International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Design and Optimization of a VLSI Architecture for Cotton Crop Disease Detection using PCA and Random Forest

[1]Priti Kumari, [2]Prof. Suresh S. Gawande
M. Tech. Scholar, Department of Electronics and Communication, RKDF College of Engineering, Bhabha University, Bhopal[1]
Guide, Department of Electronics and Communication, RKDF College of Engineering, Bhabha University, Bhopal[2]

Abstract:- Cotton is one of the most important commercial crops, and its productivity is significantly affected by various diseases that reduce crop quality and agricultural yield. Early and accurate detection of cotton crop diseases is essential for improving agricultural productivity and supporting effective disease management. Traditional disease identification methods primarily depend on manual observation and expert knowledge, which can be time-consuming, labor-intensive, and prone to human error. The proposed methodology integrates image processing, feature extraction, dimensionality reduction, machine learning, and hardware implementation to achieve an efficient and reliable disease detection system. Initially, cotton leaf images are collected and preprocessed using resizing, noise removal, and image enhancement techniques. Subsequently, relevant features are extracted from the images, and Principal Component Analysis (PCA) is applied to reduce feature dimensionality while preserving the most significant information. The optimized feature set is then provided to the Random Forest classifier for identifying and classifying different cotton leaf diseases. Furthermore, the proposed machine learning model is mapped onto a VLSI architecture to improve computational efficiency, reduce hardware complexity, and minimize power consumption and processing delay. The hardware architecture is designed by optimizing arithmetic operations, memory utilization, and parallel processing capabilities. Performance evaluation is carried out using classification metrics such as accuracy, precision, recall, and F1-score, along with hardware parameters including area utilization, power consumption, and propagation delay. The proposed PCA-Random Forest-based VLSI architecture aims to provide an accurate, low-power, area-efficient, and high-speed solution for automated cotton crop disease detection, making it suitable for real-time smart agriculture applications.

Keywords: Cotton Crop Disease Detection, VLSI Architecture, Principal Component Analysis (PCA), Random Forest, Image Processing, Feature Extraction

I. INTRODUCTION

Agriculture plays a significant role in the economic development of many countries, particularly in developing nations such as India. Cotton is one of the most important commercial crops and is widely used in the textile industry. It provides employment opportunities to millions of farmers and contributes significantly to the agricultural economy.

However, cotton production is frequently affected by various diseases caused by fungi, bacteria, viruses, and environmental conditions. Common cotton leaf diseases include bacterial blight, leaf curl disease, target spot, and Alternaria leaf spot. These diseases negatively affect plant growth, reduce crop quality, and decrease overall agricultural productivity. Therefore, early and accurate identification of cotton crop diseases is essential for improving crop management and minimizing economic losses [1, 2].

Traditional methods of cotton disease identification mainly depend on visual inspection and expert knowledge. Although these approaches can identify diseases based on visible symptoms, they are time-consuming, labour-intensive, and may produce inaccurate results due to similarities in disease symptoms. Moreover, manual inspection becomes difficult when monitoring large agricultural fields. With the advancement of artificial intelligence, machine learning, image processing, and hardware technologies, automated disease detection systems have become an effective solution for improving agricultural monitoring and decision-making [3].

Machine learning techniques provide efficient approaches for identifying and classifying different cotton crop diseases using image-based features. Among various machine learning algorithms, RF is an ensemble learning technique that combines multiple decision trees to improve classification performance. It provides better generalization, handles complex datasets, and reduces the possibility of overfitting compared with individual decision tree models. However, high-dimensional image features may increase computational complexity, memory requirements, and processing time [4, 5]. To overcome these limitations, PCA is employed as a dimensionality reduction technique. PCA transforms the original feature space into a smaller set of principal components while preserving the maximum possible variance in the data. The integration of PCA with Random Forest can reduce redundant features, improve computational efficiency, and support effective disease classification [6].

Furthermore, the implementation of machine learning algorithms on conventional computing platforms may introduce processing delays and higher power requirements, particularly in resource-constrained agricultural monitoring devices. Very Large-Scale Integration (VLSI) technology provides an efficient platform for designing compact, high-speed, and low-power hardware architectures. VLSI-based implementation enables parallel processing, optimized data handling, and efficient utilization of hardware resources. By integrating PCA-based feature reduction with a Random Forest classification architecture, the computational requirements of cotton disease detection can be reduced, making the system more suitable for hardware implementation [7, 8].

In this research, a VLSI architecture is proposed for cotton crop disease detection using PCA and Random Forest techniques. Initially, cotton leaf images are collected and preprocessed to remove noise and improve image quality [9]. Relevant image features are extracted and supplied to the PCA algorithm to reduce feature dimensionality. The reduced feature set is then classified using the Random Forest algorithm to identify different cotton crop diseases.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

II. COTTON IMAGE AND DISEASE

Cotton leaf images provide important visual information for the identification and classification of diseases affecting cotton plants. The images generally contain healthy and diseased leaves exhibiting different symptoms such as leaf_curl, fusarium_wilt, bacterial bright and healthy image. These visual characteristics can be analyzed using image processing and neural network techniques to automatically distinguish healthy plants from different disease categories.

For an automated cotton disease detection system, the image dataset is generally divided into training, validation, and testing sets. During preprocessing, images may be resized and normalized to obtain a consistent input format. Data augmentation techniques such as rotation, flipping, cropping, and brightness adjustment can also be applied to increase dataset diversity and improve model generalization. The processed images are then provided to the neural network for feature learning and disease classification.



Figure 1: Leaf_curl disease



Figure 2: Bacterial bright

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |



Figure 3: Fussarium_wilt



Figure 4: Healthy Image

III. RESEARCH METHODOLOGY

The flowchart illustrates the proposed methodology for Design and Optimization of a VLSI Architecture for Cotton Crop Disease Detection using PCA and Random Forest. The complete process consists of eight major steps, starting from library importing and dataset collection and ending with performance evaluation.

1. Import Library

The first step involves importing the necessary Python libraries required for image processing, feature extraction, dimensionality reduction, machine learning, and performance evaluation. Libraries such as NumPy, Pandas, OpenCV, Matplotlib, Scikit-learn, and Seaborn are used to perform different computational operations. These libraries provide essential functions for image manipulation, data analysis, PCA implementation, Random Forest classification, and visualization of experimental results.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

2. Cotton Disease Dataset

In this step, the cotton disease dataset is collected to train and evaluate the proposed disease detection model. The dataset contains images of cotton leaves belonging to different disease categories, including healthy and infected leaves. These images provide the input data for developing an automated cotton disease classification system. A properly labelled dataset is essential for training the machine learning model and identifying different disease patterns.

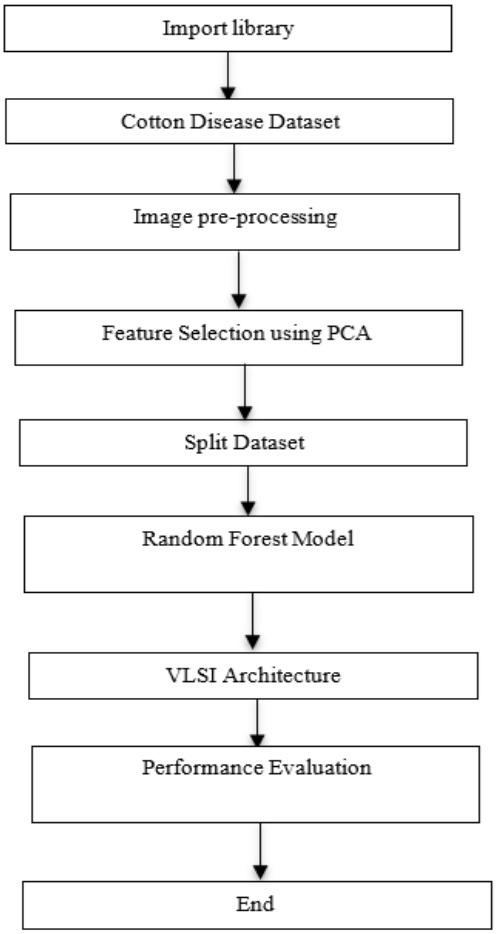

Figure 5: Flow Chart of Methodology

3. Image Pre-processing

Image preprocessing is performed to improve the quality and consistency of the collected cotton leaf images. Initially, images are resized into a fixed dimension to maintain uniformity. Noise removal, image enhancement, and pixel normalization techniques are applied to improve image quality and reduce unwanted variations. These operations help extract meaningful information from the images and improve the effectiveness of subsequent feature extraction and classification processes.

4. Feature Selection Using PCA

Principal Component Analysis (PCA) is applied to reduce the dimensionality of the extracted image features. It transforms the original high-dimensional feature space into a smaller set of

principal components while preserving the maximum possible variance in the data. PCA eliminates redundant information and reduces computational complexity. This process helps decrease memory requirements and processing time while providing optimized input features for the Random Forest classifier.

5. Split Dataset

After feature selection using PCA, the complete dataset is divided into training and testing subsets. The training dataset is used to develop and train the Random Forest classification model, whereas the testing dataset is used to evaluate its performance on previously unseen samples. Generally, the dataset can be divided into 80% training data and 20% testing data. This separation helps assess the model's generalization capability and classification reliability.

6. Random Forest Model

The Random Forest algorithm is employed to classify cotton leaf images into their respective disease categories. It is an ensemble machine learning technique that combines multiple decision trees to generate the final classification result. Each decision tree independently predicts the class, and the final output is determined through majority voting. Random Forest provides effective classification performance, handles complex feature relationships, and reduces overfitting. The trained model is subsequently considered for hardware-oriented implementation in the proposed VLSI architecture.

7. VLSI Architecture

In this step, the proposed PCA and Random Forest-based disease detection methodology is designed for VLSI hardware implementation. The computational operations of the classification system are mapped into suitable hardware structures. Different architectural optimization techniques are considered to improve processing speed, reduce power consumption, and minimize hardware resource utilization. The architecture focuses on efficient data processing, optimized arithmetic operations, and hardware resource management. The main objective is to develop a compact, low-power, and high-speed architecture suitable for real-time cotton crop disease detection.

8. Performance Evaluation

The final stage evaluates the performance of the proposed system using both classification and hardware performance parameters.

IV. STEP-BY-STEP EXPLANATION OF PROPOSED VLSI HARDWARE ARCHITECTURE

Step 1: Image Input and Memory Unit

The first step of the proposed VLSI architecture involves receiving cotton leaf images through an input interface. The input images are converted into digital pixel values and stored in the memory unit for further processing. The memory unit consists of RAM or Block RAM (BRAM), input registers, an address generator, and a data buffer. The input interface transfers pixel information to the memory, while the address generator controls the storage locations. The registers temporarily store the incoming image data and provide it to the next processing

stage. This unit ensures efficient data transfer and proper synchronization between the input image and the subsequent hardware processing operations.

Step 2: Image Pre-processing Unit

The image preprocessing unit is responsible for improving the quality of the collected cotton leaf images before feature extraction. In this stage, different image processing operations, including image resizing, noise removal, pixel normalization, and image enhancement, are performed. The hardware architecture consists of arithmetic units, filtering circuits, line buffers, pixel registers, and control logic. The resizing operation maintains a uniform image dimension, whereas normalization converts pixel values into a consistent numerical range. Noise removal techniques eliminate unwanted variations from the input images, and image enhancement improves the visibility of disease-related patterns. The processed image is then transferred to the feature extraction unit for further analysis.

Step 3: Feature Extraction Unit

The feature extraction unit converts the preprocessed cotton leaf images into numerical feature vectors that represent the important characteristics of the leaves. These characteristics may include color information, texture patterns, and shape-related features. The hardware structure consists of feature calculation circuits, multipliers, adders, accumulators, and feature registers. The arithmetic circuits perform the required mathematical operations to calculate the different image features. The extracted features are temporarily stored in registers and organized into a feature vector. This feature vector contains important information about the cotton leaf and is transferred to the PCA hardware architecture for dimensionality reduction.

Step 4: PCA Hardware Architecture

The Principal Component Analysis (PCA) hardware architecture is designed to reduce the dimensionality of the extracted feature vector while preserving the most significant information. Initially, the input feature vector is supplied to the mean subtraction unit, where the corresponding mean values are subtracted from the original features. The centered feature values are then multiplied by the PCA projection coefficients stored in the coefficient memory. The multiplier array performs multiplication operations, while the adder tree combines the multiplication results to generate the principal component values. An accumulator is used to collect the intermediate results, and the output registers store the reduced feature vector. The PCA hardware architecture mainly consists of subtractors, multipliers, adders, accumulator circuits, coefficient memory, and control units. The reduced features require less memory and computational resources than the original feature set, making the subsequent classification process more efficient.

Step 5: Hardware Optimization and Performance Evaluation

The final step involves optimizing and evaluating the complete VLSI architecture to achieve improved hardware performance. Different optimization techniques, such as pipelining, parallel processing, fixed-point arithmetic, resource sharing, and clock gating, can be applied to reduce hardware complexity, processing delay, and power consumption. Pipelining allows multiple processing stages to operate simultaneously, whereas parallel processing improves

computational speed by executing independent operations concurrently. Fixed-point arithmetic reduces the complexity of mathematical operations compared with floating-point implementation. Resource sharing minimizes hardware area by reusing arithmetic units, while clock gating reduces unnecessary switching activity. The proposed architecture is evaluated using hardware parameters such as Look-Up Tables (LUTs), Flip-Flops, slices, power consumption, propagation delay, and maximum operating frequency. Additionally, classification performance is evaluated using accuracy, precision, recall, and F1-score. These evaluation parameters help determine the effectiveness of the proposed PCA and Random Forest-based VLSI architecture for accurate, low-power, and high-speed cotton crop disease detection.

V. SIMULATION RESULT

Table 1: Recall Results

Cotton Disease		**Recall**
Healthy Image		97.18%
Cotton Disease	Cotton leaf curl	96.88%
	Bacterial bright	97.77%
	Fussarium_wilt	98.33%

Table 1 shows the recall metric of cotton disease and healthy image. Recall in PCA with RF algorithm is high than the all-other supervised machine learning algorithm. PCA with RF algorithm gives a Recall of healthy image is 97.18%, Cotton leaf curl is 96.88%, Bacterial bright is 97.77% and Fussarium_wilt is 98.33%. Fig. 6 shows the graphical representation of the different parameter.

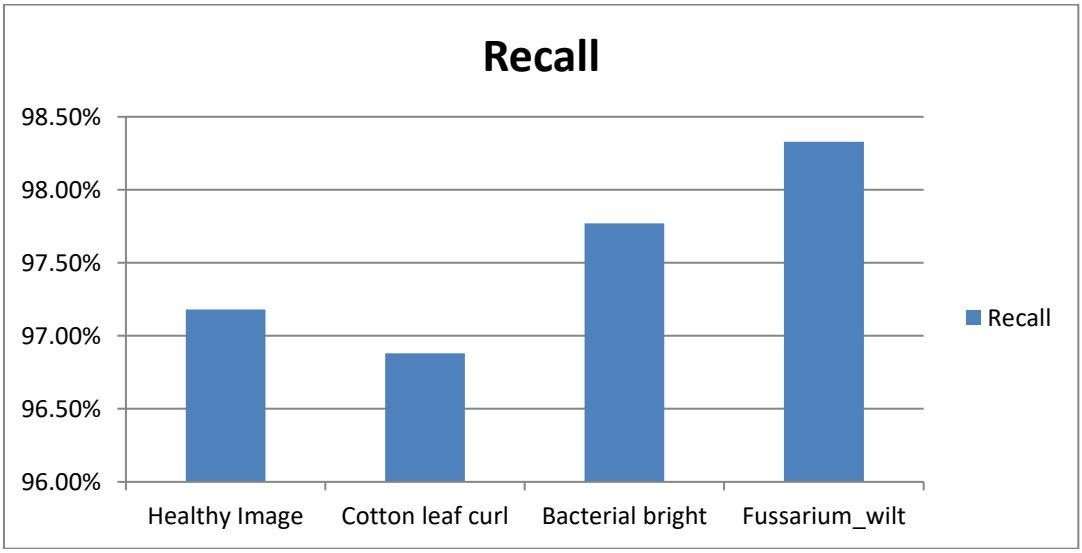

Figure 6: Graph of Recall

Table 2: Accuracy Results

Cotton Disease		Accuracy
Healthy Image		97.18%
Cotton Disease	Cotton leaf curl	96.88%
	Bacterial bright	97.77%
	Fussarium_wilt	98.33%

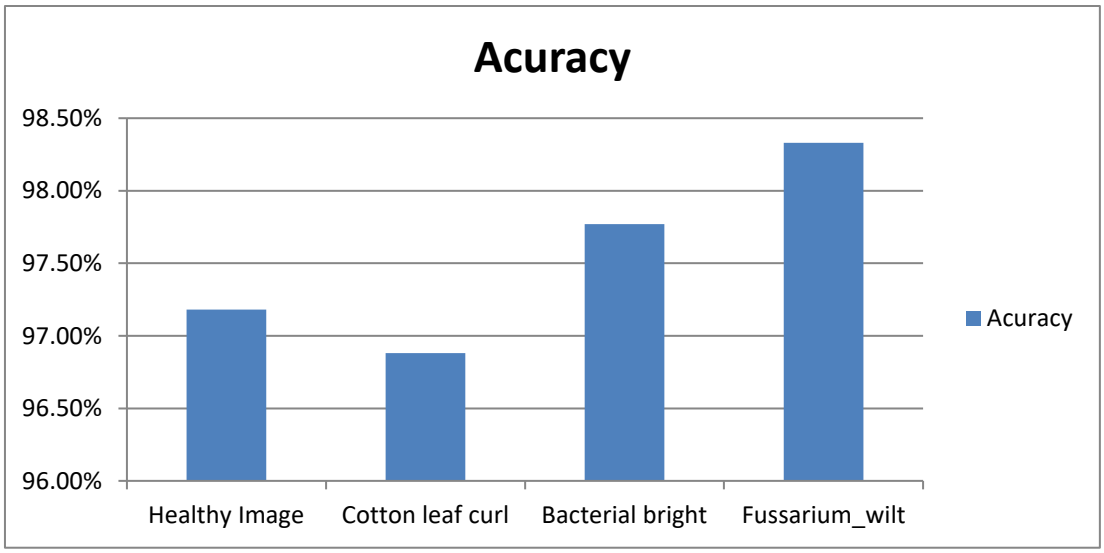

Figure 7: Graph of Accuracy

PCA with RF algorithm gives a healthy image of accuracy is 97.18%, Cotton leaf curl is 96.88%, Bacterial bright is 97.77% and Fussarium_wilt is 98.33% is represent in table 2. Fig. 7 shows the graphical representation of the different parameters.

VI. CONCLUSION

The proposed approach integrates image processing, feature extraction, dimensionality reduction, machine learning, and hardware implementation to develop an efficient and reliable cotton disease detection system. Initially, cotton leaf images are pre-processed to improve image quality and extract meaningful features. PCA is applied to reduce the dimensionality of the extracted features while preserving important information, thereby minimizing computational complexity and memory requirements. Subsequently, the reduced feature vectors are processed using the Random Forest classifier, which combines multiple decision trees to identify and classify different cotton crop diseases through the majority voting technique.

Furthermore, the proposed machine learning model is mapped into a VLSI architecture consisting of image input units, preprocessing circuits, feature extraction modules, PCA processing units, decision tree structures, threshold comparators, voting circuits, registers,

and control units. Different hardware optimization techniques, including pipelining, parallel processing, fixed-point arithmetic, and resource sharing, can be employed to improve processing speed and reduce hardware complexity. The overall architecture is evaluated using classification parameters such as accuracy and recall along with hardware parameters including power consumption, area utilization, and propagation delay. The proposed architecture aims to provide an accurate, low-power, area-efficient, and high-speed solution for automated cotton crop disease detection. This research can contribute to the development of intelligent agricultural monitoring systems and support real-time disease identification, ultimately improving cotton crop management and agricultural productivity.

REFERENCES

[1] R. Kumar *et al.*, "Hybrid Approach of Cotton Disease Detection for Enhanced Crop Health and Yield", in *IEEE Access*, vol. 12, pp. 132495-132507, 2024.

[2] I. Ahmed, G. Habib and P. K. Yadav, "An approach to identify and classify agricultural crop diseases using machine learning and deep learning techniques", *Proc. Int. Conf. Emerg. Smart Comput. Informat. (ESCI)*, pp. 1-6, Mar. 2023.

[3] S. Bondre and D. Patil, "Recent advances in agricultural disease image recognition technologies: A review", *Concurrency Computation Pract. Exper.*, vol. 35, no. 9, Apr. 2023.

[4] M. M. Islam, M. A. Talukder, M. R. A. Sarker, M. A. Uddin, A. Akhter, S. Sharmin, et al., "A deep learning model for cotton disease prediction using fine-tuning with smart web application in agriculture", *Intell. Syst. Appl.*, vol. 20, Nov. 2023.

[5] R. Mahum, H. Munir, Z.-U.-N. Mughal, M. Awais, F. S. Khan, M. Saqlain, et al., "A novel framework for potato leaf disease detection using an efficient deep learning model", *Hum. Ecol. Risk Assessment Int. J.*, vol. 29, no. 2, pp. 303-326, Feb. 2023.

[6] L. Goel and J. Nagpal, "A systematic review of recent machine learning techniques for plant disease identification and classification", *IETE Tech. Rev.*, vol. 40, no. 3, pp. 423-439, May 2023.

[7] C. Sarkar, D. Gupta, U. Gupta and B. B. Hazarika, "Leaf disease detection using machine learning and deep learning: Review and challenges", *Appl. Soft Comput.*, vol. 145, Sep. 2023.

[8] Y. Yuan, L. Chen, H.Wu, and L. Li, "Advanced agricultural disease image recognition technologies: A review," Inf. Process. Agricult., vol. 9, no. 1, pp. 48–59, Mar. 2022.

[9] V. Kathole and M. Munot, "Deep learning models for tomato plant disease detection," in Advanced Machine Intelligence and Signal Processing. Singapore: Springer, pp. 679–686, 2022.

[10] Q. Pan, J.-F. Qiao, R. Wang, H.-L. Yu, C. Wang, K. Taylor, and H.-Y. Pan, ''Intelligent diagnosis of northern corn leaf blight with deep learning model," J. Integrative Agricult., vol. 21, no. 4, pp. 1094–1105, 2022.

[11] U. Chauhan, “ResTS: Residual deep interpretable architecture for plant disease detection,” Inf. Process. Agricult., vol. 9, no. 2, pp. 212–223, Jun. 2022.
[12] Vallabhajosyula, V. Sistla, and V. K. K. Kolli, ‘‘Transfer learning-based deep ensemble neural network for plant leaf disease detection,” J. Plant Diseases Protection, vol. 129, no. 3, pp. 545–558, Jun. 2022.
[13] Chen, Z. Yuan, S. Chen, and X. Zou, “Plant disease recognition model based on improved YOLOv5,” Agronomy, vol. 12, no. 2, p. 365, Jan. 2022.
[14] Wang, X. He, K. Feng, and H. Zhu, ‘‘Plant disease detection and classification method based on the optimized lightweight YOLOv5 model”, Agriculture, vol. 12, no. 7, p. 931, Jun. 2022.
[15] J. Karthika, M. Santhose and T. Sharan, "Disease detection in cotton leaf spot using image processing", *J. Phys. Conf. Ser.*, vol. 1916, no. 1, 2021.