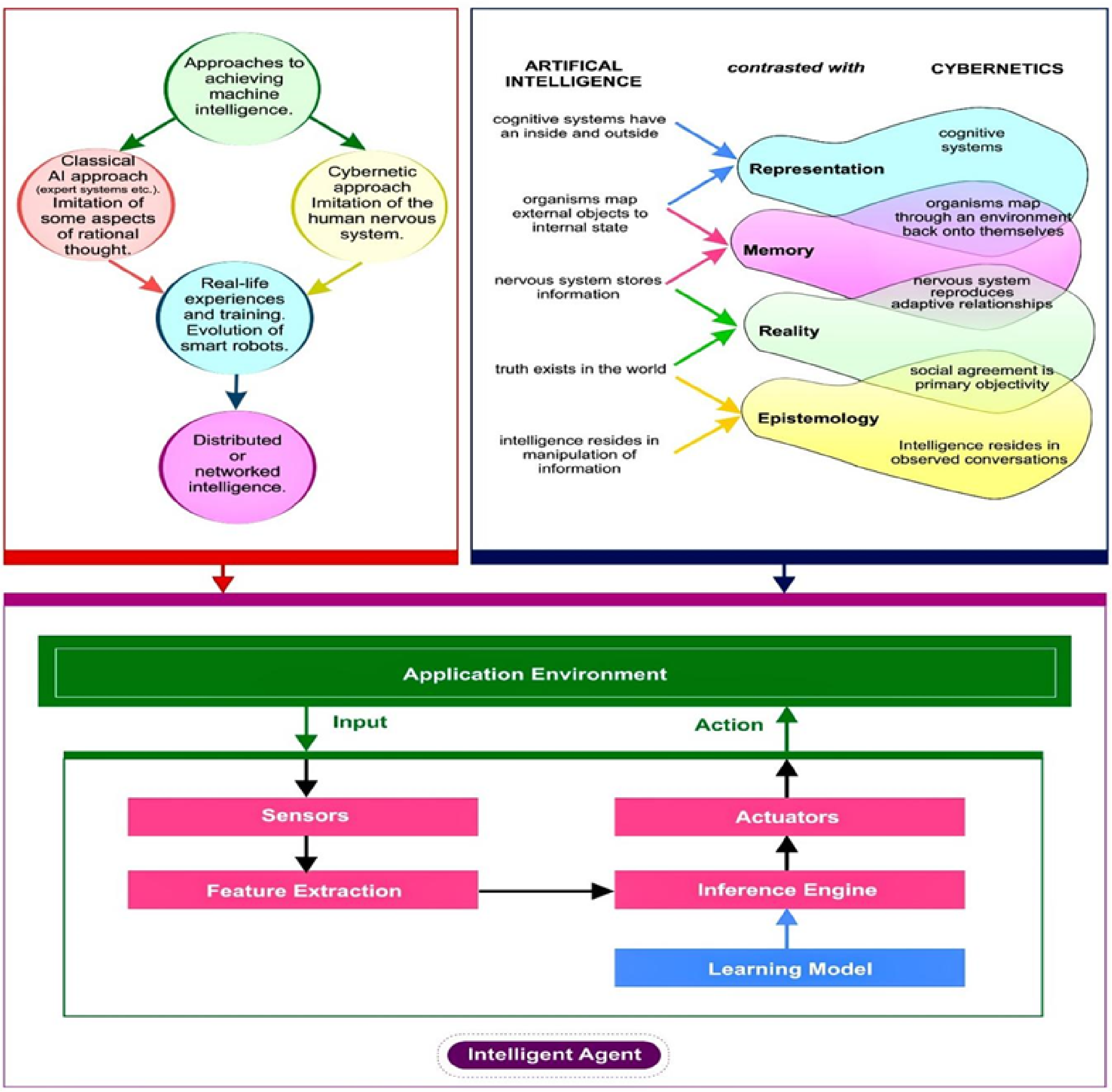


Figure 5.1 Conceptualization of cybernetics framework [130]

AI determined with DL and ML, through this equipped different applications for various crop diseases detection and prevention. Cybernetics is defined as a cumulative knowledge approach that mainly does pragmatic adjustment in complex image processing systems. Cybernetics outlines the decision-making for resource management, security, and privacy activities such as analysis or experience basis. Cybernetics deals with an advanced controlling technology that calculates comparisons and divergences present at dissimilar sampling that predict the probability of the expected hit rate and loss rate at the same timings with the same data fields. Cybernetics helps to take its activities, also this activity and its patterns can be modifiable with various controllers. Cybernetics plays a vital role to activate with the assistance of criticism to recognize complex models. In a feedback-based system, this recognition becomes a significant part of that model and can undergo any kind of managerial work by itself [136]. The comparison of intelligent multi-agent models in various cultivational applications for advanced farming figure 5.1 explains the AI-based conceptualization of the cybernetics framework.

5.2 Motivation

Our country's economy is majorly dependent on agriculture and its essential products. Nowadays intelligence technologies support finding a way to give solutions to food demands all over the world. With these technologies, humans activate self-reliant (autonomous), modernized, and smart production of required food items. Also, one more issue factor is food item security. Because of radical climate change, various crop diseases, water scarcity, and more. In the above list, various crop diseases create major problems and reduce a considerable percentage of agricultural productivity, food security, and income of small-scale farming people. So, curbing these issues at an earlier stage will raise productivity.

ML algorithms and fuzzy logic methods are providing support to raise the prediction percentage of crop disease identification, and prevention with an extraordinary accuracy rate. In this research work, our ultimate intention is to detect crop diseases with a maximum accuracy rate with the internet of things methodology. Nowadays many researchers concentrate on crop disease detection with deep learning, federal learning-based machine learning algorithms. Some deep learning models are deep brief networks (DBN), dilated residual networks (DRN), deep gated current unit (DGCU), and deep stacking network models (DSN) were used to predict several plant diseases with different dimensional inputs, GAN networks providing vital support for crop and farming field monitoring and data communication. This plant disease detection model has several problems rising on training datasets and algorithms, which can improve the preformation of image classification and overall performance with low loss function values of better convergence at real-time scenarios.

Herewith, an advanced optimized algorithm supports the LIMO model to train the rice crop disease dataset to a machine. This research works on rice crop disease detection and prevention with the help of GAN networks which provides parallel significant speed-up sample. Also, GAN is getting train data by LIMO's proposed framework which integrates the both Laurent series and Intelligent multidimensional object optimization algorithm and improves the overall hit rate and accuracy of rice crop disease detection and prevention.

5.2.1 Contribution

The foremost contribution of this research work is LIMO with GAN framework combination of laurent series, cognitive residual convolutional network, and intelligent multidimensional object optimization algorithm. The LIMO framework provides intellectual classification and optimization techniques which detect and prevent rice crop diseases very faster and more accurately by the proposed LIMO with GAN network model.

5.3 Literature Review/Comprehensive Analysis

Inclinations across this world are always innovative and creative. Around fifty-eight percent of our country's population fully depends on the cultivation process as the only option for their daily life. The Indian ministries which are supporting for agriculture and its regards have secured and assigned a goal for the agricultural department to disseminate 4763.5 billion INR by 2021-2022, values received through India brand equity foundation (IBEF) which is largest accurate information resource sharing center for an Indian agriculture, economy, investment ideas. According to the digitalization in agriculture information and study, the global market for automated agriculture productivities has assessed a value of 404.7 billion INR in the year 2016-2017. It's estimated to get raised to 1270.5 billion INR by the end of 2027-2030. The compound annual growth rate estimated that population growth is significantly increased by approximately 15% going to get raised with the current population between 2018 and 2027. Figure 5.2 demonstrates the increasing world population [137].

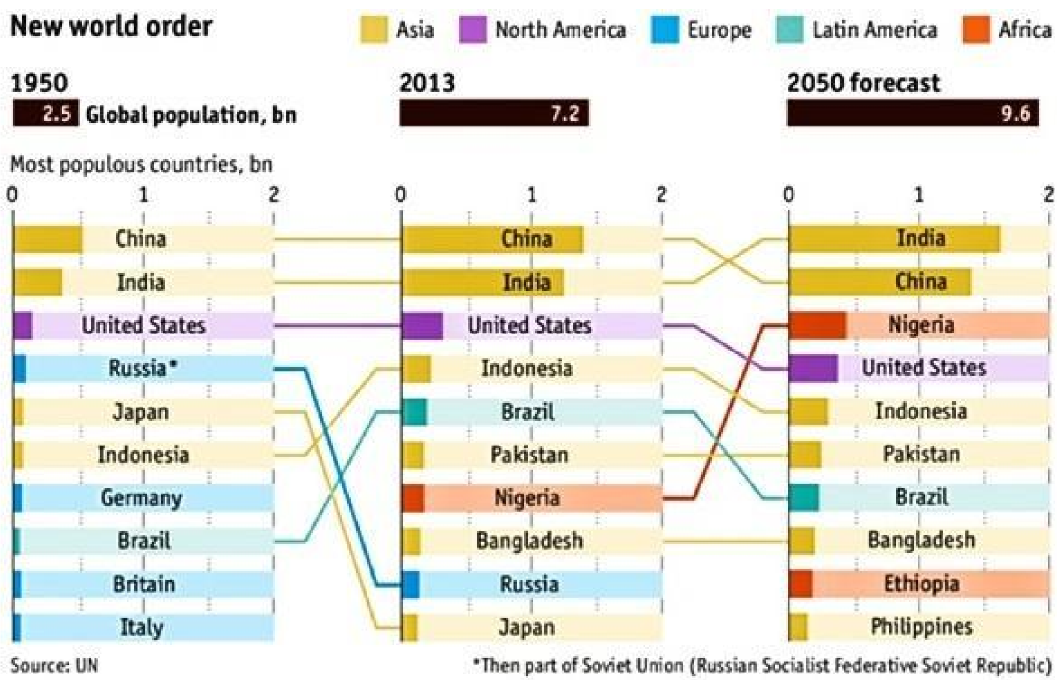


Figure 5.2 Estimating growth of world population rate [102]

The global economy of agricultural countries mainly hangs on its agricultural yield and way to reach good agriculturally based outcomes. Mainly agricultural productivity means the way of modernized farming, especially having to do detailed understanding and an effective scrutinize needed for crop disease detection and prevention, which is waging important role for increasing farming production. Crops getting diseases is quite normal, because of various natural and artificial factors. We have to pay attention to it properly to take a proper remedy. If proper prevention is provided for crop disease means, it will have measurable and crippling effects on crop yield rice plants, thereby getting a significantly decreasing percentage of agricultural productivity and production and standard of the product. In approximately the last 8 years, the demand for automated techniques like sensors, wireless networks, and clouds to detect and prevent rice plant disease methods has increased. Due to increasing space for researchers in smart agriculture and demand for farming items, agricultural plant disease detection and prevention.

Automatic rice crop disease detection and prevention techniques are playing a vital role to reduce monitoring area size before detecting plant diseases [138]. For this purpose, designed a segmentation algorithm to find rice plant diseases from some symptoms that arise in plant leaves or stern. Remote monitoring systems receive different values of crop and its field information through various sensors like temperature, humidity, moisture, crop images, total field area's gradients values. Image segmentation algorithms support the extraction of the features of fixed images like in this crop disease identification or detection system after segmentation with gradient values affected leaf identification, level of affected leaf or stern. This smart agriculture system is a fully automated-based network so all needed inputs are received through concerned sensors or electronic peripherals, this system maximum avoids manual interference. We were using plenty of sensors for various types of sensing information. Those sensors are categorized into three types, there are

- Controller-based
- Service based
- Communication-based

Controller-based sensors are controlling overall sensors and their interference which are used in that framework. Service-based sensors provide information to controller boards based on sensors such as IR, MEMS, color, rain sensor. Communication-based sensors support transmitting the values through a network. Smart agriculture systems mainly support Wireless Fidelity (WIFI) to transmit information to the target from a sensor. Sensors collect information on the farming fields, crop status, defined symptoms, environmental changes. and data gives some information to farmers to act correspondingly [139].

The IoT is one of the important technologies which needs to connect communication devices with the internet. It makes immense smart applications. The IoT has grown rapidly which is the reason many researchers are showing interest in this platform. Through this platform it is able to collect versatile information with various sensors for the application's sustainability. Various dimensional features are extracted and detect crop diseases in an effective manner. With this computational intelligence, can able to perform numerous sensors and communication devices for a smart agricultural process which improves extreme productivity. Cloud models, machine learning, and IoT materialization technologies are playing vital roles in field information collection for smart farming.

Sensors, actuators, network devices, and smart agricultural techniques support the building of an efficient farming model and its inclusive analysis of increased productivity. This analysis requires various important factors that provide inheritance information about previous functions and implementations. Crop disease detection exists with enough weighty literature review and various activities to tolerate new IoT models for farming. One more thing is the IoT platform provides user-friendly models for crop disease detection and prevention. The IoT model directly connects farmers such as the farming department with computation intelligence computation. Figure 5.3 explains the smart agriculture system.

Existing crop disease detection and prevention models maximally concentrate on anyone session in the overall system herewith we discuss restrictions and conditions. The deep learning algorithm supports a multi-context fusion network (MCFN) architected to detect problems on rice plant leaves through the IoT [140]. This model fails to detect fine-grained crops or diseases-affected crops. Also, the DL algorithm is employed with various factors including eco-friendly factors, crop types, water irrigation level, and soil fecundity to identify crop diseases [141].

This crop disease detection method effectively controls factors of farming by using various sensors with those values creating a CNN model trained dataset. This model amended overall framework is an effective and monitoring system basis with thermal camera images finding whether crop leaves are infected or not. Invented a new framework for the agricultural industry that performs edge computing for data transmission from IoT sensors to cloud storage. The main objective of this model is field monitoring, evaluating, and transmitting values to the cloud. Optimization also plays a vital role in this model, which means continuous monitoring and evaluating values communicated in the cloud on an assorted dairy circumstances basis. However, this model is absent to train dynamically the learning algorithms for miscellaneous crop diseases and symptoms.

IoT supports farm monitoring systems and predicts the prevalent diseases in crops. An IoT monitoring system reads the atmosphere values which help crops to grow properly. Wireless sensor networks provide the space to calculate and monitor the environmental data, through this evaluate the disease level of affected crops and prediction for disease spreading rate also. The main theme of this system is early prediction and taking immediate action to solve the issue. DNN segregates the crop and its disease rate of the crop then calculates the diagnosis level [142]. This DNN model uses three-level classification to differentiate the diagnosis of crop diseases from the three states mentioned healthy, semi-infected, and unhealthy using DNN Alexnet. This advanced framework provides markable improved performance on crop disease detection systems, but the main drawback of this system is the struggle with including various types of disease training data.

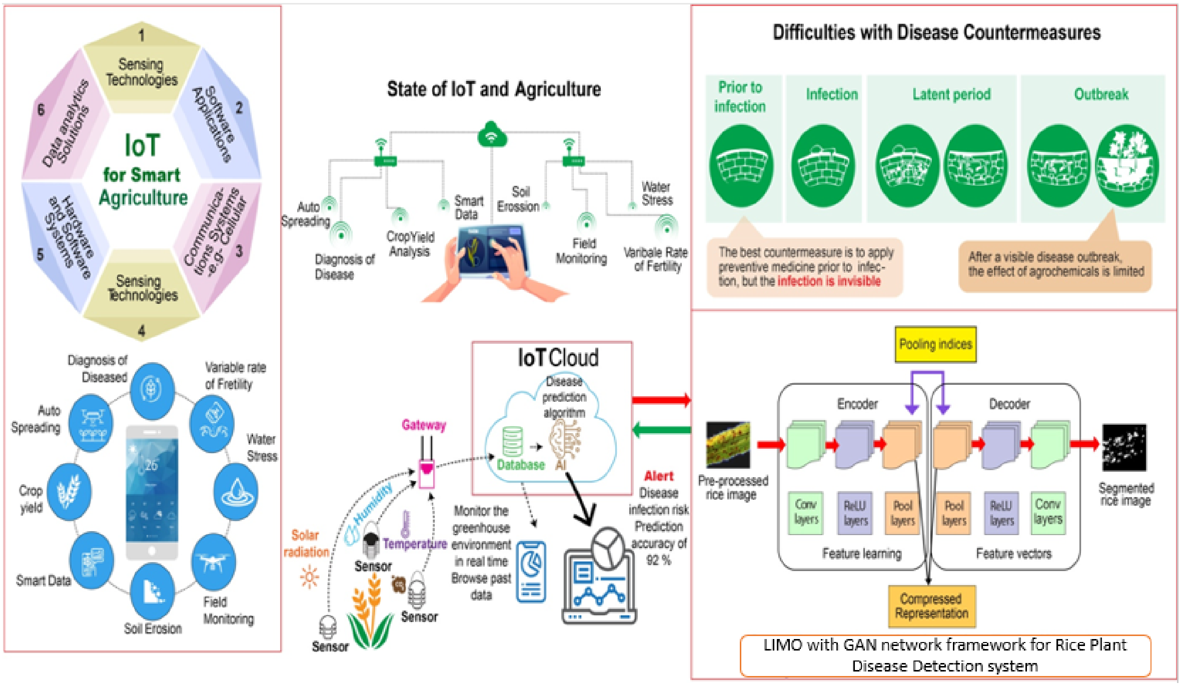


Figure 5.3 Overall structure of rice crop disease detection and its process

Streaming image inputs-based crop disease detection system to detect the crop diseases and their level was formulated with real-time crop disease detection. Video-based diseases detection model to recognize the rice plant diseases in real-time. At an initial level, the streaming image was converted into static margined multiple images. In the second step streamed multiple images diagnosed with object identification. Here this process also takes gradient values and provides mapping which is called image segmentation. At last, images were prepared for disease detection results. Video-oriented performance using different ML algorithms was positioned with important factor quality of image or video. The fast RCNN model accepts and uses blurred images or videos for disease detection with ML algorithms. RCNN and machine learning algorithms are mainly used to detect rice leaf blast crop disease on rice crop leaves [143]. This model encompasses with few levels only there are,

- Pre-processing (data transformation, data reduction, and feature engineering)
- Segmentation
- Features extraction
- Classification
- Result analysis and accuracy rate

The image detection process calculates the hue saturation value (HSV) of color and its gradient space occupied level. An image segmentation methodology extracting the preferable boundary regions and their pixel mapping patterns are calculated with different maximum range values. Hence detecting plant disease processes has to find the depth of diseases attacking the particular crop. This severity identification is also divided into three phases like networks legitimate, malicious, and an attacker. There are 1. Beginning phase such as an earlier stage which is very easy to take action against to get normal, 2. Middle phase such as infection started to spread. remedy actions somehow useful to stop the spread more, and 3. The final phase is the nastiest situation to control the disease. A computerized diagnosis framework for plant disease detection system using prototype methodology and formulated to identify the rice crop disease. The twin support vector machine model helps to improve the accuracy level classification but its absence to identify other crop diseases or symptoms.

Multi-dimensional features compensation RNN (MDFCRN) method implemented with natural language processing for effective object retrieval in the smart agriculture field [144]. MDFCRN indicates information to farming people after the identification of rice crop diseases. Also, MDFCRN improves productivity with needed information [145]. This MDFCRN model through the sensor receives plant leaf disease images that are refined with a median filter afterward the segmentation and classification process. Then, collect the information namely pixel rate, gradient values, threshold rates, and segmentation rate from images. The last process of image classification with the sine cosine algorithm with RNN (SCARNN) helps to identify the severity level of plant disease. SCARNN model advantage is used in limited resources for this disease detection process. But the disadvantage is that SCARNN missed to detect the various plant diseases. The CNN is always effective when using object identification or image comparison. For the same problem instead of CNN, we are able to use DNN which requires more input value compared with CNN for greater accuracy. Nowadays researchers focus on providing solutions for the prevention or detection of various plants and various disease detection in wild surroundings with a single framework with the help of different machine learning algorithms [146-153].

5.4 Proposed System Models

The Internet of things encompasses different types of sensors, actuators and other communication connectivity devices such as various smart devices also. All smart communication devices have communication among others for received information exchange through different channels shown in figure 5.4.

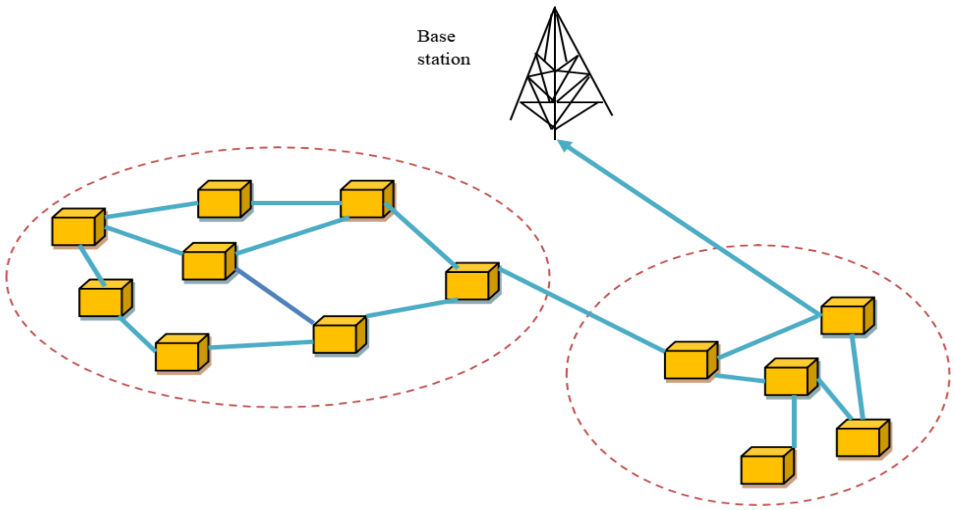


Figure 5.4 IoT network communication process

5.4.1 LIMO with GAN framework for rice crop disease detection

This research work describes an advanced object detection methodology of the LIMO framework with a GAN network for identifying rice crop diseases. LIMO and GAN framework developed with four phases, IoT communication phase, pre-processing with feature engineering, image processing with DNN phase and crop disease prevention and identification phase, finally performance calculation phase. In the IoT communication phase sensors and other devices are subject to network routing protocols. In this network communication phase path selection and decision-making are carried out by multi-objective and cognitive learning-based routing algorithms. Once communication started with MOCLEAR routing on board the unit get information about crop images. After receiving rice plant images regular image processing is performed. Feature engineering pre-processing starts the process of expert knowledge feds to the machine. Segmentation and feature extraction methods are attained input images through a sensor and IoT devices. Feature engineering to improve the image quality and transformation for comfortability. At last, received features are classified and optimized, wherein rice plant disease detection with LIMO framework. The proposed LIMO framework model is a combination of the Laurent series and intelligent multidimensional object optimization. Figure 5.5 shows the LIMO framework combined with the CRCNet network for rice crop disease detection for smart agriculture with an IoT network.

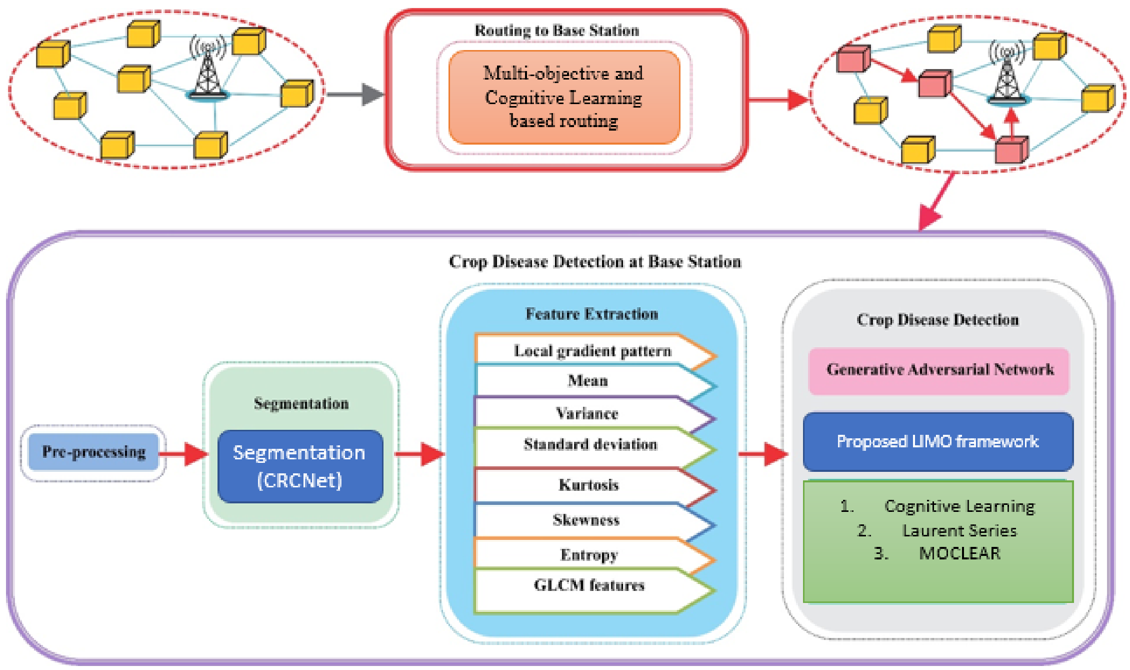


Figure 5.5 LIMO with GAN framework for rice crop disease detection using the internet of things

Assume rice crop disease dataset Rpd and $x$count of input values. Expressed as

$$Rpd = \{I_1, I_2, I_3, \ldots, I_x\} \quad (5.1)$$

where I_x indicates x^{th} plant images input.

Assume that the ith node or device holds the values about subjected field plants or crops to the IoT communication phase. Rice plant data is transmitted to an onboard unit (BS), MOCLEAR algorithm helps to find the best route for data communication. We will discuss more clearly the process of path selection namely routing algorithm and crop disease detection with cognitive learning.

5.5 Cognitive Learning-based MOCLEAR Algorithm for Data Communication

In the data, the communication phase received data is transmitted to the onboard unit by selecting the optimal path using the cognitive learning-based MOCLEAR algorithm. Effective and lossless data communication is archived with the help of the MOCLEAR algorithm which is a combination of IMO and fractional gravitational search algorithm (FGSA). IMO algorithm implements grey wolf hunting behavior (GWHB) concepts [154-161]. GWHB has three stages, approach, hunt, and attack stages. A cognitive expert position is created, and after positioning every legitimate move gets updated. Intelligent multi-objective optimization has four set solutions alpha beta omega and delta such as expert position. Expert position gets updated with the gravitational search algorithm also. The MOCLEAR algorithm is a deciding model to choose the best route for data communication to the onboard unit. This hybrid model is the most advantageous for finding n number of routes for data communication using convergence. Updating process archived which way that terms are included to IMO optimization technique with the help of FGSA algorithm. Modified MOCLEAR expression is,

$$GW(t+1) = \frac{GW_1+GW_2+GW_3+GW_4}{4} \tag{5.2}$$

Where GW_1, GW_2, $and\ GW_3$ indicate the initial and continued positions of grey wolves. Also GW_4 denotes at the tth time, grey wolves end position with FGSA. Grey wolf representation is,

$$GW_1 = GW_\alpha - I_1(D_\alpha) \tag{5.3}$$

$$GW_2 = GW_\beta - I_2(D_\beta) \tag{5.4}$$

$$GW_3 = GW_\delta - I_1(D_\delta) \tag{5.5}$$

Where, GW_α, GW_β, and GW_δ denote the optimal solution. D_α indicates the gap between the position and the evaluated expert position. D_β represents the reverse of D_α the gap between the evaluated position and the expert position is β. M_δ indicates a distance between cognitive expert δ and prediction-based position. GW_4 indicates the position of the fractional grey search algorithm with the time of t, estimation equation is,

$$GW_4 = \gamma C_j^i(t) + T^i(t+1) + \frac{1}{2}\gamma C_j^i(t-1) \tag{5.6}$$

Where, $C_j^i(t)$ denotes the position of expert i at j^{th} factor at t^{th} time, whereas the expert i's evaluated front position at j^{th} dimension in $t-1^{th}$time is indicated by $C_j^i(t-1)$. Here, $T^i(t+1)$ denotes velocity with $(t+1)^{th}$ location, while γ represents probability values from 0 to 1. Thus, rice plant data gained through sensors or cameras and other communication devices are given to the onboard unit, where rice plant disease detection and prevention is implemented.

5.5.1 Identification/detection of rice crop diseases

The rice crop disease detection and prevention with sensed rice crop information through IoT. After receiving values or images followed by the processes are image transformation, denoising, data reduction, image segmentation, features extraction, enhancement, image classification with optimization, and performance calculation for future machine training purposes. Image segmentation is achieved with CRCNet architecture for image localization. It is attained with segments that support significant and efficient object detection of feature extraction. At last, LIMO proposed a framework performed with GAN network MOCLEAR that trained the model with the LIMO algorithm. A detailed design of LIMO and MOCLEAR for rice crop disease detection will be discussed in an impending position of this research work.

5.5.2 Image pre-processing with image enhancement

First, the image acquisition process collected all field and crop-oriented images and posting to pre-processing. Then, the Image pre-processing method helps segment creations to design pixel masking of every object which is present in an image. It means an image is converted into a suitable image. In simple explanation pre-processing supports image smoothening, transformation, and image reduction to improve image quality for better object detection. Pre-processing plays an important role in image processing, it increases the probability to detect objects accurately. Image enhancement calculating noise ratio level, dimensional value, HSV, smoothening, grey scale value, size, brightness values, and sharpening images. Moreover, at the time of pre-processing data reduction supports reducing unnecessary noise and information from an image. After that, image restoration reduces noise and blur and orients things from an input image. This process increases the appearance of the rice crop image with better quality, which increases the accuracy of the rice plant disease detection and prevention process.

5.5.3 Features extraction and segmentation

After the image enhancement and restoration process then segmentation, which is supported to offer high dimensional segmented information. Segmentation is partitioning which makes it easier for the image reader to understand the object's boundaries. For the rice plant disease detection model, a cognitive network is provided to detect the boundaries range of disease information collected from each segment of crop images. Segmentation is partitioning the images to enable the user to detect the object and extract the important feature which decides the types of disease and disease stages using a cognitive network. Moreover, the cognitive expert provides support to decide on information extracted from each pixel. Divide and conquer method provisions to cognitive experts, operational images convert it into multi segments partitioning and extract the best fit in it. The LIMO framework classifies every pixel and its values based on cognitive decision-making with a pre-determined group. CRCNet is contained in the expert's direction of encoder and decoder with feature learning and feature vectors. Also, the cognitive network creates intellectual segments to improve the accuracy of a disease detection system. Consider the intellectual segments created from input crop images,

$$IS = \{IS_1, IS_2, \ldots, IS_e, \ldots, IS_f\} \tag{5.7}$$

where f denotes overall segments surviving on input rice plant image, IS_e represents e^{th} segment of the input. Feature learning such as encoder includes convolutional layers, ReLU layers, and pool layers-based segmentation avoiding the curse of dimensionality issues. Machine learning starts its process of training sets creation with weights values (w). The feature learning process consists of corresponding feature vector rates for the reconstruction of the segmented image. Convolutional layer converts the images (300X300) with RGB values (300X300X3) and then continuously does convolution processes with filtering, slicing, pooling, and padding. After pooling values send to feature vectors with softmax class values to create the probabilities of multi classes for image pixels for both types of pixels. Next, an element-based CNN activation function of ReLU is adapted with sub-sampled. ReLU is used to confirm the efficiency and implication of the ReLU layer when maintaining a set of vast networks and their weights. In pool layers, no need to train bias and neuron weight values, this process decreases the train data complexity.

The max pooling and padding process starts the convolution with different slicing and integrates regional filters. Then the reverse process of feature learning is supported to sub-sample procedural feature maps with trained map values and trained through maximum pooling. Feature mapping methods convolved with comfortable feature vectors and filters such as slice rate to increase the map DenseNet. At last, creates final feature vector values which expose the final vector class values to the probability vector. Softmax classifier is a self-governing classifier used to segregate image pixels. The final result created by softmax classes is SC, which represents the number of softmax classes. Estimated segmentation boundaries with SC and it provides a high probability vector rate for every pixel. The CRCNet supports segmentation slicing and padding for generating segments. CRCNet is effective as it takes filters, slice rate, max pooling, feature maps, and vectors used for better accuracy.

5.5.4 Image feature extractions

One receives the segments of a particular image and then has to start to acquire features from the segmented image. The extracted features are categorized into various types which helps to classify crop disease identification accurately and quickly. Here we were calculating various types of features and their extracted values. They are 1. Entropy, 2. Inverse difference moment, 3. Angular sec moment, 4. LGP (Local Gradient Pattern), and 5. Variance. For rice plant disease detection mainly collecting GLCM, statistical and texture features to make results betterment and improve detection accuracy [162-164]. Feature extraction method calculated the above-mentioned feature values from an input image and its description following eight steps:

- Mean: Mean values calculated with convolutional-based numbers pixels. The mean value (T_1) calculated from the sum of segments present in input is divided by the total values of segment values.

$$Mean_{seg} = \frac{1}{\left|d\left(S_f\right)\right|} \times \sum_{f=1}^{\left|d\left(S_f\right)\right|} d\left(S_f\right) \tag{5.8}$$

where f represents the total segment in an input image, $d\left(S_f\right)$ denotes single-segment pixel rates, $\left|and\ d\left(S_f\right)\right|$ is the cumulative value of pixels present in the segment.

- Variance: It is calculating pixel spreading rate with mean values $Mean_{seg}$.

$$Variance_{seg} = \frac{\sum_{f=1}^{IS} * \left|S_f - \frac{1}{\left|d\left(S_f\right)\right|} \times \sum_{f=1}^{\left|d\left(S_f\right)\right|} d\left(S_f\right)\right|}{(IS)} \tag{5.9}$$

- Standard deviation: SD supports descriptive statistical measures; an advanced variance is SD which estimates random deviation expressed as an amplitude rate.

$$Standard\ Deviation_{seg} = \sqrt{\frac{\sum_{f=1}^{IS} * \left|S_f - \frac{1}{\left|d\left(S_f\right)\right|} \times \sum_{f=1}^{\left|d\left(S_f\right)\right|} d\left(S_f\right)\right|}{(IS)}} \tag{5.10}$$

- Skewness: from variance collected amount of variability and through skewness provide direction of V. Skewness is the lack of symmetry for frequency distribution such as affected level of plant leaf and its boundary sharp using a numerical value.
- Kurtosis: Kurtosis is measuring outliers and lacks outlier values. Kurtosis provides sharpened values of probability and its deviation like peak, saturation, and low.
- Entropy: $Entropy_{prob}$measures the count of bits needed for encoding image data such as conditional entropy. It provides the value of complexity in a certain portion of an image. It measures that is operated to find certain and uncertain information about an image. This value castoffs between mutual data in every entropy. Entropy calculates the variation

between adjacent pixels or clusters such as segments. Additionally, it is defined as a corresponding complexity level based on trained values. Entropy image segments are computerized with analytical quantities to get image and pixel information. Also, entropy calculates an actual probability of hit rate and miss rate. Later, entropy is appraised to attain the probability of intensity level.

$$Entropy_{prob} =- \ Q \log log\ (\ Q) \tag{5.11}$$

where Q denotes the probability of encoding image information and discrete value.

- Local Gradient Pattern: LGP acclimates the filter value and curve representation of gradient patterns in a group of pixels [42] in current research. Local gradient pattern calculations,

$$T_7^{q,l} = LGP_{\delta,\alpha}^{u^l,v^l} = \sum_{n=0}^{\beta-1} \eta\left[\bar{\mu}_n - \bar{\mu}^*\right] \times 2^n \tag{5.12}$$

$$\eta(s) = \{0 \quad ; \quad s < 0\ 1 \quad ; \quad otherwise \tag{5.13}$$

where μ^* and $\bar{\mu}_n$ denotes mean gradient and center pixel point $\left(\lambda^*\right)$ situated at $\left(u^l, v^l\right)$ it is neighboring pixels (λ) shows as,

$$\bar{\mu}_n = \left|\lambda - \lambda^*\right| \tag{5.14}$$

Where λ^* dynamic threshold fixing with cognitive values for code generation. The range of color values was plotted and it's expressed as [3×γ] 'γ' represents bins count.

- GLCM statistical texture feature: GLCM defines statistical 2nd order values and texture features are developed by distinctive greyscale input images using location, position, and spatial information. Each pixel rate calculates cooccurrence combination frequencies for improving the quality of feature extraction. GLCM defines combination frequencies of all possible occurrences that are calculated with image grey-level value. GLCM is used to maintain the pixel position and explicit pixel location stored through the two-dimensional (2D) array.

$$GLCM_{feq} = P(h_1, h_2, ..., h_t) = \frac{A_m(h_1, h_2, ..., h_t)}{\sum_{z_1}^{N_h-1} \sum_{z_2}^{N_h-1} A_m(h_1, h_2, ..., h_t)} \tag{5.15}$$

where, N_h represent the total grey levels in an input image, it is calculated with t^{th} order GLCM statistical feature rate. z_1 and z_2 denote the first and second positions and locations of gray level values. A_m describes the Angular moment and Inverse difference moment of every pixel value. $P(h_1, h_2, ..., h_t)$ represents a population of explicit pixel positions.

5.5.5 Implementation of image decoding

In this image decoder each group of explicit pixels or segments is mean, variance, standard deviation, skewness, kurtosis, entropy, local gradient pattern, and grey level co-occurrence matrix. In equation 5.15, a set of important features are exposed from the input image, they are GLCM, statistical, and texture image features. Equation 5.16 image decoding is the processing of received feature values from the image encoder which provides proper information to take a final decision to predict crop diseases. Unique attributes and vital attributes such as features are having max prediction rate to increase the accuracy of the optimization algorithm and cognitive learning algorithm by doing image enhancement, restoration, and image classification. This extraction process receives values from each segment present in an input image.

$$Img\_dec = \{Attr\} \tag{5.16}$$

Attr= $Mean_{seg}$, $Variace_{seg}$, SD_{seg}, $skew_{sharo}$, $kurtosis_{peak}$, $Entropy_{prob}$, LGP_{centre_value}, $GLCM_{feq}$

Image decoder Img_dec inserts values into GAN, which classifies with statistical texture values, which softmax classes are trained. The rice crop is segregated by the classifier as a level of infection occurs on the rice plant in a corresponding input image.

5.6 LIMO with GAN network for Rice Crop Disease Prevention and Detection

GAN network working combined with AI and ML algorithms especially, we are recommending cognitive learning algorithms which have AI-based decision-making systems based on the environment and use. GAN network involves two types of neural network models, generator and discriminator. Generators make a prediction or approximate samples on an original or expected outcome basis. Discriminator finds the variation from regular activities. This process between generator and discriminator models leads up to the levels of perfection. Various optimum values are received through different algorithms which support the training of crop disease images and compiling the framework. Here an AI-based decision-making system supports choosing optimal values based on prediction and fuzzy basis. The main objective of the optimization technique in the neural network is to fine-tune the neuron weights to improve the hit rate ratio. Figure 5.6 explains the flow of LIMO with the GAN network to detect rice plant diseases.

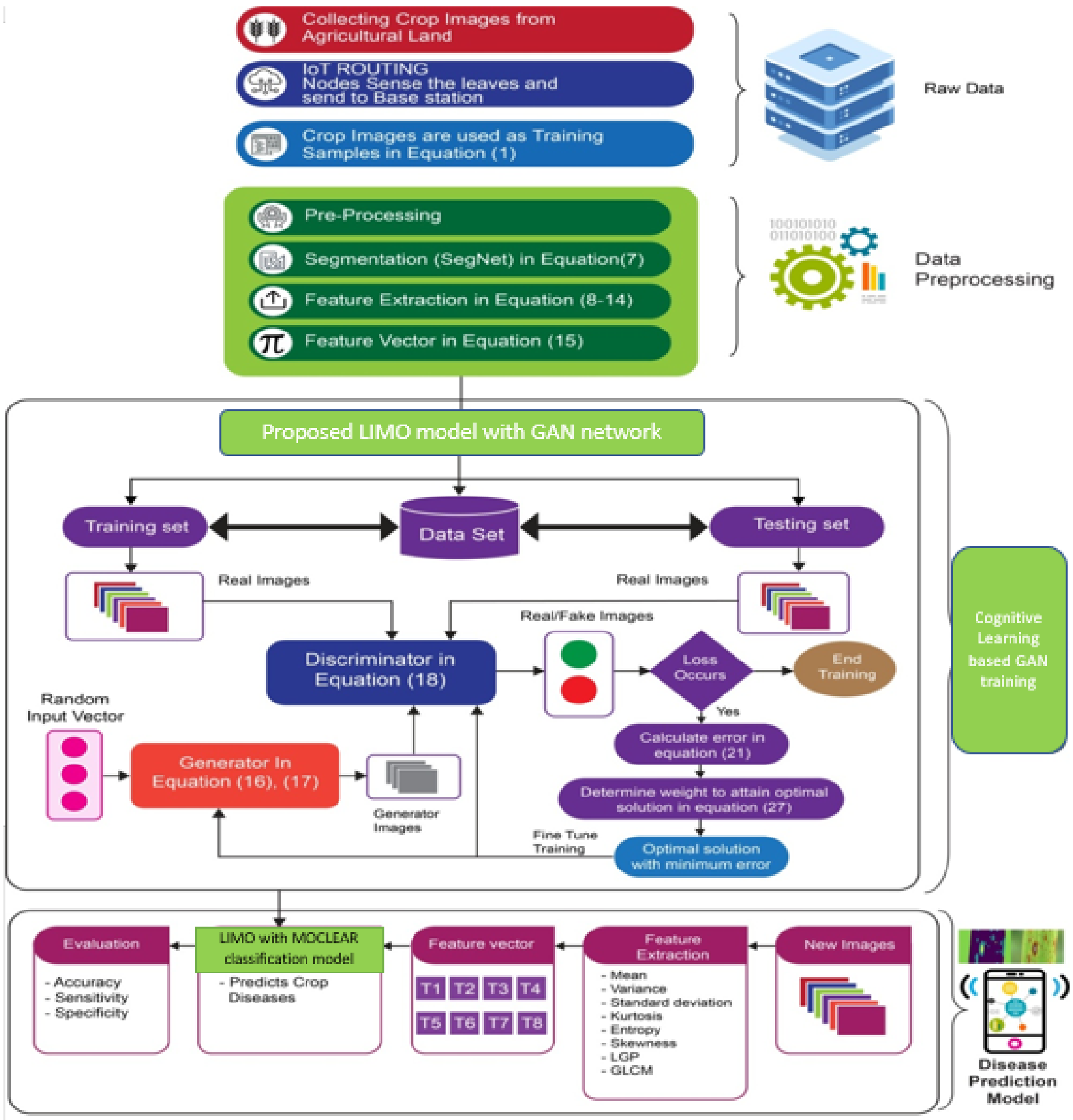


Figure 5.6 Rice plant disease detection using LIMO with GAN framework

Rice plant disease detection using an image decoder such as feature vector which follows eight steps to do feature extraction and forwarded data to the LIMO classification framework which is combined with the GAN network. LIMO trains the dataset with help of feature-extracted attributes and their values. This proposed LIMO framework to identify the rice crop disease with intellectual multi-object optimization and MOCLEAR. Laurent series implements the methods of complex variables with the addition of infinite terms and extensions. For image processing and object detection, this LIMO framework is an effective and powerful detection method because of computational integration and infinite sum creation. Moreover, this LIMO series is a one step process that monitors with high dimensional terms. This laurent series supports the derivation of imaginary upper error and convergence. Laurent serious, DNN and

AI-based optimization (IMO) enables a high classification accuracy rate compared to existing optimizations. IMO optimization employs three functions propagation, diversion, and contravention which improve the quality of high-ordered solutions on optimization issues. Also, the IMO model maintains stability between manipulation and investigation. IMO with Laurent seriously increases the optimal way for conjunction process, most favorable ground solutions, and symmetric between trained and test data. Here integrated IMO optimization and Laurent series to raise the overall throughput significantly. Will discuss the GAN network and its architecture used in this rice plant disease detection system.

5.7 Generative Adversarial Network structural design

The image decoder Img_{dec} gives feature extracted values into IMO classification to identify rice crop disease. GAN is a DL classifier that obtains an optimal process on object detection of rice plant disease. This network performs with high precision and recall values for object detection based on high-order complexity and its phenomenon. The previous method attempts matching with the latter process of comparing or identifying no matching data such as fake data in real-time databases/datasets. It confirms the generative, discriminative, and combinational values made at the time of training inputs. The generative component provides wised data training with cognitive learning and joint distribution for prediction betterment. Sample data are mapped with a generative adversarial network following joint distribution. The rice plant diseases are accurately identified at the time of mapping. Take data points as Img_{dec}in which the image decoder provides values to a classifier. Assume r represents high ordered random variable, E_a represents the generative joint distribution model, and E_{data} defines the real-time distribution random variable E_o. Equation 5.17 and 5.18 shows a generative mapping of an encoder feature vectors Img_{dec} which is epitomized G(T) in a symbolic representation. Moreover, GAN functions of posterior, likelihood, and conditional probability represent generative $I(.)$ and discrimination function $J(.)$. For the classifier, the feature vector value of $R(loss,\ predict)$ as,

$$R(loss, predict) = U_{O\sim V_{data}}[\log log\, J\,(W)] + Q_{O\sim V_g}[\log log\,(1\,-\,J(W))\,] \tag{5.17}$$

$$R(loss,\ predict) = Q_{O\sim V_{data}}[\log log\, J\,(W)] + Q_{O\sim V_s}[\log log\,(1\,-\,J(I(w)))\,] \tag{5.18}$$

In this 5.17 equation, the ReLU activation function is represented by $J(W)$. The discriminator model provides support to convolutional neurons and computational outcomes. Then $I(w)$ estimates the synthetic joint distribution value. Here, $Q_{I\sim K}$ denotes fuzzy and prediction set generation with random information K distribution samples. Also calculates the loss function rate of LIMO classification by cross-validation and entropy [134]. Discriminator loss functions as,

$$Loss_{fun} = -\frac{1}{\vartheta}\sum_{\rho=1}^{\vartheta} G_{\rho} \log log\left(J\left(W_{\rho}\right)\right) - \frac{1}{\vartheta}\sum_{\rho=1}^{\vartheta}\left(1 - G_{\vartheta}\right) \log log\left(1 - J\left(W_{\vartheta}\right)\right) \tag{5.19}$$

Above equation, ϑ represents a count of samples. The generative model helps to reduce entropy rate with discriminator gain and loss function $Loss_{fun}$ as,

$$Loss_{fun} = \max_{P} . R(J, I) \tag{5.20}$$

GAN network is trained on various types of rice crop disease datasets using the proposed framework LIMO.

5.8 The proposed LIMO with GAN network Intelligent Multidimensional Object Optimization Algorithm

Figure 5.6 describes the bias and weights for LIMO classification made to undergo learning with experience data by the LIMO framework to create an optimal solution. LIMO with GAN network adapts by combinational laurent series with multi-dimensional object optimization to choose the better weights for effective classification. For the rice plant disease detection system used, IMO optimization is proposed through this research work. LIMO proposed framework gives a training number of possible data meantime sorting based on the optimal solutions. The combination of the laurent series and IMO optimization method used to detect the object and iteration-based learning system provide deep knowledge to the classifier to classify disease types and levels with maximum accuracy rate. Step-by-step procedure for rice plant disease detection,

Initialization: at first initiates collective solutions,

$$Y = \{Y_1, Y_2, \cdots, Y_{\nu}, \cdots, Y_{\kappa}\}; 1 \le \nu \le \kappa \tag{5.21}$$

where κ denotes the count of solutions, Y_{ν} indicates the v^{th} solution of Y.

Computational values of hit and miss rate: To find the optimal solution is attained with miss rate and hit rate. LIMO support divided and conquer-based huge problem splits into a small problem then collective small solutions top ordered based optimal solution is finalized. Mean square error helps to finalize the best solution with the following formula,

$$MS_{err} = \frac{1}{c}\sum_{d=1}^{c}\left[\xi_d - \xi_d^{*}\right]^2 \tag{5.22}$$

where ξ_d denotes the expected overall solutions prediction basis, ξ_d^{*} is fuzzy and prediction-based LIMO expected outcomes and c represents a count of samples $1 < d \leq c$. Through various combinations of solutions and their optimal solution, IMO is compared and chooses a closer value of a trained set. Then LIMO model,

$$Y_x^{'} = Y_x + Rand(-1, 1)\lambda X_x \tag{5.23}$$

The size of x^{th} search space for high ordered dimensional value. For global problems obtained a globally optimal solution to identify rice plant disease detection with laurent series-based deep solutions. For updating pixel location values, segmentation boundary rate and position values are used to calculate a high optimal solution with the proposed LIMO framework, where apply following values in equation a1=1.3591, a2=0.6795, a3=0.2259, a4=0.0555, a5=0.0104, a6=1.38, and a7=9.9.

$$Y_x^{'} = 0.5Y_x + a1Y_x^{y-1} - a1Y_x^{y-2} + a2Y_x^{y-3} - a3Y_x^{y-4} + a4Y_x^{y-5} - a5Y_x^{y-6} + a6e^{-3}Y_x^{y-7} - a7e^{-5}Y_x^{y-8} \tag{5.24}$$

$$Y_x = 2\left[Y_x^{'} - a1Y_x^{y-1} + a1Y_x^{y-2} - a2Y_x^{y-3} + a3Y_x^{y-4} - a4Y_x^{y-5} + a5Y_x^{y-6} - a6e^{-3}Y_x^{y-7} + a7e^{-5}Y_x^{y-8}\right] \tag{5.25}$$

Rewrite equation (5.22) with equation (5.24),

$$Y_x^{'} = 2\left[Y_x^{'} - a1Y_x^{y-1} + a1Y_x^{y-2} - a2Y_x^{y-3} + a3Y_x^{y-4} - a4Y_x^{y-5} + a5Y_x^{y-6} - a6e^{-3}Y_x^{y-7} + a7e^{-5}Y_x^{y-8}\right] \tag{5.26}$$

$$2Y_x^{'} - Y_x^{'} = 2\left[a1Y_x^{y-1} - a1Y_x^{y-2} + a2Y_x^{y-3} - a3Y_x^{y-4} + a4Y_x^{y-5} - a5Y_x^{y-6} + a6e^{-3}Y_x^{y-7} - a7e^{-5}Y_x^{y-8}\right] \tag{5.27}$$

The updated expression for the proposed framework LIMO is,

$$Y_x^{'} = 2\left[a1Y_x^{y-1} - a1Y_x^{y-2} + a2Y_x^{y-3} - a3Y_x^{y-4} + a4Y_x^{y-5} - a5Y_x^{y-6} + a6e^{-3}Y_x^{y-7} - a7e^{-5}Y_x^{y-8}\right]$$

(5.28)

Accuracy, precision-recall values for optimal solutions are optimal solution received after many analyses with MSE error rate, positive and negative values calculated with hit and miss rate modified for a training set in the GAN network. The following pseudo-code describes the LIMO framework,

Pseudo code: Proposed LIMO framework algorithm

Input: *Optimal segment, feature selection values for collection of optimal solutions such* $Y = \{Y_1, Y_2, \cdots, Y_\nu, \cdots, Y_\kappa\}$; $1 \le \nu \le \kappa$

Output: Optimal solution M^* *(LIMO Decision making and classification)*

begin

Initialization of generative and discriminative values, κsolutions (RAND)

Computation of hit rate and error rate

LIMO initiate cognitive decision making from Img_dec*, gen(k), dic(k)*

while *condition fails, input values forward to pre-processing else if* ***do***

for *solution*$Y \in P$***do***

PROMOTE Y solutions to new $Y^{'}$*, IMO decision*

if $(Loss_{fun} > p(y - y`))$

if $the\ function\ of\ updated\ solutions\ compared\ with\ LIMO\ solutions)$ ***then***

exit $Y^{'}$ *into solution*

Updating $from\ Y^* to Y^{'}$

Replace$Y$*with* $Y^{'}$

else

$updated\ solution\ (Y.l) = Optimal\_func(Loss_{fun}, Img_{dec},\ gen(k),\ dic(k));$

if update solution $Iter(val,\ t, 0) = 0$ ***then***

Calculates miss_rate, hit_rate and MSE

Update solutions, MSE rate

$$MS_{err} = \frac{1}{c}\sum_{d=1}^{c}\left[\xi_d - \xi_d^*\right]^2$$

return Y`

end

5.9 Performance Metrics and Analysis

It provides comprehensive detail on the performance evaluation of our proposed LIMO classification against typical strategies using a rice crop disease dataset such as pre-templated that calculates different factors such as precision rate, recall values, F1-score, accuracy, sensitivity, and specificity percentage. Additionally calculating energy consumption by sensor network nodes and throughput different numbers of repetitions. Also, the proposed framework LIMO was analyzed with different factors and it was implemented with Matlab software which operated on experimental setup such as high-end model, win 10 operating system, i5 processor with 8GB RAM storage, for other image and dataset storage used SSD. Network parameters are nodes check-up rate 500 counts, adhoc on-demand multiple vector protocol (AOMDV) used for communication. Technical parameters activation function ReLU, epoch rate 85, machine learning rate 0.00158, and kernel range 4-6.

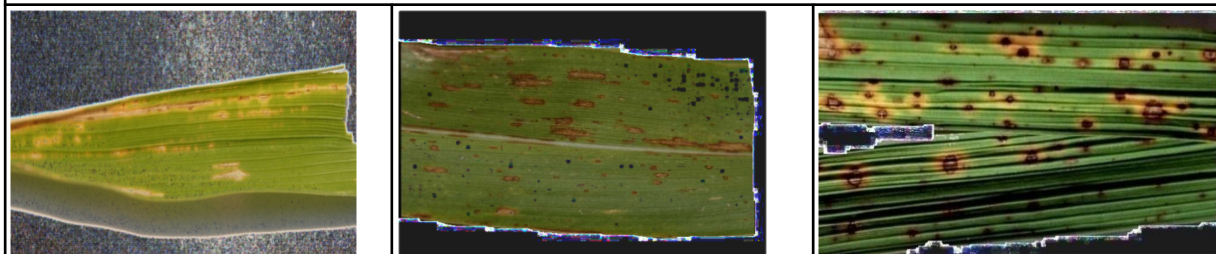
5.9.1 Dataset description

Result analysis and its performance metrics calculated with rice crop diseases data set. The rice crop disease dataset contains various rice crop disease images such as example: bacterial blight, BHP, bacterial leaf streak, leaf blast, collar blast, false smut, hipsa, rice grassy stunt, rice ragged stunt, brown spot, leaf scald, narrow brown spot, around 2400 various crop disease image dataset used for this experiment. This dataset is used in three stages: first pre-processing, second image enhancement and segmentation, last one is classification methodology, and finally analysis with calculated values from above.

5.9.2 Performance measures and experimental results

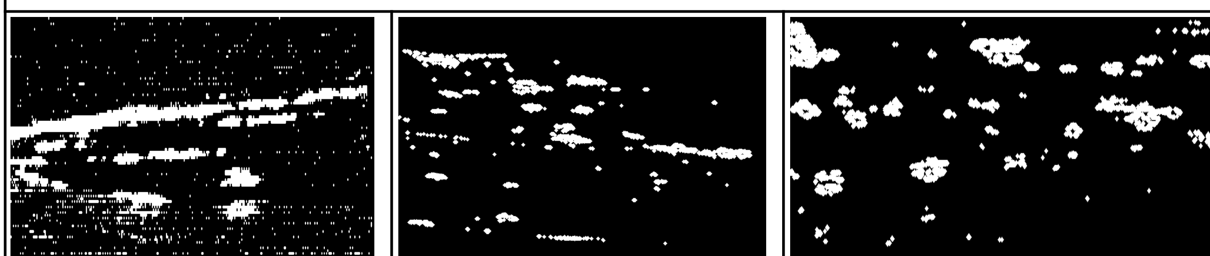
Measuring the efficiency of LIMO with the GAN framework is to estimate the precision, recall, f1 score, accuracy, sensitivity, and specificity rates of the proposed system. Figure 5.8 shows the experimental rice plant disease results conquered with the LIMO framework based GAN network using rice crop disease dataset. Figure 5.7(a) explains the various rice crop disease images obtained from the rice crop disease dataset, figure 5.7(b) shows an image pre-processing with data reduction and feature engineering contains from input values, figure 5.7(c) describes the ground truth values of input images segmented input obtained boundary values by CRCNet is represented in figure 5.7(d), and LGP extracted image is depicted in figure 5.7(e) final segmentation process.

(a)

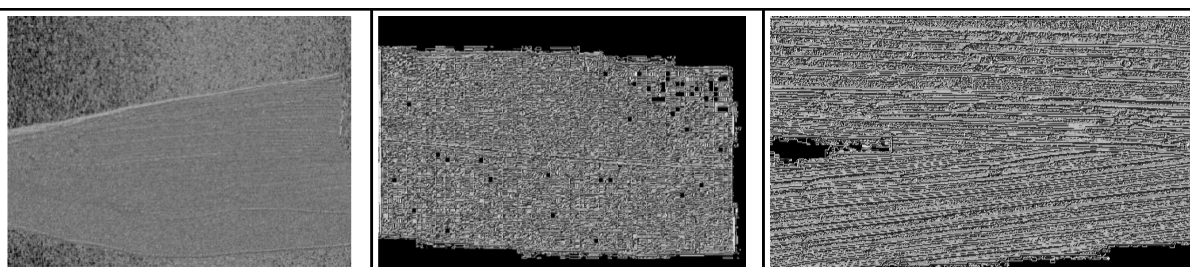


(b)

(c)



(d)



(e)

Figure 5.7 Experimental study values of LIMO combined with CRCNet using Matlab tool a) real input image b) multi-objective based pre-processing input image c) Ground Truth input image d) LIMO Segmentation values e) Extracted local gradient pattern.

5.9.3 Experimental analysis

This performance analysis focused on calculating highly accurate prediction and recall with the proposed LIMO in the GAN network to express its efficacy in standings of accuracy, recall, precision, sensitivity, and specificity metrics. This performance analysis process with diverse dimensionalities like various epoch values like 3000 to 5000 and different batch sizes range from 100 to 400.

- Experimental analysis with fluctuating epoch values.

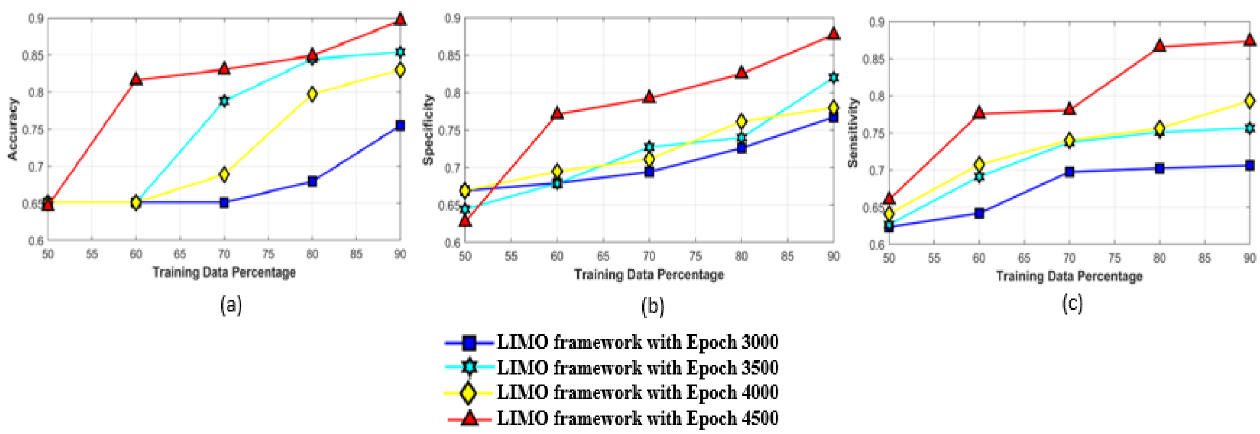


Figure 5.8 Performance metrics of LIMO framework with GAN networks, various epoch values using for a) Accuracy b) Specificity c) Sensitivity

The experimental analysis of LIMO with a generative network, to calculate the percentage accuracy rate, sensitivity values, and specificity values using various epoch values illustrated in figure 5.8. The input image-based data analysis of rice plant disease detection with the proposed framework LIMO such as IMO with GAN network using accuracy metrics is exposed in figure 5.8(a). In this analysis first test case of 80% training information, the accuracy rate calculated by LIMO performance on hit rate with epoch=3000 means the accuracy of 75.4%, epoch=3500 means 83.6% accuracy, for epoch=4000 accuracy is 85.1%, and epoch=4500 means 89% accuracy, batch sizes are randomly taken such as needs of requirements. The rice crop disease detection analysis of the proposed LIMO model using sensitivity rate is portrayed in figure 5.9(b). here, with 80% training values, the sensitivity rate calculated by the proposed framework LIMO by epoch=3000 means 75.3%, epoch value 3500 get 76.1% sensitivity, epoch=4000 means 79%, and epoch values 4500 means 86.8%. For specificity analysis with the proposed framework LIMO is represented in figure 5.9(c), Analysis performs with 80% training values, the specificity calculated with LIMO with epoch values 3000 means 77.6%, epoch rate of 3500 means got 76.2%, epoch is 4000 means 82.9% and finally epoch value is 4500 means get 88.3% specificity rate.

- Experimental measure with various batch values

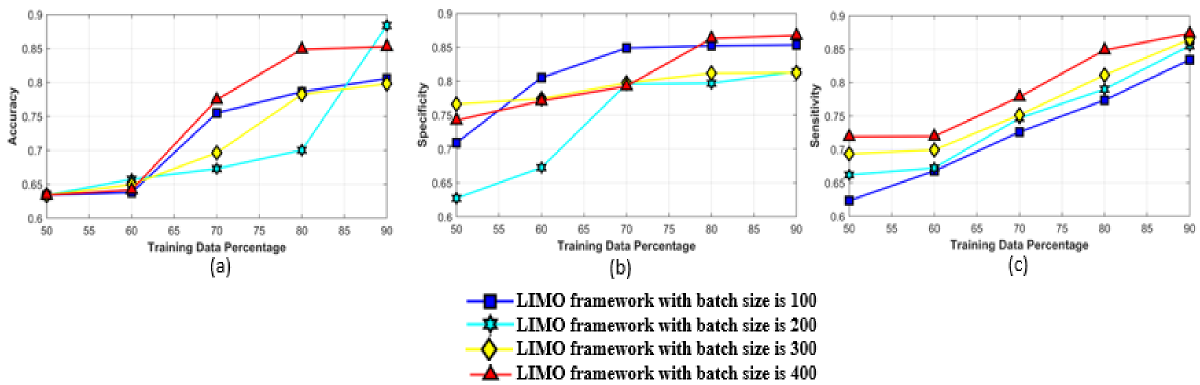


Figure 5.9 Various batch size-based performance analysis (a) Accuracy (b) Specificity (c) Sensitivity values of LIMO model

Figure 5.9 reveals the performance analysis of LIMO with GAN network with different batch sizes. For various batch sizes to calculate the LIMO model's accuracy, sensitivity, and specificity rates. For accuracy analysis, taken 80% trained to set, with this performance analysis calculated accuracy rate is obtained by the LIMO model using various batch sizes. Figure 5.10(a) shows the analysis of a batch size is 100 means 88.1%, 200 means 85.3%, a batch size of 300 means the accuracy rate is 86%, and finally, 400 batch size achieved 89.3% accuracy.

This performance analysis shows the LIMO model's sensitivity values shown in figure 5.9(b). For accuracy analysis, taken 80% trained to set, with this performance analysis calculated accuracy rate is obtained by the LIMO model using various batch sizes. At the time of batch size, 100 with our proposed LIMO framework's sensitivity rate is 85.7%, batch size 200 means 87% sensitivity, batch size 300 gives 88.1% and finally batch size 400 means our LIMO model gives 88.9%. This performance analysis shows the LIMO model's specificity rates analyzed with various batch sizes as shown in figure 5.10(c). For specificity analysis, taken 80% trained to set, with this performance analysis calculated specificity rate is obtained by the LIMO model using various batch sizes. At the time of batch size, 100 with our proposed LIMO framework's specificity rate is 86.3%, batch size 200 means 87.4% specificity, batch size 300 gives 88.8% and finally batch size 400 means our LIMO model gives 89.1%.

5.9.4 Comparative analysis

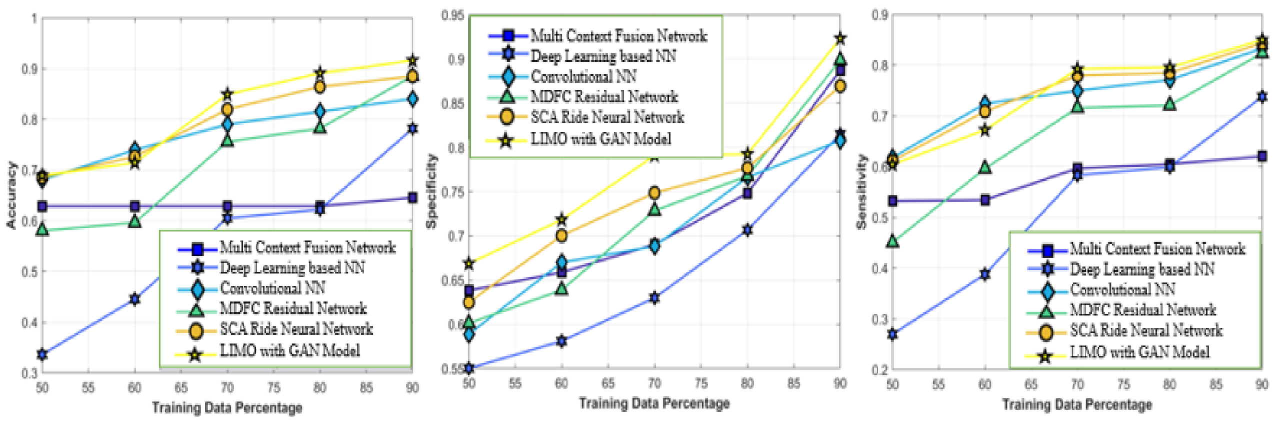
This research work modified a few methodologies to analyze the LIMO model with some existing models like multi-task fully convolutional network, DNN, CNN, MDFCRN, SCARNN, and finally our proposed LIMO with GAN framework. These existing models compared with LIMO used three metrics analyzed accuracy, sensitivity, and specificity with various levels of data training model. Additionally calculated precision, recall, and throughput values undergo many iterations for deep analysis.

- Multi-task fully convolutional network: the main objective of this MFCN model deployed with smart agriculture is especially used to identify badly affected crops quickly. The convolution model supports the extraction of unique crop features dynamically which collects generative and discriminative features from the 50k crop disease dataset. With this MFCN measures contextual values by image acquisition method which improves the quality of classification of disease crops especially measuring false positive rates through cross-validation. MFCN collects generative, discriminative, visual, unique, and contextual features to increase the quality of outcomes.
- Deep neural network: DNN model's main objective maximum number of neurons with weights processed to get high-priority feature factors from rice crops for instance creation, plant disease, the fertility rate of the field, circumstance rates, and irrigation methods.
- Convolutional neural network: This model archives comparison with a trained set and test set for rice leaf disease detection and monitoring. CNN provides various combinational outcomes and hit accurate decisions. It recommends the optimal solution to get accurate object detection.
- Multi-dimensional feature compensation with residual network: This network model is used to identify crop diseases through a compensation algorithm in the compensation layer. With the help of multi-dimensional feature collection, this model improves the detection rate of accuracy for species, fine-grained and coarse-grained crops.
- Sine cosine algorithm-ride NN: This model is normally initiated with population, position, and cluster-based image classification. The architectural design of the model collects rice leaf images with various simulation executions. SCA-RNN model provides field monitoring and automated disease identification with the help of an IoT sink node.

In image pre-processing, to improve the input data quality, the median filter was too comfortable and had proper dimensional data for further image processing. Also, feature extraction and image segmentation are executed with high-ordered feature extraction and center pixel point calculation basis. Then SCA-RNN model finalize and identify crop diseases using the image classification method.

- LIMO with GAN framework: Through this proposed research work LIMO supports identifying crop diseases and their level very effectively which helps to make advanced smart agriculture for better productivity.

Experimental analysis of training information, figure 5.10 shows the analysis results obtained with existing models such as MFCN, MDFCRN, DNN, CNN, and SCARNN models training methodology and its outcomes such as accuracy, specificity, and sensitivity compared with the LIMO data training method.

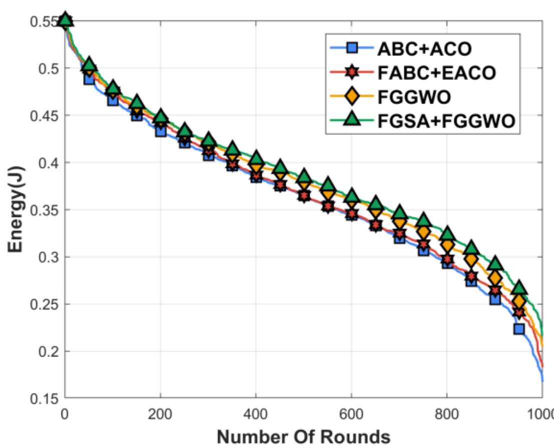


(a) (b) (c)

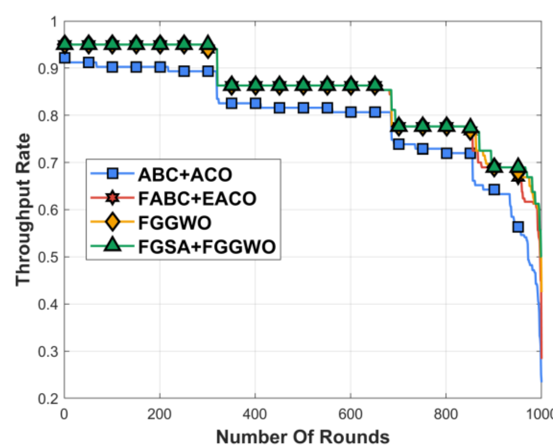
Figure 5.10 Experimental analysis of 50% to 80% training set of the detection system (a) Accuracy, (b) Sensitivity, (c) Specificity by LIMO model compared with various existing models

In this experimental analysis, figure 5.10(a) portrayed 50% to 80% of training data attained accuracy levels with the LIMO framework and other existing models. At the end of the analysis when the 50% training set is completed, the accuracy percentage is LIMO=70.5, MCFN=62.9, DNN=58, CNN=68.1, MDFCRN=59, and SCARNN=68.5%. At the time of 80% training completed, the accuracy rate is LIMO=90.6, MCFN=64.5, DNN=78.2, CNN=84%, MDFCRN=88.5, and SCARNN=88.6%. Figure 5.10(b) explains the sensitivity factors comparison of LIMO with other existing models. At the time the 50% training process was completed, the sensitivity percentage of the LIMO model was 61% also other models are MFCN was 53.2, DNN was 30, CNN was 61.5, MDFCRN was 45.2, and SCARNN was 61.1%. Comparison of 90% training data sensitivity rate was LIMO=85, MCFN=62, DNN=73.8, CNN=83.3, MDFCRN=82.4, and SCARNN=84.2%. Figure 5.11(c) shows the specificity factors evaluation of the proposed classification model with other existing models. At the time of 50% training process completed, the specificity percentage of the LIMO model has attained only 66.7% also other models are MFCN was 64, DNN was 55.1, CNN was 58.9, MDFCRN was 60.4 and SCARNN was 63.1%. Comparison of 90% training data sensitivity rate was LIMO=92.3, MCFN=88.7, DNN=81.5, CNN=81, MDFCRN=90 and SCARNN=87%.

Number of iteration and complexity rate with nodes energy and throughput, figure 5.11 describes the actual outcomes of overall throughput and communication node energy values calculated with n count of iterations. Figure 5.11(a) portrays the performance metrics of the energy consumption parameter 200 times, energy calculated by the ABC with ant colony optimization (ACO), fast artificial bee colony (FABC) plus Efficient Ant Colony Optimization (EACO), MOCLEAR without FGSA, and FGSA with MOCLEAR were 43.3, 44.3, 44.7, and 44.8 respectively.



(a)



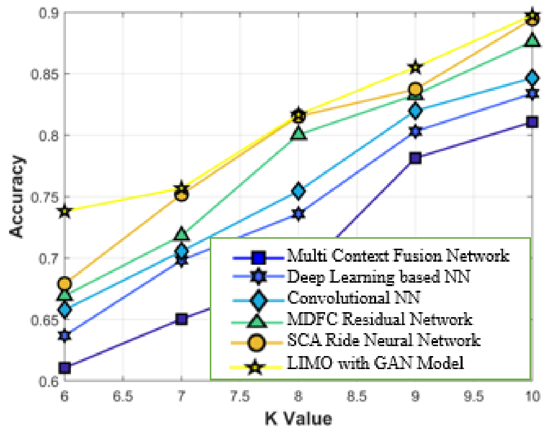
(b)

Figure 5.11 Experimental result of node energy and throughput with different iterations level.

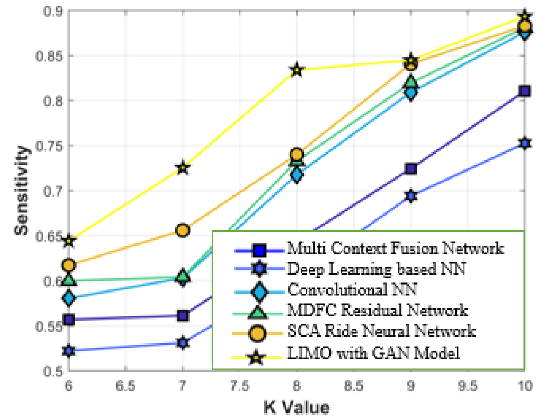
Repeated the same with 1000 iterations values computed, artificial bee colony with ant colony optimization, FABC and EACO, MOCLEAR, and FGSA with MOCLEAR were 16.7, 18.3, 20.5, and 21.6 respectively. Figure 5.11(b) portrays the performance metrics of the overall throughput parameter with 200 times, energy calculated by the ABC with ant colony optimization (ACO), fast artificial bee colony (FABC) plus efficient ant colony optimization (EACO), MOCLEAR without FGSA, and FGSA with MOCLEAR were 90.3, 93, 94.5, and 95 respectively. Repeated the same with 1000 iterations values computed, artificial bee colony with ant colony optimization, FABC and EACO, MOCLEAR, and FGSA with MOCLEAR were 23.4, 28.3, 42.5, and 50 respectively.

- **Cross-validation experimental analysis**

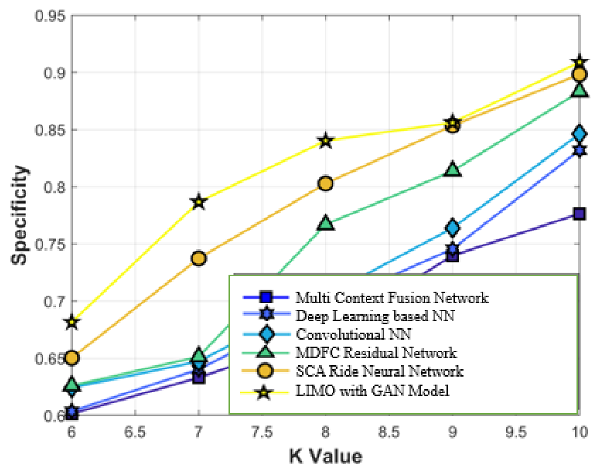
Cross-validation such as k-fold validation provides dominant wary calculated values against an overfitting sample. This cross-validation method calculates ground truth values with hyper-parameters of our proposed classifier models and accepts proper training set values. Figure 5.12 portrays experimental results obtained with cross-validation k-fold parameters as a percentage of accuracy, sensitivity values, and specificity rates. Performance analysis of accuracy rate compared with various existing classification models figure 5.12(a). Fold value k=6 means our proposed LIMO framework accuracy rate is 73.8 existing models like MFCN, DNN, CNN, MDFCRN, and SCARNN accuracy rate is 61.1, 63.7, 65.8, 66.9, and 67.9 respectively. Fold value k=10 means our proposed LIMO model accuracy rate is 89.8 existing models like MFCN, DNN, CNN, MDFCRN, and SCARNN accuracy rate is 81.1, 83.4, 84.7, 87.6, and 88.5. Figure 5.12(b) portrays the performance analysis of sensitivity values LIMO with various existing models. Fold value k=6 percentage of sensitivity values of the models by MFCN, DNN, CNN, MDFCRN, SCARNN and our proposed model LIMO with GAN network are 56.1, 53.1, 60.3, 60.4, 65.6, and 72.5 respectively. Next round with fold value k=9 percentage of sensitivity values of the models by MFCN, DNN, CNN, MDFCRN, SCARNN, and our proposed model LIMO with GAN network are 75.4, 81.4, 80.9, 87.4, 88.1, and 89.7. Experimental analysis figure 5.12(c) of specificity values LIMO with various existing models. Fold value k=6 percentage of specificity values of the existing models by MFCN, DNN, CNN, MDFCRN, SCARNN and our proposed model LIMO with GAN network are 60.1, 60.4, 62.4, 62.6, 65, and 68.2 respectively. Next round with fold value k=10 percentage of specificity values of the models by MFCN, DNN, CNN, MDFCRN, SCARNN, and our proposed model LIMO with GAN network are 77.6, 83.2, 84.6, 88.3, 89.8, and 90.9 respectively.



(a)



(b)



(c)

Figure 5.12 Cross validation's experimental analysis with (a) Accuracy k-fold values (b) Sensitivity k-fold values (c) Specificity k-fold values

5.10 Result Analysis

Performance analysis of our modified LIMO model was providing better values of accuracy, specificity, and sensitivity. Additionally, artificially based multi-object optimization archived excellent energy consumption rate and throughput. The performance is measured with training data, count of iterations, throughput calculation, and node energy consumption.

a) Machine learning data analysis

Table 5.1 our proposed LIMO model with a GAN network gets the highest accuracy rate is 91.6, whereas the accuracy rate compared with existing classification models obtained by MFCN, DNN, CNN, MDFNRN, and SCARNN were 64.5, 78.2, 84, 88.5, and 88.5. A comparison of these accuracy values reveals our proposed LIMO framework provides 29.585%

betterment than the MFCN model. LIMO with GAN network attained the highest sensitivity rate is 85%, whereas the sensitivity rate compared with existing classification and optimization models obtained by MFCN, DNN, CNN, MDFNRN, and SCARNN were 62, 73.8, 80.6, 81.2 and 83.3. Overall comparison with sensitivity rate is 27.05% better than other (MFCN) existing models. LIMO with GAN network attained the highest specificity rate is 92.3%, whereas the specificity rate compared with existing classification and optimization models obtained by MFCN, DNN, CNN, MDFNRN, and SCARNN were 88.7, 81.5, 83.6, 86.6, and 80.8. Overall comparison with specificity rate is 3.1% increased than other MFCN existing models. LIMO framework model accomplished by maximum performance analysis values compared with existing rice plant disease detection models. From this research work observation, LIMO provides increased optimization capacity with artificial intelligence in DNN. Experimental results show that LIMO with a GAN network can enlarge the classification accuracy rate and control overfitting issues. Accuracy and prediction rate got significant improvement since our research work is providing different suggestions from existing GAN alone used models. These GAN networks are used to improve image resolution values controlled with the cognitive learning method. Finally, LIMO reached maximum and excellent numbers different from existing prediction models.

Table 5.1 Performance analysis of accuracy, sensitivity, and specificity for trained data

Performance analysis percentage (%)	Multi-task fully convolutional network	Deep neural network	Convolutional neural network	Multidimensional Feature Compensation RN	Sine Cosine algorithm-RNN	***Proposed Laurent-IMO-based GAN***
Accuracy	64.5	78.2	84	88.5	88.5	**91.6**
Sensitivity	62	73.8	83.3	82.4	84.3	**85**
Specificity	88.7	81.5	80.8	89.9	86.9	**92.3**

b) Number of iterations and complexity rate with nodes energy and throughput

Table 5.2 describes the performance metrics of the node's energy and overall throughput values with n numbers of iterations. This LIMO model calculates energy and throughput graph values against various counts of iterations. Our proposed model's highest energy consumption is 21.6 was proficient by LIMO with GAN network, whereas the energy rate attained with ABC combination of ant colony optimization, FABC plus EACO, and LIMO with MOCLEAR were

16.7, 18.3, and 20.5. The highest throughput value 50 was consummated with FGSA with MOCLEAR, whereas overall throughput of ABC plus ant colony optimization, FABC combines with EACO, and MOCLEAR was 23.4, 28.3, and 42.5. Because of this LIMO provides betterment routing and deterministic path selection.

Table 5.2 An Energy and throughput analysis metrics and matrix

Metrics	ABC+ACO	FABC+EACO	MONCLER	**FGSA+MOCLEAR**
Energy	16.7	18.3	20.5	**21.6**
Throughput rate	23.4	28.3	42.5	**50**

c) Cross-validation experimental analysis

Table 5.3 portrays the performance analysis of various parameters compared with existing rice plant disease detection models using the cross-validation method. The proposed framework LIMO with GAN model attained an extreme accuracy is 89.8%, whereas accuracy rates attained by MFCN, DNN, CNN, MDFCRN, and SCARNN are 81.1, 83.4, 84.7, 87.6, and 88.5, respectively.

Table 5.3 Performance metrics of accuracy, sensitivity, and specificity comparisons

Performance analysis (percentage %)	Multi-task fully convolutional network	Deep neural network	Convolutional neural network	Multi-dimensional feature compensation RN	sine cosine algorithm-RNN	**Proposed Laurent-IMO-based GAN**
Accuracy	81.1	83.4	84.7	87.6	88.5	**91.2**
Sensitivity	81	75.3	87.6	88.1	88.3	**89.8**
Specificity	77.6	83.2	84.6	88.3	89.8	**90.9**

LIMO model attained a maximum sensitivity rate of 89.8% through the GAN network. Our proposed model with a GAN network gets a high sensitivity rate is 89.3, where the overall sensitivity rates attained by various existing systems are MFCN, DNN, CNN, MDFCRN, and SCARNNs are 81, 75.3, 87.6, 88.1, and 88.3 respectively. The extreme specificity rate of 90.9 is attained with our proposed LIMO-based GAN model, whereas the specificity rates estimated with existing models by MFCN, DNN, CNN, MDFCRN, and SCARNN are 77.6, 83.2, 84.6, 88.3, and 89.8 respectively.

CHAPTER 6
COMPARATIVE ANALYSIS

6.1 Overview

This objective introduces an advanced particle swarm optimization (APSO) by attribute subset selection algorithm based population methodology with the laurent series with intelligent multidimensional object optimization for rice plant disease detection (SIOC+LIMO) model. The main objective of this proposed SIOC+LIMO framework is used for early detection and optimal classification of different kinds of rice plant diseases. The proposed this intelligence multidimensional based object detection for rice plant disease detection technique encompasses gaussian filtering (GF) based preprocessing, multilevel thresholding Otsu with cognitive attribute selection (SIOC-CAS) based segmentation, fusion based feature extraction such as CRCNet, GAN, CAS with ResNet50, LIMO based classification. The design of CAS to describe the best possible threshold values of the Otsu-based segmentation approach helps to improve the overall classification performance.

6.2 SIOC with LIMO Framework

In this work, a new SIOC+LIMO system is derived to identify and classify different classes of RPDs. The proposed SIOC+LIMO technique includes SIOC-CAS based segmentation, GF-based preprocessing, LIMO based classification, and fusion-based feature extraction. Figure 6.1 portrays the complete working model of the proposed classifier SIOC+LIMO model.

6.3 GF based Preprocessing

In this work, the GF technique is utilized to remove the noise that exists in the plant leaf images. Several areas of study such as medicine, astronomy, geography, and so on, have commonly used the concept of DIP. This area frequently requires real‑time results, with efficacy. The performance of two-dimensional gaussian filters is extensively employed for noise removal and smoothing. It needs large numbers of computation resources and its efficacy in the application has been an inspiring study field. The convolutional operator is the gaussian filter and diverse dimensional conception of gaussian smoothing and lightening can be attained using cognitive convolutional methods.

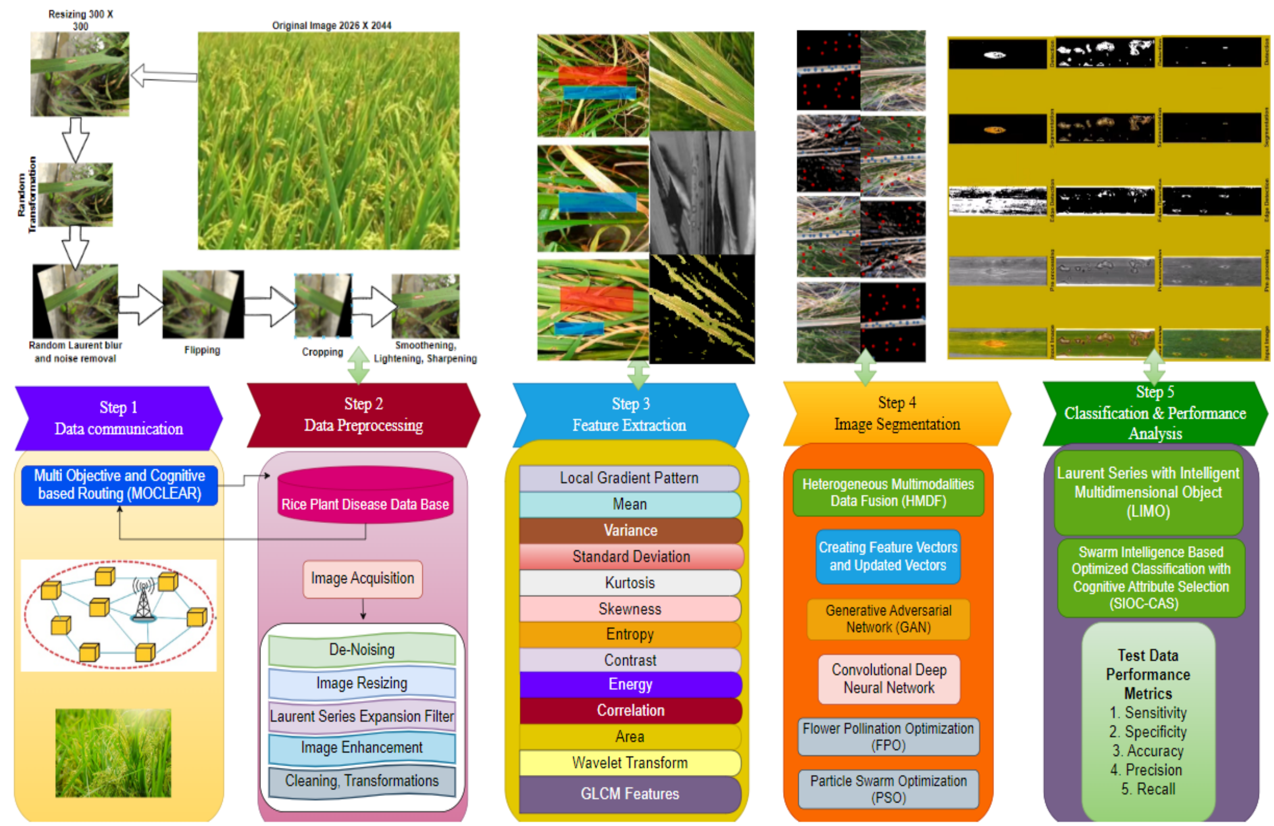


Figure 6.1 Overall process of SIOC+LIMO model

SIOC-CAS based Segmentation

During the segmentation process, the SIOC-CAS segmentation model is utilized for vital performance factors from affected portions of rice leaves in images. The best solution to the multilevel thresholding problems is Otsu's approach, an unsupervised and non-parametric model whose aim is to increase the between class variance for selecting the optimum threshold. Where k represents the number of thresholds, Otsu's model would select and optimize the $\left[t_1, t_2, ..., t\right]$ threshold. By increasing the subsequent objective function, the threshold subdivides the images to $k + 1$ classes represented as $\left[C_0, C_1, C_2, ..., C\right]$.

$$f(\left[t_1, t_2, ... , t_k\right] = \sum_{i=0}^{t_1-1} P_i, \mu_0 = \sum_{i=0}^{t_1-1} \frac{P_i}{\omega_0} \sigma_1^2 = \omega_1(\mu_1 - \mu_T)^2, \omega_1 = \sum_{i=t_1}^{t_1-1} P_i, \mu_1 = \sum_{i=t_1}^{t_2-1} \frac{i.P_i}{\omega_1} \sigma_2^2 = \omega_2(\mu_2 - \mu_T)^2. \sigma_k^2 = \omega_k(\mu_k - \mu_T)^2, \omega_k = \sum_{i=t_k}^{L-1} P_i, \mu_2 = \sum_{i=t_k}^{L-1} \frac{i.P_i}{\omega_k} \quad (6.1)$$

In equation 6.1 $f(\left[t_1, t_2, ..., t_k\right])$ denotes the objective function, $\sigma_0^2, \sigma_1^2, \sigma_2^2, ..., \sigma_k^2$ represent the variance of distinct classes of an image, $\omega_0, \omega_1, \omega_2, ..., \omega_k$ signifies the class likelihoods, and $\mu_0, \mu_1, \mu_2, ..., \mu_k$ indicates the standard level of the segmented classes $C_0, C_1, C_2, ..., C_k$

correspondingly. $\sum_{J=1}^{k} \omega_j = 1$ and $\sum_{j=1}^{k} \omega_j \mu_j = \mu_T$, μ_T represents the average intensity in an image. To determine the threshold values, the CAS is applied to it. The CAS is a metaheuristic model and follows the population initialization with the aim of the optimization method. Consider the overall amount of the population as C and accessible in the searching space of D. The population initialization is created in the dimension alongside arbitrary initiation in the search space can be given by,

$$a^i = L_j + rand\, X\, (U_j - L_j) \tag{6.2}$$

The early vector of the i_{th} CAS is represented as a^i. The lower limit and upper limit of the search space are provided as L_j and U_j in the j_{th} parameter respectively. $rand$ represents the arbitrarily created value that falls in the range of zero and one. The better capacity of the cognitive model to search in the searching space is formulated by,

$$\rho = \delta exp^{(-\alpha t/R)} \tag{6.3}$$

Let ρ represents the variable utilized in the iteration and reduce by the growing iteration. δ, $\alpha \& \beta$ is the predetermined parameter used to deal with the exploitation and exploration phases. The rotating centered coordinate utilized to update the location of CAS in the search space is expressed by,

$$a\, rand_r^i = m * ac_r^i \tag{6.4}$$

$a\, rand_r^i$ represents the rotating center coordinate of the CAS. mis utilized for denoting the rotation matrix and ac_r^i is utilized for denoting the centering coordinate in r_{th} iterations. The inertia weight of the iteration is formulated by,

$$W = \left(1 - \frac{r}{R}\right)^{\left(\lambda\sqrt{\frac{r}{R}}\right)} \tag{6.5}$$

Where W represents the weight of inertia, λ indicates the arbitrary value utilized for controlling the exploitation capability. The λ value is equivalent to one. The acceleration rate of the CAS is expressed by,

$$y = 2590 * \left(1 - exp^{-\log log\,(r)}\right) \tag{6.6}$$

whereas y is utilized to determine the acceleration of CAS. From the pseudo-code, it is noted that the CAS initializes the optimization model and the CAS location is upgraded by,

$$a_{r+1}^i = a.\, rand_r^i - a_r^{-i} \tag{6.7}$$

$$G^{i}_{r+1} = a^{-i}_{r} + \left(\left(v^{ij}_{r}\right)^2 - \left(v^{ij}_{r-1}\right)^2\right)/(2y) \qquad (6.8)$$

Where G^{i}_{r+1} represents the global optimal position of the CAS and v^{ij}_{r} denotes the novel velocity of the r_{th} CAS. When the CAS goes outside of the searching area, it would be returned to the constraint that has previously been defined. The fitness function is evaluated in every round for predicting the CAS with optimum fitness. Therefore, the fitness function is utilized for finding optimal CAS that catches the prey in the first place. This process is continued till it fulfills the total rounds iteration cycles.

6.4 Fusion-based Feature Extraction

In these sections, the fusion-based feature extraction process is carried out using three DL models like DenseNet, GAN, and Inception with ResNet50. The CNG structure outperformed the $ResNet50$, VGG 16, U-Net, $ResNet101$, and $ResNet152$ on ImageNet. The CNG method has its foundation in preceding methods such as the inception V3 and the original Inception. Followed by a similar architecture, the GAN network has its major differences in the usage of cognitive residual convolution rather than conventional residual connections (RC), and convolutions like those presented on ResNet methods. The structure can be made by thirty-six convolution layers forming the feature extraction base. The thirty six convolution layers are designed for fourteen models where the first and last one doesn't have a residual connection.

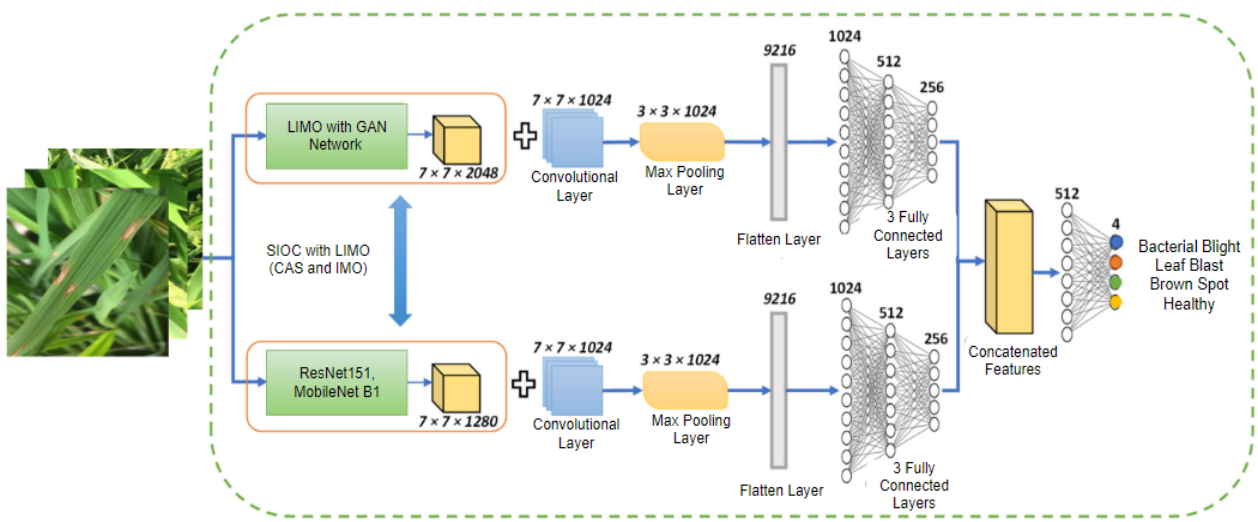


Figure 6.2 Architecture of convolutional model with GAN network

The cognitive convolutional is a spatial convolutional which independently executes on all the channels in the input, after that, pointwise convolution, a 1×1 convolutions, projecting the channel's output by the cognitive convolution to a novel channel space. Initially, the last output of CRCNet and classical convolution are similar, but it can be noted that the algorithm that can be implemented via CRCNet looks more extensive and complicated. The CRCNet implements less operation, reducing the computation costs of the network. Using CRCNet we can get a deeper algorithm which is very effective when compared to a wider one. Conversely, although the GAN with CAS method has many parameters it is very faster and more efficient when compared to the VGG 16 method. Discriminator equator models are distinct when compared to the ones discovered in the Inception methods due to the incorporation of CRCNet. Figure 6.2 demonstrated the overall architecture of the convolutional model with the GAN network. Once this method was utilized for classification on ImageNet, the last output layer comprised of thousand neurons with activation function Softmax for predicting likelihood for thousand classes. The CRCNet network has been proposed for improving the real-time efficiency of the DL method in constrained hardware conditions. This network could decrease the number of variables without sacrificing precision. Prior research has demonstrated that CRCNet requires 1/33 of a parameter of visual geometry group-16 (VGG 16) for achieving similar classification performance in imageNet 1000 classification task.

6.5 Performance Validation

Inspect the rice plant disease classification outcomes and their analysis on the SICO+LIMO framework. The outcomes are analyzed with various diverse dimensions and factors. The dataset comprises three class labels namely bacterial blight (BB), rice blast (RB), brown spot (BS), false smut (FS), and tungro (TG) besides, the augmented dataset holds 495, 487, 478, 200, 418, and 223 images under BB, RB, BS, FS, and TG.

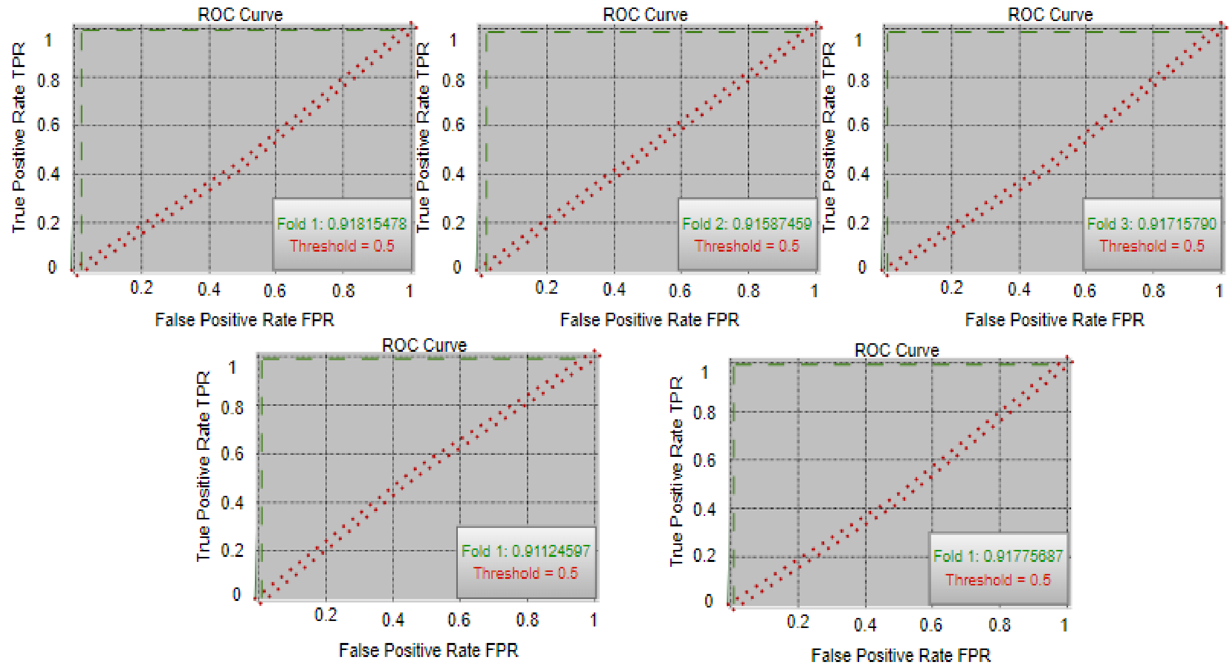


Figure 6.3 ROC analysis of SIOC+LIMO model

Table 6.1 offers a comprehensive result of the SICO+LIMO system on the applied five folds. The results established that the SIOC+LIMO algorithm has classified different types of images.

Table 6.1 Result analysis of SIOC+LIMO model with varying folds

Methods	Precision	Sensitivity	Specificity	Accuracy	F-Score	Recall	H.Loss
			Fold-1				
BB	0.9300	0.9350	0.9347	0.9348	0.9325	-	-
RB	0.9345	0.9338	0.9374	0.9330	0.9391	-	-
BS	0.9343	0.9395	0.9322	0.9313	0.9369	-	-
FS	0.9346	0.9400	0.9374	0.9383	0.9373	-	-
TG	0.9391	0.9384	0.9349	0.9396	0.9337	-	-
Average	**0.9306**	**0.9304**	**0.9348**	**0.9380**	**0.9305**	**0.9393**	**0.0604**
			Fold-2				
BB	0.9350	0.9350	0.9373	0.9365	0.9350	-	-
RB	0.9346	0.9401	0.9374	0.9383	0.9373	-	-
BS	0.94011	0.9347	0.9407	0.9383	0.9374	-	-
FS	0.9347	0.9347	0.9374	0.9365	0.9347	-	-

TG	0.946	0.9346	0.9374	0.9365	0.9346	-	-
Average	**0.9365**	**0.9366**	**0.9383**	**0.9377**	**0.9366**	**0.9365**	**0.0635**
Fold-3							
BB	0.9395	0.9379	0.9373	0.9348	0.9325	-	-
RB	0.9391	0.9384	0.9349	0.9396	0.9337	-	-
BS	0.9341	0.9395	0.9370	0.9378	0.9317	-	-
FS	0.9350	0.9350	0.9373	0.9365	0.9350	-	-
TG	0.9392	0.9346	0.9349	0.9348	0.9319	-	-
Average	**0.9386**	**0.9386**	**0.9365**	**0.9307**	**0.9321**	**0.9391**	**0.0608**
Fold-4							
BB	0.9350	0.9350	0.9373	0.9365	0.9350	-	-
RB	0.9346	0.9346	0.9374	0.9365	0.9346	-	-
BS	0.9395	0.9395	0.9348	0.9330	0.9395	-	-
FS	0.9343	0.9395	0.9322	0.9313	0.9369	-	-
TG	0.9346	0.9311	0.9374	0.9383	0.9373	-	-
Average	**0.9330**	**0.9345**	**0.9365**	**0.9394**	**0.9330**	**0.9041**	**0.0604**
Fold-5							
BB	0.9376	0.9350	0.9299	0.9383	0.9375	-	-
RB	0.9392	0.9346	0.9349	0.9348	0.9319	-	-
BS	0.9347	0.9347	0.9374	0.9365	0.9347	-	-
FS	0.9350	0.9350	0.9373	0.9365	0.9350	-	-
TG	0.9392	0.9346	0.9349	0.9348	0.9319	-	-
Average	**0.9347**	**0.9348**	**0.9374**	**0.9365**	**0.9347**	**0.9322**	**0.0678**

For example, cross-validation fold-1, SIOC+LIMO framework has classified BB, RB, BS, FS, and TG with the accuracy of 0.9348, 0.9330, 0.9313, 0.9383, and 0.9396 respectively. Followed by fold-3, the SIOC+LIMO approach has classified the BB, RB, BS, FS, and TG with the accuracy of 0.9348, 0.9396, 0.9378, 0.9365, and 0.9348 correspondingly. Likewise, with fold-5, the SIOC+LIMO system has classified the BB, RB, BS, FS, and TG with the accuracy of 0.9383, 0.9348, 0.9365, 0.9365, and 0.9348 respectively. Table 6.2 and Figures 6.4 to 6.9 depicts the analysis of the result on the combination of the SIOC with SVM and GAN with LIMO techniques under five folds. For example, with fold-1, the SIOC+LIMO approach has led to a better outcome with a precision probability is 0.9396, sensitivity value of 0.9394, specificity is 0.9348, accuracy value is 0.9330, F1 measure is 0.9395, recall value is 0.9393, and hamming loss rate is 0.0604. In addition, with fold-2, the SIOC+LIMO approach has reached optimum performance with the precision=0.9365, sensitivity=0.9366, specificity=0.9383, accuracy=0.9377, F1 measure=0.9366, recall=0.9365, and hamming loss rate=0.0635. Followed with fold-3, the SIOC+LIMO manner has resulted in an improved outcome with precision is 0.9386, sensitivity is 0.9386, specificity is 0.9331, accuracy is 0.9307, F1 measure is 0.9336, recall is 0.9391, and hamming loss rate is 0.0608. Also, with fold-4, the SIOC+LIMO method has attained optimum efficiency with the precision=0.9330, sensitivity=0.9345, specificity=0.9365, accuracy=0.9354, F1-score=0.9330, recall=0.9396, and hamming loss rate=0.0604. Moreover, with fold-5, the SIOC+LIMO technique has obtained optimal performance with a precision value is 0.9347, sensitivity value is 0.9348, specificity value is 0.9374, accuracy value is 0.9365, F1 measure rate is 0.9347, recall of 0.9322, and hamming loss rate is 0.0678.

Table 6.2 Result analysis of SIOC+LIMO model with various measures

No.of Fold	Precision	Sensitivity	Specificity	Accuracy	F-Score	Recall	Hamming
Fold-1	0.9306	0.9304	0.9348	0.9380	0.9305	0.9393	0.0604
Fold-2	0.9365	0.9366	0.9383	0.9377	0.9366	0.9365	0.0635
Fold-3	0.9386	0.9386	0.9365	0.9307	0.9321	0.9391	0.0608
Fold-4	0.9330	0.9345	0.9365	0.9394	0.9330	0.9041	0.0604
Fold-5	0.9347	0.9348	0.9374	0.9365	0.9347	0.9322	0.0678
Average	**0.9333**	**0.9334**	**0.9367**	**0.9356**	**0.9334**	**0.9302**	**0.0625**

Furthermore, figure 6.5 inspects the results of the SIOC+LIMO model with recent methods based on sensitivity. The outcomes showcased that MFCN, SCARNN, and MCFN ResNet models have gained lower sensitivity of 68.02%, 63.30%, and 65% respectively.

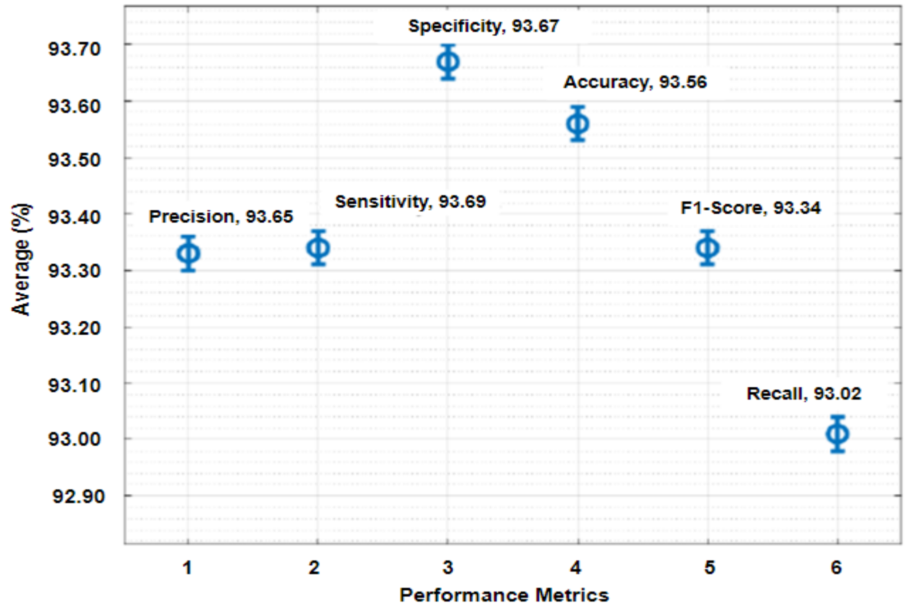


Figure 6.4 Average analysis of SIOC+LIMO model with varying measures

In line with this, the MDFN ResNet model has resulted in a slightly increased sensitivity of 73.46%. With that, the DCNN, SIOC+SVM, and GAN+LIMO techniques have accomplished moderate sensitivity of 83.70%, 86.66%, and 90.00% correspondingly. Moreover, the proposed GAN+LIMO technique has resulted in a near-optimal sensitivity. However, the proposed SIOC+LIMO technique has increased sensitivity by 99.34%.

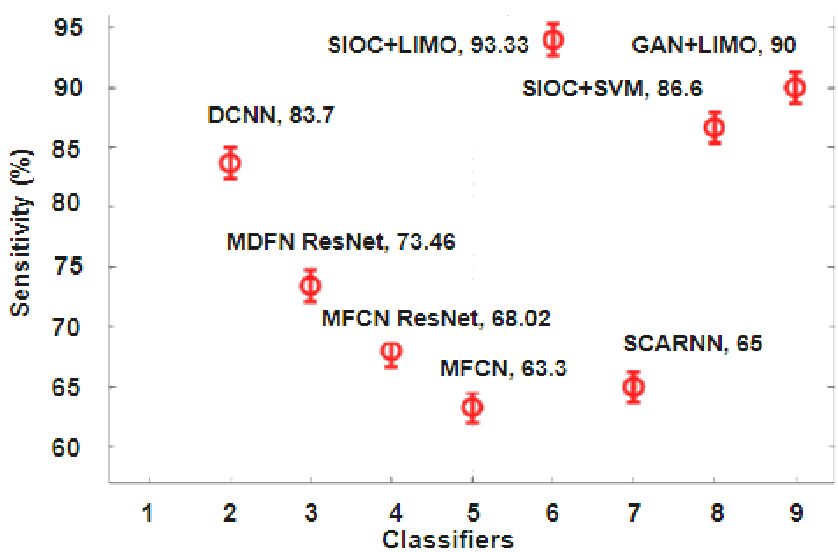


Figure 6.5 Sensitivity analysis of SIOC+LIMO technique with existing approaches

Figure 6.6 examines the specificity of the SIOC+LIMO method with recent approaches. The outcomes demonstrated that the MFCN, MFCN ResNet, and SCARNN models have reached minimum specificity of 88%, 90.4%, and 89.6% correspondingly. Also, the MDFN ResNet manner has occasioned a somewhat enhanced specificity of 91.9%. Also, the GAN+LIMO framework has accomplished moderate specificity values of 92.79%. The SIOC+SVM model has resulted in a near-optimum specificity of 92.80%. At last, the projected SIOC+LIMO methodology has a maximum specificity of 93.97%.

Figure 6.7 scrutinizes the precision of the SIOC+LIMO algorithm with existing manners. The results showed that the MFCN, MFCN ResNet, and SCARNN models have gained lower precision of 87.46%, 86.5%, and 85.2% correspondingly. Along with that, the DNN model has resulted in a somewhat enhanced precision of 74.91%. Followed by, the SIFT-SVM, MDFN ResNet, and MFCN techniques have accomplished moderate precision of 91.3%, 89.3%, and 88% correspondingly. Moreover, the GAN+LIMO methods have resulted in a near-optimal precision of 92.8%. Eventually, the presented SIOC+LIMO technique has increased precision to 93.80%.

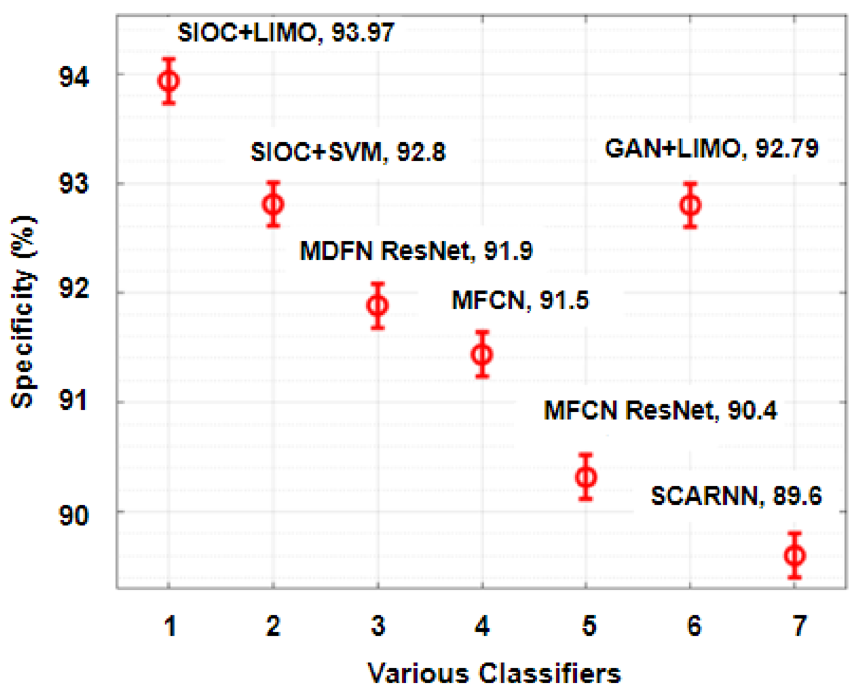


Figure 6.6 Specificity analysis values of SIOC+LIMO framework compared with existing classification algorithms

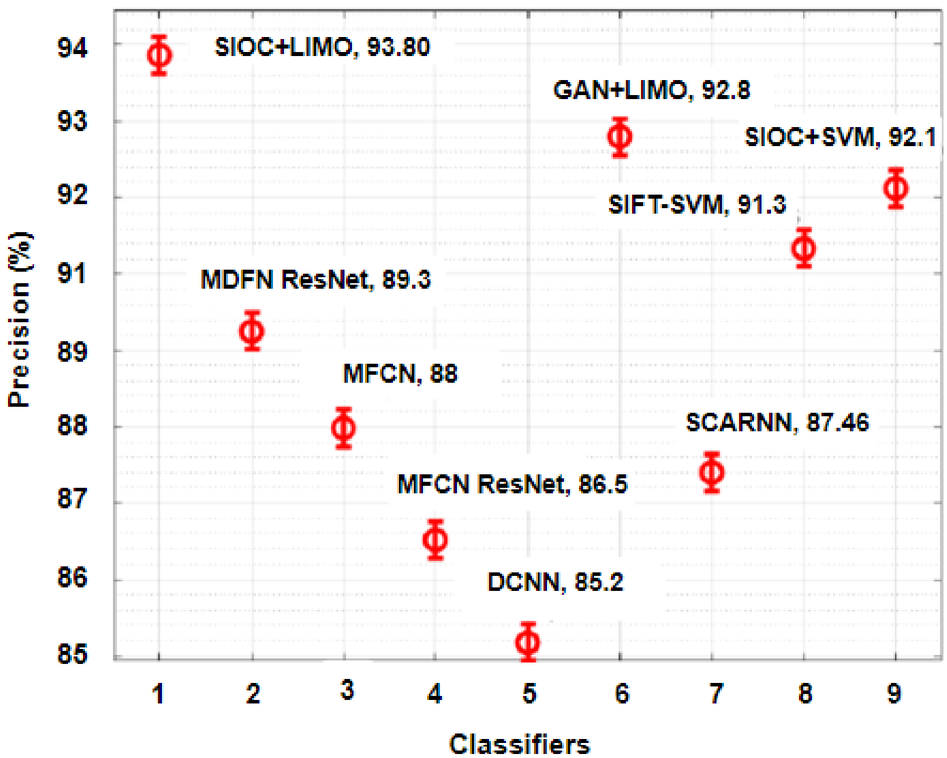


Figure 6.7 Precision values of SIOC+LIMO framework with various classifiers

Figure 6.8 inspects the accuracy of the SIOC+LIMO framework with various recent classification models. The outcomes showcased that the SVM, ANN, and KNN models have gained lower accuracy of 91.2%, 90%, and 88% respectively. In line with this, the MFCN, XGBooster, SIFT-SVM, and Ensemble model has resulted in a slightly superior accuracy of 92.7%, 92.6%, 92.89%, 92.2%, and 92%. Also, the SIOC+SVM, MFCN ResNet, DCNN, and MDFN ResNet techniques have accomplished moderate accuracy of 93.2%, 92.8%, 92.79%, and 92.6% correspondingly. Moreover, the GAN+LIMO accuracy of 93.6%. However, the presented SIOC+LIMO approach has a maximum accuracy of 93.97%.

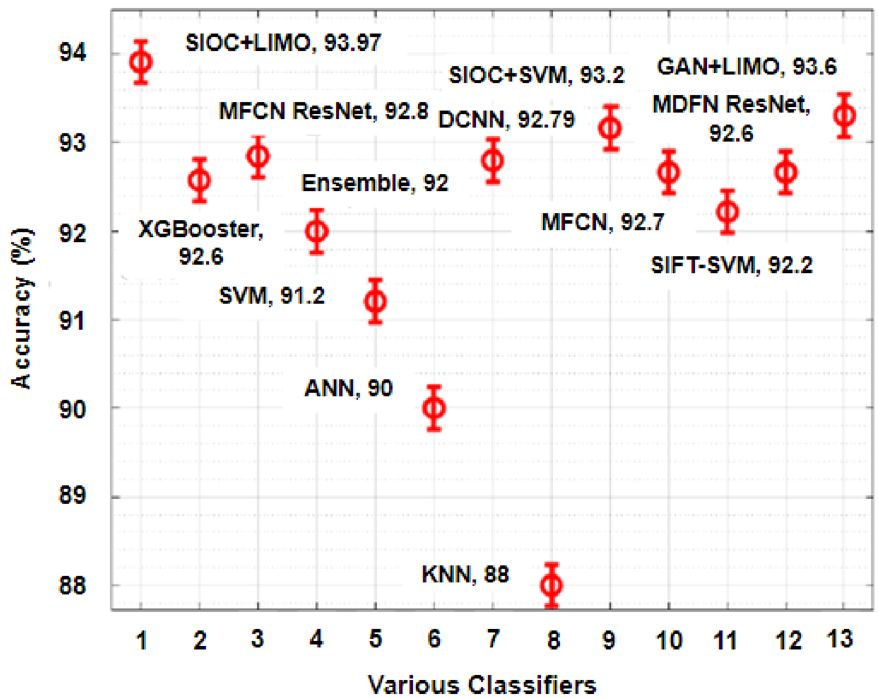


Figure 6.8 Accuracy of SIOC+LIMO framework with various disease detection classifiers

Figure 6.9 studies the F1-measure of the SIOC+LIMO technique with recent methodology. The outcomes portrayed that SVM, DCNN, and MFCN algorithms have achieved minimum F1 scores of 89.4%, 87.6%, and 87% correspondingly. Likewise, the DNN model has resulted in a somewhat increased F1 score of 81.51%. Eventually, the SIOC_SVM, MFCN ResNet, SIFT-SVM, and MDFN ResNettechniques have accomplished moderate F1 scores of 92%, 91.8%, 91.4%, and 90.3% correspondingly. Furthermore, the GAN+LIMO technique has resulted in a near-optimum F1 score of 92.8%. But, the proposed SICO+LIMO methodology has improved the F1 score by 93.8%.

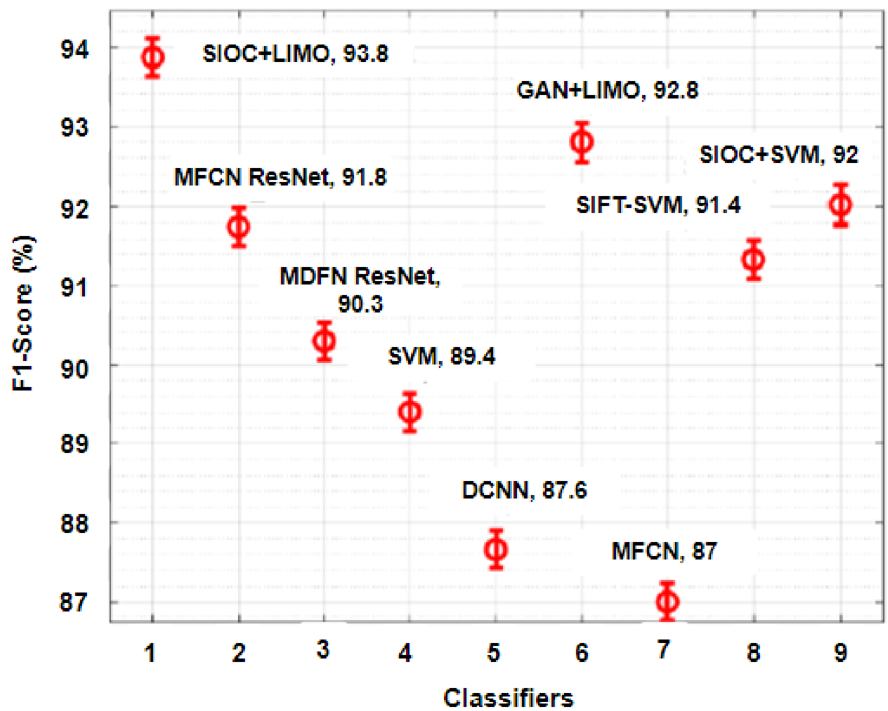


Figure 6.9 F1-score measure of SIOC+LIMO model with various machine learning algorithms

Table 6.3 Various detection and classification model metrics compared with SIOC+LIMO

Performance metrics	**Sensitivity value (%)**	**Specificity value (%)**	**Precision rate (%)**	**Accuracy level (%)**	**F1-score measure (%)**	**Computation time(s)**
SIOC+SVM	86.6	92.80	92.1	93.2	92	11
GAN+LIMO	90	92.79	92.8	93.6	92.8	10
Proposed SIOC+LIMO	**93.3**	**93.97**	**93.8**	**93.97**	**93.8**	**8**

Table 6.3 demonstrates the detailed comparative result analysis of the three proposed models for rice plant disease classification. The practical performance analysis indicated that the proposed SIOC+LIMO framework has outdone the other two methods with maximum accuracy of 93.97%. It is also noticed that the GAN+LIMO and SIOC+SVM models have shown reasonable computation times of 11s and 10s respectively. However, the LIMO model has gained a lower computation time of 8s. Therefore, it is found that the SIOC+LIMO technique has outperformed the other two proposed models.

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENTS

7.1 Conclusion

The performance of the attribute selection algorithms with One all random population initialization is evaluated in terms of accuracy gain, dimensionality reduction, and computational time. PSO with Greedy leap and one to all random initialization achieve an accuracy gain in the range of 1.76% to 40.05% and a dimensionality reduction in the range of 62.31% to 90% and out performs both PSO with Greedy reset and FPO with the Greedy crossover. Also, PSO with Greedy leap requires less computational time compared to PSO with Greedy reset and FPO with Greedy crossover. All three proposed attribute selection algorithms with One to all random initialization are compared with other randomized metaheuristic algorithms employed as search strategies in the process of attribute selection. SIOC-CAS algorithm provides the best outcomes with intelligent feature selections in terms of both the objectives compared with existing methodologies for features selection in most of the instances and in case of poor performance, the difference is marginal. The predictive performance of proposed algorithms is validated with six classifiers and the results find that the proposed wrappers are general to different classification algorithms. All the proposed attribute selection methodologies with One to all random population initialization are applied as a pre-processing technique to classification in the process of solving agriculture-based pattern recognition problems. Rice leaf disease identification is solved using DIP and ML techniques. It is observed that on applying the proposed pre-processing technique, an accuracy gains of 9% and dimensionality reduction of 93.71% is achieved with a computational time of 2357s over random population initialization that takes 4025s for computation such as PSO with Greedy leap.

This research work formulated a new framework LIMO with GAN network for rice crop disease identification and prevention in the IoT platform. Through this framework, the IoT network nodes are collecting values from the rice crop leaves through a sensor, although the sensed values were communicated with a base station through MOCLEAR. This MOCLEAR method selects the optimal route between transmission nodes in which nodes participated in communication. Therefore, an advanced optimal route such as path selection with an intelligent decision-making algorithm for the betterment of IoT network communication. With the sensor values, our proposed laurent with IMO algorithm performs an effective classification for identifying crop disease and its levels. Through this research framework, modifications to pre-processing with data reduction and feature engineering to improve the quality of object detection. Then MOCLEAR method for decision-making betterment for path selection. Cognitive residual convolutional network was revised to segment the sensed values with appropriate feature selections to improve the efficiency of crop disease identification. LIMO classification was implemented with the LIMO Optimization. This AI-based optimization framework combined with the GAN network revealed enhanced performance metrics with better sensitivity, specificity, and accuracy values of 89.8%, 90.9%, and 91.2% respectively.

In this research work, a fine-tuning hybrid model termed SIOC+LIMO was evaluated and compared with various rice plant disease detection classifiers. SIOC+LIMO framework, the cognitive tuning of parameters such as activation function and loss function was done. This shows improvement over the results of the second stage analysis in terms of accuracy and trainable parameters. The accuracy, sensitivity, specificity, precision, recall, and F1-Score for SIOC+LIMO in our work for data fusion and multi-modality classification. Also, their performance is evaluated based on the performance parameter. This SIOC+LIMO model was providing results with the best filtering methodologies, attribute selection model (CAS), various combinations of optimization techniques such as particle swarm optimization, flower pollination algorithm, intelligence multimodality object optimization, adaptive network availability such as ResNet50, CRCNet, GAN, U-Net, and finally various dimensional image input. Also additionally introduced HMDF for data fusion and the MOCLEAR algorithm used for data communication. Above mentioned features improve the quality of the RPD system, very quick detection, and optimal classification. For SIOC+LIMO accuracy is 93.97%, sensitivity value is 93.3%, specificity value is 93.97%, the precision rate is 93.8%, recall is 93.9%, and F1 score is 93.8%.

7.2 Scope of the Future Enhancement

Research on attribute selection is vast. Much research can be carried out to explore the possible ways of handling the problem of the dimensionality curse and research on attribute selection is endless. To cite a few in biological datasets, many attributes can be a few hundred or more than thousands belonging to the category of large-scale problems. This research work focuses on a classification problem with a maximum of 166 attributes. Since attribute selection is a necessary step before classification, in the future, it is interesting to examine the capability of proposed attribute selection algorithms on large-scale problems. Attribute selection demands a discrete problem space such as count of attributes. Binary encoding is employed to represent the solution vectors in problem space. The remedy vector is D dimensional and the problem becomes complex if the attribute space is large. Hence, an encoding scheme needs to be designed for large-scale problems. All the benchmark datasets taken from the UCI repository for validating the proposed attribute selection algorithms represent a two-class problem. It is necessary to extend the application of a proposed algorithm to a multi-class problem. The population initialization technique is independent of the dimensionality of attribute space. However, restriction on population size makes it less attractive in large-scale problems owing to computational complexity. Hence, population initialization, to address large-scale problems needs to be devised. In the future, to go with various deep learning algorithms which can be suitable for unstructured more noisy images to improve the exact rice crop disease prevention and detection. Moreover, crop disease datasets for learning can be received from plenty of different sources, dimensional captured values, a range of geographical boundaries, and different environments.

REFERENCES

[1] M. H. Saleem, J. Potgieter and K. M. Arif, (2022). A Performance-Optimized Deep Learning-Based Plant Disease Detection Approach for Horticultural Crops of New Zealand, *IEEE Access*, vol. 10, pp. 89798-89822. doi: 10.1109/ACCESS.2022.3201104.

[2] S. K. Noon, M. Amjad, M. A. Qureshi and A. Mannan, (2022). Handling Severity Levels of Multiple Co-Occurring Cotton Plant Diseases Using Improved YOLOX Model, *IEEE Access*, vol. 10, pp. 134811-134825. doi: 10.1109/ACCESS.2022.3232751.

[3] G. Delnevo, R. Girau, C. Ceccarini and C. Prandi, (2022). A Deep Learning and Social IoT Approach for Plants Disease Prediction Toward a Sustainable Agriculture, *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7243-7250. doi: 10.1109/JIOT.2021.3097379.

[4] L. Li, S. Zhang and B. Wang, (2021). Plant Disease Detection and Classification by Deep Learning—A Review, *IEEE Access*, vol. 9, pp. 56683-56698. doi: 10.1109/ACCESS.2021.3069646.

[5] Novikov D.A. (2016). Systems Theory and Systems Analysis. Systems Engineering. In: *Cybernetics. Studies in Systems, Decision and Control*, vol 47. Springer, Cham. https://doi.org/10.1007/978-3-319-27397-6_4

[6] Andres-F Jimeneza, Pedro-F Cardenas, Antonio Canales, Fabian Jimenez, Alfonso Portacio, (2020). A survey on intelligent agents and multi-agents for irrigation scheduling, *Computers and Electronics in Agriculture*, vol. 176, 105474. https://doi.org/10.1016/j.compag.2020.105474

[7] Russell, Stuart J.; Norvig, Peter (2003). Artificial Intelligence: A Modern Approach (2nd ed.). *Upper Saddle River, New Jersey*: Prentice Hall. ISBN 0-13-790395-2.

[8] Mahmoud Abbasi, Marta Plaza-Hernández, Javier Prieto, Juan M. Corchado, (2022). Security in the Internet of Things Application Layer: Requirements, Threats, and Solutions, *IEEE Access*, Vol 10, pp 97197 – 97216.

[9] Andres-F Jimenezb, Pedro-F Cardenas, Fabian Jimenez, Antonio Ruiz-Canales, Angel López, (2020). A cyber-physical intelligent agent for irrigation scheduling in horticultural crops, *Computers and Electronics in Agriculture*, vol. 178, 105777. https://doi.org/10.1016/j.compag.2020.105777

[10] India Brand Equity Foundation (IBEF), (2020). Agriculture in India: Information about Indian Agriculture & Its Importance, IBEF-Department of Commerce, Ministry of Commerce and Industry, Government of India. Online Source: https://www.ibef.org/industry/agriculture-india.aspx

[11] Andrew, M., (2020). Smart Farming in 2020: How IoT sensors are creating a more efficient precision agriculture industry. https://www.businessinsider.com/smart-farming-iot-agriculture?IR=T (accessed 8.18.20).

[12] Architectural documentaries, (2016). World Population 1800-2050, http://archidocu.com/wp-content/uploads/2016/12/World-Population-1800-2050.jpg

[13] Jeff Desjardins, (2017). This animation compares the population growth of India and china, World Economic Forum, 99-93 route de la capite, ch-1223. https://www.weforum.org/agenda/2017/08/this-animation-compares-the-population-growth-of-india-and-china.

[14] Singh, V., Misra, A.K., (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Process Agriculture* 4, 41–49. https://doi.org/https://doi.org/10.1016/j.inpa.2016.10.005

[15] Ramesh, S., Rajaram, B., (2018). IoT based crop disease identification system using optimization techniques. *ARPN J. Eng. Appl. Sci*. 13, 1392–1395.

[16] Sharma, R., Kamble, S.S., Gunasekaran, A., Kumar, V., Kumar, A., (2020). A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Comput. Oper. Res.* 119, 104926. https://doi.org/https://doi.org/10.1016/j.cor.2020.104926

[17] Thorat, A., Kumari, S., Valakunde, N.D., (2017). An IoT based smart solution for leaf disease detection, in: *2017 International Conference on Big Data, IoT and Data Science* (BID). pp. 193–198. https://doi.org/10.1109/BID.2017.8336597

[18] Wahabzada, M., Mahlein, A.-K., Bauckhage, C., Steiner, U., Oerke, E.-C., Kersting, K., (2016). Plant Phenotyping using Probabilistic Topic Models: Uncovering the Hyperspectral Language of Plants. *Sci. Rep*. 6, 22482. https://doi.org/10.1038/srep22482

[19] Khattab, A., Habib, S., Ismail, H., Zayan, S., Fahmy, Y., Khairy, M., (2019). An IoT-based cognitive monitoring system for early plant disease forecast. *Comput. Electron. Agric*. 166, 105028. https://doi.org/10.1016/j.compag.2019.105028

[20] Zhao, Y., Liu, L., Xie, C., Wang, R., Wang, F., Bu, Y., Zhang, S., (2020). An effective automatic system deployed in agricultural Internet of Things using Multi-Context Fusion Network towards crop disease recognition in the wild. *Appl. Soft Comput*. 89, 106128. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106128

[21] Mohammed, H.R., (2020). Mapping Paddy Rice Fields Using Landsat and Sentinel Radar Images in Urban Areas for Agriculture Planning. *J. Plan. Dev*. 41, 1–28.

[22] Shende, V., Mahendra, P., (2019). IoT BASED PREVENTIVE CROP DISEASE MODEL USING IP AND CNN. *Int. Res. J. Eng. Technol*. 6, 1270–1272.

[23] Alonso, R., Sittón-Candanedo, I., García, Ó., Prieto, J., Rodríguez, S., (2019). An intelligent Edge-IoT platform for monitoring livestock and crops in a dairy farming scenario. *Ad Hoc Networks* 98, 102047. https://doi.org/10.1016/j.adhoc.2019.102047

[24] Wolfert, S., Goense, D., Sørensen, C.A.G., (2014). A Future Internet Collaboration Platform for Safe and Healthy Food from Farm to Fork, in: 2014 *Annual SRII Global Conference*. pp. 266–273. https://doi.org/10.1109/SRII.2014.47

[25] Pujari, J.D., Yakkundimath, R., Byadgi, A.S., (2015). Image Processing Based Detection of Fungal Diseases in Plants. *Procedia Comput. Sci*. 46, 1802–1808. https://doi.org/https://doi.org/10.1016/j.procs.2015.02.137

[26] Khatua, P.K., Ramachandaramurthy, V.K., Kasinathan, P., Yong, J.Y., Pasupuleti, J., Rajagopalan, A., (2020). Application and assessment of internet of things toward the sustainability of energy systems: Challenges and issues. *Sustain. Cities Soc*. 53, 101957. https://doi.org/https://doi.org/10.1016/j.scs.2019.101957

[27] Gupta, A.K., Gupta, K., Jadhav, J., Deolekar, R. V, Nerurkar, A., Deshpande, S., (2019). Plant Disease Prediction using Deep Learning and IoT, in: 2019 *6th International Conference on Computing for Sustainable Global Development* (INDIACom). pp. 902–907.

[28] Atole, R.R., Park, D., (2018). A Multiclass Deep Convolutional Neural Network Classifier for Detection of Common Rice Plant Anomalies. *Int. J. Adv. Comput. Sci. Appl*. 9, 67–70.

[29] Li, D., Wang, R., Xie, C., Liu, L., Zhang, J., Li, R., Wang, F., Zhou, M., Liu, W., (2020). A Recognition Method for Rice Plant Diseases and Pests Video Detection Based on Deep Convolutional Neural Network. *Sensors* 20, 578. https://doi.org/10.3390/s20030578

[30] Abu Bakar, M.N., Abdullah, A.H., Abdul Rahim, N., Yazid, H., Misman, S.N., Masnan, M.J., (2018). Rice leaf blast disease detection using multi-level colour image thresholding. *J. Telecommun. Electron. Comput. Eng*. 10, 1–6.

[31] Chawal, B., Panday, S., (2019). Rice Plant Disease Detection using Twin Support Vector Machine (TSVM). *J. Sci. Eng*. 7. https://doi.org/10.3126/jsce.v7i0.26794

[32] Dang, K.B., Burkhard, B., Windhorst, W., Müller, F., (2019). Application of a hybrid neural-fuzzy inference system for mapping crop suitability areas and predicting rice yields. *Environmental Model Software*. 114, 166–180. https://doi.org/https://doi.org/10.1016/j.envsoft.2019.01.015

[33] Zheng, Y.-J., (2015). Water wave optimization: A new nature-inspired metaheuristic. *Comput. Oper. Res*. 55, 1–11. https://doi.org/https://doi.org/10.1016/j.cor.2014.10.008

[34] Dhumane, A. V, Prasad, R.S., (2018). Fractional Gravitational Grey Wolf Optimization to Multi-Path Data Transmission in IoT. *Wirel. Pers. Commun*. 102, 411–436. https://doi.org/10.1007/s11277-018-5850-y

[35] Jiang, F., Gong, M., Zhan, T., Fan, X., (2020). A Semi supervised GAN-Based Multiple Change Detection Framework in Multi-Spectral Images. *IEEE Geosci. Remote Sens*. 17, 1223–1227. https://doi.org/10.1109/LGRS.2019.2941318

[36] Mishra, M., Choudhury, P and Pati, B., (2020). Modified ride-NN optimizer for the IoT based plant disease detection. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/https://doi.org/10.1007/s12652-020-02051-6

[37] Paul Shekonya Kanda, Kewen Xia, Olanrewaju Hazzan Sanusi, (2021). A Deep Learning-Based Recognition Technique for Plant Leaf Classification, IEEE Access, vol 9, pp 162590 – 162613.

[38] Zhong, L., Hu, L., Zhou, H., (2019). Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* 221, 430–443. https://doi.org/https://doi.org/10.1016/j.rse.2018.11.032

[39] Richey, B., Majumder, S., Shirvaikar, M., Kehtarnavaz, N., (2020). Real-time detection of maize crop disease via a deep learning-based smartphone app, in: *Proc.SPIE*. https://doi.org/10.1117/12.2557317

[40] Rodríguez, L., Castillo, O., García, M. (2020). A new randomness approach based on sine waves to improve performance in metaheuristic algorithms. *Soft Comput*, 24, 11989–12011. https://doi.org/10.1007/s00500-019-04641-9

[41] Rodríguez, L., Castillo, O., Soria, J., Melin, P., Valdez, F., Gonzalez, C.I., Martinez, G.E and Soto, J., (2017). A fuzzy hierarchical operator in the grey wolf optimizer algorithm, *Applied Soft Computing*, 57: 315-328. https://doi.org/https://doi.org/10.1016/j.asoc.2017.03.048

[42] Rodríguez, L., Castillo, O., García, M and Soria, J., (2019). Constrained Real-Parameter Optimization Using the Firefly Algorithm and the Grey Wolf Optimizer. *Studies in Computational Intelligence, Springer,* Cham, 827. 155-167. https://doi.org/https://doi.org/10.1007/978-3-030-34135-0_11

[43] Sanchez, D.F., Melin, P and Castillo, O., (2017). A Grey Wolf Optimizer for Modular Granular Neural Networks for Human Recognition. Computational Intelligence and Neuroscience, 8: 1-26. https://doi.org/https://doi.org/10.1155/2017/4180510

[44] Badrinarayanan, V., Kendall, A., Cipolla, R., (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

[45] Largeron, C., Moulin, C., Géry, M., (2011). Entropy Based Feature Selection for Text Categorization, *in: Proceedings of the 2011 ACM Symposium on Applied Computing*, SAC11. Association for Computing Machinery, New York, NY, USA, pp. 924–928. https://doi.org/10.1145/1982185.1982389

[46] Jun, B., Choi, I., Kim, D., (2013). Local Transform Features and Hybridization for Accurate Face and Human Detection. *IEEE Trans. Pattern Anal. Mach. Intell*. 35, 1423–1436.

[47] Gao, Y., Kong, B., Mosalam, K.M., (2019). Deep leaf-bootstrapping generative adversarial network for structural image data augmentation. *Comput. Civ. Infrastruct. Eng*. 34, 755–773. https://doi.org/10.1111/mice.12458

[48] AlameluMangai, S., Sankar, B.R., Alagarsamy, K., (2014). Taylor Series Prediction of Time Series Data with Error Propagated by Artificial Neural Network. *Int. J. Comput. Appl*. 89. https://doi.org/10.5120/15470-4112

[49] Prajapati, H., Shah, J., Dabhi, V., (2017). Rice Leaf Diseases Data Set. *Intell. Decis. Technol.* 11, 357–373. https://doi.org/10.3233/IDT-170301

[50] Hu, W-J., Fan, J., Du, Y-X., Li, B-S., Xiong, N and Bekkering, E., (2020). MDFC–ResNet: An Agricultural IoT System to Accurately Recognize Crop Diseases. *IEEE Access*, 8:115287 – 115298. https://doi.org/https://doi.org/10.1109/ACCESS.2020.3001237

[51] Rutuja R. Patil, Sumit Kumar, (2022). Rice-Fusion: A Multimodality Data Fusion Framework for Rice Disease, *IEEE Access*, vol 10, issue 2169-3536, pp 5207 – 5222, DOI: 10.1109/ACCESS.2022.3140815.

[52] Xue, B, Zhang, M & Browne, (2014), Particle swarm optimization for feature selection in classification: Novel initialization and updating mechanisms, *Applied Soft Computing*, vol. 18, pp. 261-276.

[53] Junde Chen, Weirong Chen, Adan Zeb, Shuangyuan Yang, Defu Zhang, (2022). Lightweight Inception Networks for the Recognition and Detection of Rice Plant Diseases, *IEEE Sensor Journals*. Vol 22, issue 14, pp 14628 – 14638, DOI: 10.1109/JSEN.2022.3182304.

[54] Zawabaa, HM, Hassanien, AE, Emary, E, Yamany, W & Parv, B (2015). Hybrid flower pollination algorithm with rough sets for feature selection, *Proceedings of International Conference on Computer Engineering*, pp. 278-283.

[55] Wen-Liang Chen, Yi-Bing Lin, Fung-Ling Ng, Chun-You Liu, Yun-Wei Lin, (2020). RiceTalk: Rice Blast Detection Using Internet of Things and Artificial Intelligence Technologies, *IEEE Internet of Things Journal*, vol 7, issue 2, pp 1001 – 1010. DOI: 10.1109/JIOT.2019.2947624.

[56] Zhang, Y, Gong, D, Hu, Y & Zhang, (2015), Feature selection algorithm based on bare bones particle swarm optimization, *Neurocomputing*, vol. 148, no. 1, pp. 150-157.

[57] Guoxiong Zhou, Wenzhuo Zhang, Aibin Chen, Mingfang He, Xueshuo Ma, (2019). Rapid Detection of Rice Disease Based on FCM-KM and Faster R-CNN Fusion, *IEEE Access*, vol 7, pp 143190 – 143206. DOI: 10.1109/ACCESS.2019.2943454.

[58] Zhang, Y, Wang, S, Phillips, P & Ji, G, (2014). Binary PSO with mutation operator for feature selection using decision tree applied to spam detection, *Knowledge-Based Systems*, vol. 64, pp. 22-31.

[59] Patil, R.R. Kumar, S. Chiwhane, S. Rani, R. Pippal, S.K. (2023). An Artificial-Intelligence-Based Novel Rice Grade Model for Severity Estimation of Rice Diseases. *Agriculture* 2023, 13, 47. https://doi.org/10.3390/agriculture13010047

[60] Chuang, LY, Tsai, SW & Yang, CH, (2011), Improved binary particle swarm optimization using catfish effect for feature selection, *Expert Systems with Applications*, vol. 38, no. 10, pp. 12699-12707.

[61] Billah, MM, Bhuiyan, MM & Ashraf, MA (2014), Computer Based Tools to Locate and Measure the Disease Infected Area of Rice Leaves, *Progressive Agriculture*, vol. 18, no. 2, pp. 209-213.

[62] Hussain, M. Al-Aqrabi, H, Munawar, M. Hill, R, (2022). Feature Mapping for Rice Leaf Defect Detection Based on a Custom Convolutional Architecture. *Foods*, 11, 3914. https://doi.org/10.3390/foods11233914

[63] Chandrashekar, G &Sahin, F (2014). A survey on feature selection methods, *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16- 28.

[64] Sengupta, S. Dutta, A. Abdelmohsen, S.A.M. Alyousef, H.A. Rahimi-Gorji., M. (2022). Development of a Rice Plant Disease Classification Model in Big Data Environment. *Bioengineering*, 2022, 9, 758. https://doi.org/10.3390/bioengineering9120758

[65] Emary, E, Zawbaa, HM & Hassanien, AE (2016). Binary grey wolf optimization approaches for feature selection, *Neurocomputing*, vol. 172, pp. 371-381.

[66] Jain, S. Sahni, R. Khargonkar, T. Gupta, H. Verma, O.P. Sharma, T.K. Bhardwaj, T. Agarwal, S. Kim, H.(2022). Automatic Rice Disease Detection and Assistance Framework

Using Deep Learning and a Chatbot. *Electronics*, 2022, 11, 2110. https://doi.org/10.3390/electronics11142110

[67] Jiang, G, Wang, X & Wang, Z (2015). Rice Disease Spots Extraction Based on Machine Vision, *International Journal of u-and e-Service, Science and Technology*, vol. 8, no. 3, pp. 211-220.

[68] Zhang, L.; Xie, L.; Wang, Z.; Huang, C (2022). Cascade Parallel Random Forest Algorithm for Predicting Rice Diseases in Big Data Analysis. *Electronics*, 2022, 11, 1079. https://doi.org/10.3390/electronics11071079

[69] Tabakhi, S, Moradi, P & Akhlaghian, F (2014). An unsupervised feature selection algorithm based on ant colony optimization, *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 112- 123.

[70] Rodrigues, D, Yang, XS, De Souza, AN & Papa, JP (2014). Binary flower pollination algorithm and its application to feature selection, *Recent Advances in Swarm Intelligence and Evolutionary Computation*, Springer, pp. 85-100.

[71] Tabakhi, S & Moradi, P (2015). Relevance-redundancy feature selection based on ant colony optimization, *Pattern recognition*, vol. 48, no. 9, pp. 2798-2811.

[72] Moayedikia, A, Jensen, R, Wiil, UK &Forsati, R (2015). Weighted bee colony algorithm for discrete optimization problems with application to feature selection, *Engineering Applications of Artificial Intelligence*, vol. 44, pp. 153-167.

[73] H. Samma, S. A. Suandi, N. A. Ismail, S. Sulaiman and L. L. Ping, (2021). Evolving Pre-Trained CNN Using Two-Layers Optimizer for Road Damage Detection from Drone Images, *in IEEE Access*, vol. 9, pp. 158215-158226. doi: 10.1109/ACCESS.2021.3131231.

[74] Kazimipour, B, Li, X & Qin, AK (2014). A review of population initialization techniques for evolutionary algorithms, *Proceedings on IEEE Congress on Evolutionary Computation*, pp. 2585-2592.

[75] H. Xue, Y. Bai, H. Hu, T. Xu and H. Liang (2019). A Novel Hybrid Model Based on TVIW-PSO-GSA Algorithm and Support Vector Machine for Classification Problems, in IEEE Access, vol. 7, pp. 27789-27801. doi: 10.1109/ACCESS.2019.2897644.

[76] S. Elsayed, R. Sarker and C. A. CoelloCoello (2017). Sequence-Based Deterministic Initialization for Evolutionary Algorithms, *in IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2911-2923, Sept. 2017, doi: 10.1109/TCYB.2016.2630722.

[77] Wang, Z.; Qin, J.; Hu, Z.; He, J. Tang, D (2022). Multi-Objective Antenna Design Based on BP Neural Network Surrogate Model Optimized by Improved Sparrow Search *Algorithm. Appl. Sci.* 12, 12543. https://doi.org/10.3390/app122412543

[78] Jones, R.A.C. (2021). Global Plant Virus Disease Pandemics and Epidemics. *Plants*, 10, 233. https://doi.org/10.3390/plants10020233

[79] Talukder, S (2011). Mathematical modelling and applications of particle swarm optimization, *Doctoral dissertation*, Blekinge Institute of Technology.

[80] Vieira, SM, Mendonça, LF, Farinha, GJ & Sousa, JM (2013). Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients, *Applied Soft Computing*, vol. 13, no. 8, pp. 3494-3504.

[81] Z. Daoqing and J. Mingyan, (2020). Parallel discrete lion swarm optimization algorithm for solving traveling salesman problem, *Journal of Systems Engineering and Electronics*, vol. 31, no. 4, pp. 751-760. doi: 10.23919/JSEE.2020.000050.

[82] H. Lin and P. Li, (2012). Classifying circuit performance using active-learning guided support vector machines, 2012 *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, San Jose, CA, USA, pp. 187-194.

[83] Billah, MM, Bhuiyan, MM & Ashraf, MA (2014). Computer Based Tools to Locate and Measure the Disease Infected Area of Rice Leaves, *Progressive Agriculture*, vol. 18, no. 2, pp. 209-213.

[84] M. Chichignoud and S. Loustau, (2014). Adaptive Noisy Clustering, *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7279-7292. doi: 10.1109/TIT.2014.2356577.

[85] Olson, E (2011). Particle shape factors and their use in image analysis- part 1: Theory, *Journal of GXP Compliance*, vol. 15, no. 3, pp. 85-89.

[86] Nasir, AFA, Rahman, MNA & Mamat, AR (2012). A study of image processing in agriculture application under high performance computing environment, *International Journal of Computer Science and Telecommunications*, vol. 3, no. 8, pp.120-124.

[87] G. Yang, G. Chen, Y. He, Z. Yan, Y. Guo and J. Ding (2020). Self-Supervised Collaborative Multi-Network for Fine-Grained Visual Categorization of Tomato Diseases, *IEEE Access*, vol. 8, pp. 211912-211923. doi: 10.1109/ACCESS.2020.3039345.

[88] S., Sachan, R., & Rajpal, D. (2020). Deep Convolutional Neural Network based Detection System for Real-time Corn Plant Disease Recognition. *Procedia Computer Science*, 167.

[89] Anjna, Sood, M., & Singh, P. K. (2020). Hybrid System for Detection and Classification of Plant Disease Using Qualitative Texture Features Analysis. *Procedia Computer Science*, 167.

[90] Agarwal, M., Singh, A., Arjaria, S., Sinha, A., & Gupta, S. (2020). ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network, *Procedia Computer Science*, 172-183.

[91] Ji, M., Zhang, L., & Wu, Q. (2020). Automatic grape leaf diseases identification via United Model based on multiple convolutional neural networks. *Information Processing in Agriculture*.

[92] Rahman, C. R., Arko, P. S., Ali, M. E., Iqbal Khan, M. A., Apon, S. H., Nowrin, F., & Wasif, A. (2020). Identification and recognition of rice diseases and pests using convolutional neural networks. *Biosystems Engineering*.

[93] Picon, A., Seitz, M., Alvarez-Gila, A., Mohnke, P., Ortiz-Barredo, A., & Echazarra, J. (2019). Crop conditional Convolutional Neural Networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. *Computers and Electronics in Agriculture*, 167(September).

[94] Lee, S. H., Goëau, H., Bonnet, P., & Joly, A. (2020). New perspectives on plant disease characterization based on deep learning. *Computers and Electronics in Agriculture*, 170.

[95] Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). Image Processing Techniques for Diagnosing Rice Plant Disease: A Survey. *Procedia Computer Science*, 167(2019).

[96] Anjna, Sood, M., & Singh, P. K. (2020). Hybrid System for Detection and Classification of Plant Disease Using Qualitative Texture Features Analysis. *Procedia Computer Science*, 167(2019).

[97] Hu, G., Yang, X., Zhang, Y., & Wan, M. (2019). Identification of tea leaf diseases by using an improved deep convolutional neural network. *Sustainable Computing: Informatics and Systems*.

[98] Chen, J., Chen, J., Zhang, D., Sun, Y., & Nanehkaran, Y. A. (2020). Using deep transfer learning for image-based plant disease identification. *Computers and Electronics in Agriculture*, 173(March).

[99] Agarwal, M., Gupta, S. K., & Biswas, K. K. (2020). Development of Efficient CNN model for Tomato crop disease identification. *Sustainable Computing: Informatics and Systems*.

[100] Arnal Barbedo, J. G. (2019). Plant disease identification from individual lesions and spots using deep learning, *Biosystems Engineering*, 180.

[101] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection, *Frontiers in Plant Science.*

[102] Kaur, S., Pandey, S., & Goel, S. (2018). Semi-automatic leaf disease detection and classification system for soybean culture. *IET Image Processing*, 12(6).

[103] KC, K., Yin, Z., Wu, M., & Wu, Z. (2019). Depth wise separable convolution architectures for plant disease classification. *Computers and Electronics in Agriculture*, 165.

[104] Khan, M. A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., & Saba, T. (2018). CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features, *Computers and Electronics in Agriculture*, 155.

[105] Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). Deep feature based rice leaf disease identification using support vector machine, *Computers and Electronics in Agriculture*, 175.

[106] Zhang, S., Zhang, S., Zhang, C., Wang, X., & Shi, Y. (2019). Cucumber leaf disease identification with global pooling dilated convolutional neural network. *Computers and Electronics in Agriculture.*

[107] Sholihati, R. A., Sulistijono, I. A., Risnumawan, A., & Kusumawati, E. (2020). Potato Leaf Disease Classification Using Deep Learning Approach. *IES 2020 - International Electronics Symposium: The Role of Autonomous and Intelligent Systems for Human Life and Comfort*, 392–397.

[108] Madhulatha, G., &Ramadevi, O. (2020). Recognition of plant diseases using convolutional neural network, *Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC* 2020, 738–743.

[109] Gayathri, S., Wise, D. C. J. W., Shamini, P. B., & Muthukumaran, N. (2020). Image Analysis and Detection of Tea Leaf Disease using Deep Learning, *Proceedings of the International Conference on Electronics and Sustainable Communication Systems*, ICESC, Icesc, 398–403.

[110] Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture.*

[111] Jasim, M. A., & Al-Tuwaijari, J. M. (2020). Plant Leaf Diseases Detection and Classification Using Image Processing and Deep Learning Techniques, *Proceedings of the 2020 International Conference on Computer Science and Software Engineering*, CSASE, 259–265.

[112] Ashok, S., Kishore, G., Rajesh, V., Suchitra, S., Gino Sophia, S. G., & Pavithra, B. (2020). Tomato leaf disease detection using deep learning techniques. *Proceedings of the 5th International Conference on Communication and Electronics Systems*, ICCES 2020, Icces, 979–983.

[113] Jiang, Di., Li, F., Yang, Y., & Yu, S. (2020). A Tomato Leaf Diseases Classification Method Based on Deep Learning. *Proceedings of the 32nd Chinese Control and Decision Conference*, CCDC 2020, 1446–1450.

[114] Mathulaprangsan, S., Lanthong, K., Jetpipattanapong, D., Sateanpattanakul, S., & Patarapuwadol, S. (2020). Rice Diseases Recognition Using Effective Deep Learning Models, *2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering*, ECTI DAMT and NCON 2020.

[115] Biswas, R., Basu, A., Nandy, A., Deb, A., Chowdhury, R., & Chanda, D. (2020). Identification of Pathological Disease in Plants using Deep Neural Networks - Distribution of Open VINOTM Toolkit, *Indo - Taiwan 2nd International Conference on Computing, Analytics and Networks, Indo-Taiwan ICAN 2020* - Proceedings, 45–48.

[116] Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). Image Processing Techniques for Diagnosing Rice Plant Disease: A Survey, *Procedia Computer Science*, 167, 516-530.

[117] Chapaneri, R., Desai, M., Goyal, A., Ghose, S., & Das, S. (2020, April). Plant Disease Detection: A Comprehensive Survey, In 2020 3rd *International Conference on Communication System, Computing and IT Applications* (CSCITA) (pp. 220-225). IEEE.

[118] Bera, T., Das, A., Sil, J., & Das, A. K. (2019). A survey on rice plant disease identification using image processing and data mining techniques, *In Emerging Technologies in Data Mining and Information Security* (pp. 365-376). Springer, Singapore.

[119] Pinki, F. T., Khatun, N., & Islam, S. M. (2017, December). Content based paddy leaf disease recognition and remedy prediction using support vector machine, In 2017 20th *International Conference of Computer and Information Technology* (ICCIT) (pp. 1-5). IEEE.

[120] Ghyar, B. S., & Birajdar, G. K. (2017, November). Computer vision-based approach to detect rice leaf diseases using texture and color descriptors, In 2017 *International Conference on Inventive Computing and Informatics* (ICICI) (pp. 1074-1078). IEEE.

[121] Zhang, J., Yan, L., & Hou, J. (2018, July). Recognition of rice leaf diseases based on salient characteristics. In 2018 13th *World Congress on Intelligent Control and Automation* (WCICA) (pp. 801-806). IEEE.

[122] Shah, J. P., Prajapati, H. B., & Dabhi, V. K. (2016, March). A survey on detection and classification of rice plant diseases. In 2016 *IEEE International Conference on Current Trends in Advanced Computing* (ICCTAC) (pp. 1-8). IEEE.

[123] Pothen, M. E., & Pai, M. L. (2020, March). Detection of Rice Leaf Diseases Using Image Processing, In 2020 *Fourth International Conference on Computing Methodologies and Communication* (ICCMC) (pp. 424-430). IEEE.

[124] Rani, B., Ojha, S., Kaushal, H., Kumar, S, Roy Chowdhury, T., Karmakar, S., & Jana, B. (2020). A Study on Detection of Brown Spot Disease in Rice Plant Using Machine Learning Technique: International Conference on Recent Trends in Artificial Intelligence, IoT, *Smart Cities & Application* (ICAISC 2020). IoT, Smart Cities & Application (ICAISC 2020) (July 9, 2020).

[125] Dataset: https://doi.org/10.6084/m9.figshare.21975287.v1

[126] Dataset: https://doi.org/10.6084/m9.figshare.21975239.v1

[127] Dataset: https://doi.org/10.6084/m9.figshare.21975236.v1

[128] Dataset: https://doi.org/10.6084/m9.figshare.22262632.v1

[129] Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). Deep feature based rice leaf disease identification using support vector machine, *Computers and Electronics in Agriculture*, 175, 105527.

[130] Yasmin, G., Das, A. K., & Ghosal, A. (2017, July). A hierarchical stratagem for rice leaf disease distinction, In 2017 *International Conference on Intelligent Computing, Instrumentation and Control Technologies* (ICICICT) (pp. 1177-1183). IEEE.

[131] Suresha, M., Shreekanth, K. N., & Thirumalesh, B. V. (2017, April). Recognition of diseases in paddy leaves using knn classifier, *In 2017 2nd International Conference for Convergence in Technology* (I2CT) (pp. 663-666). IEEE.

[132] Joshi, A. A., & Jadhav, B. D. (2016, December). Monitoring and controlling rice diseases using Image processing techniques, In 2016 *International Conference on Computing, Analytics and Security Trends* (CAST) (pp. 471-476). IEEE.

[133] Ahmed, K., Shahidi, T. R., Alam, S. M. I., & Momen, S. (2019, December). Rice leaf disease detection using machine learning techniques, In 2019 *International Conference on Sustainable Technologies for Industry* 4.0 (STI) (pp. 1-5). IEEE.

[134] Devi, T. G., Srinivasan, A., Sudha, S., & Narasimhan, D. (2019). Web enabled paddy disease detection using Compressed Sensing, *Mathematical Biosciences and Engineering*, 16(6), 7719-7733.

[135] Ramesh, S., & Vydeki, D. (2018, September). Rice blast disease detection and classification using machine learning algorithm, In 2018 *2nd International Conference on Micro-Electronics and Telecommunication Engineering* (ICMETE) (pp. 255-259). IEEE.

[136] Orillo, J. W., Cruz, J. D., Agapito, L., Satimbre, P. J., & Valenzuela, I. (2014, November). Identification of diseases in rice plant (oryza sativa) using back propagation Artificial Neural Network, In 2014 *International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management* (HNICEM) (pp. 1-6). IEEE.

[137] Lu, Y., Yi, S., Zeng, N., Liu, Y., & Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks, *Neurocomputing*, 267, 378-384.

[138] Asfarian, A., Herdiyeni, Y., Rauf, A., & Mutaqin, K. H. (2013, November). Paddy diseases identification with texture analysis using fractal descriptors based on fourier spectrum. In 2013 *International Conference on Computer, Control, Informatics and Its Applications* (IC3INA) (pp. 77-81). IEEE.

[139] Majid, K., Herdiyeni, Y., & Rauf, A. (2013, September). I-PEDIA: Mobile application for paddy disease identification using fuzzy entropy and probabilistic neural network, In 2013 *International Conference on Advanced Computer Science and Information Systems* (ICACSIS) (pp. 403-406). IEEE.

[140] Karunanithi, A., Singh, A.S., Kannapiran, T. (2022). Enhanced hybrid neural networks (CoAtNet) for paddy crops disease detection and classification, *Revue d'Intelligence Artificielle*, Vol. 36, No. 5, pp. 671-679. https://doi.org/10.18280/ria.360503

[141] Agustina, E., Pratomo, I., Wibawa, A. D., & Rahayu, S. (2017, August). Expert system for diagnosis pests and diseases of the rice plant using forward chaining and certainty factor method, In 2017 *International Seminar on Intelligent Technology and Its Applications* (ISITIA) (pp. 266-270). IEEE.

[142] Devi, D. A., & Muthukannan, K. (2014, May). Analysis of segmentation scheme for diseased rice leaves, In 2014 *IEEE International Conference on Advanced Communications, Control and Computing Technologies* (pp. 1374-1378). IEEE.

[143] Tian, L., Wan, Z., Li, D., Jiang, J., Yao, X., Cao, Q. & Cheng, T. (2018, July). Detecting Rice Blast Disease Using Model Inverted Biochemical Variables from Close-Range Reflectance Imagery of Fresh Leaves. In IGARSS 2018-2018 *IEEE International Geoscience and Remote Sensing Symposium* (pp. 2749-2752). IEEE.

[144] Ihsanuddin, E. W. H., &Rahmatulloh, A. (2019). Identification of Bacterial Leaf Blight and Brown Spot Disease in Rice Plants with Image Processing Approach. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika* (JITEKI), 5(2), 59-67.

[145] Khatua, P. K., Ramachandaramurthy, V. K., Kasinathan, P., Yong, J. Y., Pasupuleti, J., & Rajagopalan, A. (2020). Application and assessment of internet of things toward the

sustainability of energy systems: Challenges and issues, *Sustainable Cities and Society*, 53, 101957.

[146] Gupta, A.K., Gupta, K., Jadhav, J., Deolekar, R. V, Nerurkar, A., Deshpande, S. (2019). Plant Disease Prediction using Deep Learning and IoT. 6th *International Conference on Computing for Sustainable Global Development* (INDIACom), pp. 902–907.

[147] Atole, R. R., & Park, D. (2018). A multiclass deep convolutional neural network classifier for detection of common rice plant anomalies, *International Journal of Advanced Computer Science and Applications*, 9(1).

[148] Li, D., Wang, R., Xie, C., Liu, L., Zhang, J., Li, R., & Liu, W. (2020). A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network, *Sensors*, 20(3), 578. https://doi.org/10.3390/s20030578

[149] Bakar, M. A., Abdullah, A. H., Rahim, N. A., Yazid, H., Misman, S. N., & Masnan, M. J. (2018). Rice Leaf Blast Disease Detection Using Multi-Level Colour Image Thresholding, *Journal of Telecommunication, Electronic and Computer Engineering* (JTEC), 10(1-15), 1-6.

[150] Chawal, B., & Panday, S. P. (2019). Rice Plant Disease Detection using Twin Support Vector Machine (TSVM), *Journal of Science and Engineering*, 7, 61-69. https://doi.org/10.3126/jsce.v7i0.26794

[151] Zhou, C., Zhang, Z., Zhou, S., Xing, J., Wu, Q., & Song, J. (2021). Grape leaf spot identification under limited samples by fine grained-GAN, *IEEE Access*, 9, 100480-100489.

[152] Dang, K. B., Burkhard, B., Windhorst, W., & Müller, F. (2019). Application of a hybrid neural-fuzzy inference system for mapping crop suitability areas and predicting rice yields, *Environmental Modelling & Software*, 114, 166-180. https://doi.org/https://doi.org/10.1016/j.envsoft.2019.01.015

[153] A. K and A. S. Singh. (2021). Detection of Paddy Crops Diseases and Early Diagnosis Using Faster Regional Convolutional Neural Networks, *International Conference on Advance Computing and Innovative Technologies in Engineering* (ICACITE), pp. 898-902. doi: 10.1109/ICACITE51222.2021.9404759

[154] Dhumane, A. V, Prasad, R.S. (2018). Fractional Gravitational Grey Wolf Optimization to Multi-Path Data Transmission in IoT, *Wirel. Pers. Commun.* 102, 411–436. https://doi.org/10.1007/s11277-018-5850-y

[155] Jiang, F., Gong, M., Zhan, T., & Fan, X. (2019). A semisupervised GAN-based multiple change detection framework in multi-spectral images, *IEEE Geoscience and Remote Sensing Letters*, 17(7), 1223-1227. https://doi.org/10.1109/LGRS.2019.2941318

[156] Mishra, M., Choudhury, P and Pati, B. (2020). Modified ride-NN optimizer for the IoT based plant disease detection, *Journal of Ambient Intelligence and Humanized Computing*, 12, 691-703.

[157] Orchi, H., Sadik, M., & Khaldoun, M. (2021). On using artificial intelligence and the internet of things for crop disease detection: A contemporary survey, *Agriculture*, 12(1), 9.

[158] Paul Shekonya Kanda, Kewen Xia, Olanrewaju Hazzan Sanusi. (2021). A Deep Learning-Based Recognition Technique for Plant Leaf Classification, *IEEE Access*, 9, 162590 – 162613.

[159] Amara, J., Bouaziz, B., & Algergawy, A. (2017, March). A Deep Learning-based Approach for Banana Leaf Diseases Classification, In BTW (workshops) (Vol. 266, pp. 79-88).

[160] Zhong, L., Hu, L., & Zhou, H. (2019). Deep learning based multi-temporal crop classification, *Remote sensing of environment*, 221, 430-443. https://doi.org/https://doi.org/10.1016/j.rse.2018.11.032

[161] Richey, B., Majumder, S., Shirvaikar, M., & Kehtarnavaz, N. (2020, April). Real-time detection of maize crop disease via a deep learning-based smartphone app, *In Real-Time Image Processing and Deep Learning 2020* (Vol. 11401, pp. 23-29). SPIE. https://doi.org/10.1117/12.2557317

[162] Rodríguez, L., Castillo, O., García, M., & Soria, J. (2020). A new randomness approach based on sine waves to improve performance in metaheuristic algorithms, *Soft Computing*, 24(16), 11989-12011.

[163] K., A. ., Shanker Singh, A., & K., T. (2022). An Automatic Rice Plant Disease Detection Model Built With Unstructured Data Using IMDT Tiling and CNN Cognitive Object Detection, *International Journal on Recent and Innovation Trends in Computing and Communication*, 10(12), 65–75. https://doi.org/10.17762/ijritcc.v10i12.5887

[164] Emary, E, Zawbaa, HM & Hassanien, (2016). Binary grey wolf optimization approaches for feature selection, *Neurocomputing*, vol. 172, pp. 371-381.