Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 5, 2026

SOYBEAN LEAF DISEASE CLASSIFICATION USING MOBILENETV2 WITH LIGHTWEIGHT DEEP LEARNING IN RESOURCE CONSTRAINED AGRICULTURAL ENVIRONMENT

Qamer Un Nisa[1], Muhamamd Usman Javeed[*2], Asim Ali Rao[3], Muhammad Nauman[4], Waheed Yousuf Ramay[5], Zaira Marriam[6], Rabia Rasool[7]

[1]*Department of Computer Science, University of Agriculture Faisalabad, Pakistan*
[*2]*Department of Computer Science, COMSATS University of Islamabad, Sahiwal, Pakistan*
[3]*Department of Computer Science, Muhammad Nawaz Sharif University of Engineering and Technology, Multan, Pakistan*
[4]*Department of Computer Science and Artificial Intelligence, Southwest Jiatong University Chengdu China,*
[5]*Department of Computer Science, Cholistan University of Veterinary and Animal Sciences, Bahwalpur, Pakistan*
[6,7]*Department of Computer Science, University of Okara, Pakistan*

[1]nisaqamerun524@gmail.com, [*2]usmanjavveed@gmail.com, [3]asimalirao@mnsuet.edu.pk, [4]mr.nauman.edu@gmail.com, [5]waheedramay@cuvas.edu.pk, [6]zairamarriam1234@gmail.com, [7]rabiarasool4028028@gmail.com

DOI: https://doi.org/10.5281/zenodo.20131425

Keywords
Agricultural AI, Convolutional Neural Networks, MobileNetV2, Soybean Disease Detection, Transfer Learning

Article History
Received: 14 March 2026
Accepted: 22 April 2026
Published: 12 May 2026

Corresponding Author: *
Muhamamd Usman Javeed

Abstract

Soybean is one of the most significant oilseed crops in the world and is highly susceptible to a number of diseases of the leaves that cause losses in quality and yield. Early and accurate detection of disease is vital for successful early disease management and sustainable agriculture production. In this study, a lightweight deep learning framework MobileNetV2 is proposed for an automatic soybean leaf disease diagnosis system in resource-poor agricultural settings. Images of 2000 soybean leaves were downloaded from agricultural fields of the Pakistan region and a region-specific dataset named SDD-2025 was developed. The data is split into three disease classes: Charcoal Rot, Mosaic Virus and Pest Infestation. Pre-processing methods such as rotation, flipping, zooming and brightness adjustment were used to augment the data to help prevent overfitting and to make the model more generalizable. The feature extraction and classification was performed using transfer learning with a pretrained MobileNetV2 architecture. The dataset was split in the ratio 70:15:15 for training, validation and testing respectively. The experimental results were evaluated by calculating accuracy, precision, recall, F1 score and ROC analysis. The proposed model successfully classified the soybean diseases with 96.14% accuracy, which shows the effectiveness of lightweight CNNs to recognize soybean diseases. The trained framework yields an automated and efficient early disease diagnosis solution, which requires minimal human involvement. Moreover, the low weight of the model makes it well suited for implementing real-time agricultural monitoring and smart farming applications. Although the study results are positive, the number of records and geographical diversity are limited. In order to make the framework more robust and generic, it *will be crucial to expand the data set, add more illness classes, and test it under various climatic conditions.*

1. INTRODUCTION

Agriculture Department is the main supplier of food to a rapidly growing population. Agriculture diseases, nevertheless, endanger the continuity of this highly prized resource. Agricultural production losses and forest economy are immense. Farmers can recover some portions of $11 million losses caused by soya bean rust (a plant disease of soybeans) just by cleansing 20% of the infected crop. (R.J. Hay 2015). Soybeans are pressed for oil and used-up for their protein as an Asian staple. Soya bean ingredients are increasing in importance as food and drink companies attempt to reduce their expenses while holding their protein supplies for human intake steady or improving them. Soybeans have been long held in esteem as a top-notch oil seed based on their great protein calm and other alimental values (abundant mineral salts and vitamins A, B, and D). So far, we have found that Pakistan Usage consumed 1.09 million tons of pure oil, 29% of which was self-produced domestically and 71% was imported. Every environment, at every stage of growth, and under every mechanical, physiological, and biological stress can retard the regular growth and expansion of oil seeds crops. A combination of several abiotic and biotic factors such as pollutants, pests, and diseases represent serious though variable threats to cropping depending upon region, season, and affected plant component. Due to its high protein complacent (40 percent), content for oil (20 percent) and meal industry based on the seed, soybean is conveniently used in human and animal nutriment and even in industrial applications. Soya bean oil has been pressed and tested and found to have no toxins. Unsaturated fatty acids that are rich in the oil are responsible for its lowering effect on cholesterol levels (C. Melenotte n.d.). Regrettably, edible oil production in Pakistan is not enough to meet the demands. Palm oil is imported from Malaysia while soybean oil is imported from America and Argentina. Importation of edible oil is an expensive undertaking that is valued at 1 billion and is second most expensive import after fossil fuels. In Pakistan, effort is being made in cultivating non-conventional oil seed crops like soybeans and sunflowers in an attempt to meet the growing demands in the country that cannot be met by conventional oil-seed crops like rape seed and mustard ray. A shortage of edible oils has existed since the 1970s, despite the fact that per capita consumption and local edible oil production have both increased rapidly along with the global population.

Macrophomina phaseout Carbon rot in soybean is due to fungus Architectonic sabbatical and Sclerotic sabbatical (I. K. J. Rashid n.d.), (I. K. J. Rashid n.d.), (M. Souhait n.d.). Plants with weaker vitality had their seeds infected with diseases like chromium, cryptosporidium, and metallic element. Soybean seeds were infected with macrophage phaseolina due to a precipitation deficit along with excessive heat during early soybean growth. (M. I. J. Rashid n.d.) Therefore, there are countless methods for plant pathology detection, and early catching and diagnosis of works simplicity is imperative for the implementation of appropriate measures. Where there are no observable symptoms or symptoms are not manifested until too far gone to do any good, a detailed analysis, frequently involving powerful microscopes, must be conducted! In other instances, signs can only be manifested in invisible areas (B. G. Blackburn n.d.). Symptoms due to diseases can be manifested on leaves, stems, fruits, seeds, etc. This study focuses on detection and soybean plant disease organizing via signs borne by diseases on leaves of plants. Diagnosis, or at least preliminary screening, is almost invariably done by a human person. (B. G. Blackburn n.d.) Conventional crop disease detection is done by skilled pathologists and producers, who look at leaves to spot symptoms. This procedure is manual work intensive, costly, and inconsistent due to the fact that human detection is relied upon. The value of this process may also be eroded due to environmental distractions as well as the result of optical illusions such as that due to parallax effect.

Thus, instead of time-consuming and possibly harmful eye assessments of plant health, sensing remote sensing application have also been found to be a consistent method for separating ill from healthy leaves. Soybean diseases have been automated with digital image processing [8, 9], pattern recognition (Kluger n.d.), and computer vision as per a series of researchers. Furthermore, a series of various varieties of crops and fruits are investigated with the latest techniques, such as grapes, pomegranates tomatoes, and maize. Convolutional neural networks have been applied recently in object identification and image categorization amongst others. Deep neural networks (DNNs) like the convolutional kind are increasingly being utilised to process images in ways that mimic the human visual system. There have been many proposed CNN designs for application in object perception. The LeNet and Alex Net models are among those used as standards for diverse tasks (A. Kirubarajan n.d.). Soybean (Glycine max) is leguminous growth in the worldwide, serving as a valuable beginning of protein and oil for human intake and animal feed. However, soya bean production faces significant challenges from various biotic and abiotic stressors, including fungal, bacterial, and viral diseases that affect crop yield and quality. Among these diseases, leaf diseases such as soybean rust, Septoria leaf spot, and bacterial blight are particularly detrimental, causing substantial economic losses to farmers and posing threats to global food security. Efficient approaches to disease management such as, crop rotation, chemical medication, and plant propagation for resistance. Mannual methods methods of disease identification rely on eye's r by agronomists and laboratory-based techniques such as microscopy and pathogen isolation, a process that can demand significant time, manual work and individual interpretation.

In deep learning techniques for automated disease diagnosis in agriculture. These methods offer the potential to revolutionize disease detection by analyzing large-scale image datasets and extracting discriminative features from plant images with unprecedented accuracy and efficiency. Deep learning, in particular, has achieved significant happening in a range of computer imagery applications such as, including object recognition, segmentation, and classification for automated plant disease diagnosis. The objective of this thesis is to use a deep learning-based classification system for identifying soybean leaf diseases from digital images. By leveraging deep learning architectures and techniques, the proposed system aims to accurately classify soybean leaf images into different disease categories, enabling rapid and reliable diagnosis in agricultural settings. Through empirical evaluation and validation using real-world datasets, this research seeks to demonstrate the effectiveness and practical applicability of deep learning in soybean disease management, ultimately contributing to the sustainability and resilience of soybean production systems.

With the advancement of computational systems, more computer vision approaches have been applied to plant disease classification. Agricultural prediction and categorization problems (Y. LeCun n.d.) are progressively being tackle with the use of artificial powerfulness and data mining method. Image process know-how were employed by multiple groups of person to recognize species and diagnose diseases by recover the Region of Interest. (ROI). Decision Tree (DT), k-Nearest Neighbor , Support Vector Machine (SVM), utilized, and their performances were plumbed using tenfold cross-validation (Q. Rao and J. Frtunikj n.d.).The recent emergence of deep learning within the realm of machine learning (ML) has yielded groundbreaking advancements in a variety of scientific disciplines, including computer vision, drug design, and bioinformatics. Without the need for manually constructed features, raw data. The past few years have witnessed rapid growth in deep learning-related methods in the field of computer vision (H. Sun and R. Grishman n.d.). In this inquiry I will give "Method for Soybean simplicity detection from Pakistani. We will develop the Soybean dataset with the help of Agriculture university, Faisalabad sub-campus okara's pathologist.

However, despite the potential of deep learning in plant disease diagnosis, there are challenges to overcome, including the availability of annotated datasets, model generalization across different environments and disease severities, and interpretability of model predictions. Tackling these issues calls for collaborative efforts between different fields of plant pathologists, computer scientists, and agronomists to build a slid and reliable deep learning-based classification systems for soybean leaf diseases. By leveraging the potential of technology, it is possible to improve disease management practices, ensure effective resource distribution and ultimately enhance soybean productivity and food security globally.

This study addresses the gap of lightweight and locally relevant soybean disease detection systems by proposing a MobileNetV2-based classification model trained on a dataset collected from Pakistani agricultural fields. The main contributions are:

i. Development of a region-specific soybean disease dataset.
ii. Adaptation of a lightweight CNN for efficient classification.
iii. Comprehensive evaluation using multiple performance metrics.
iv. Analysis of model suitability for deployment in resource-constrained environments.

The organization of this material is as follows: Section 2 explores prior research, Section 3 outlines the experimental framework and methodological approach, Section 4 presents the findings along with analysis, and Section 5 delivers concluding insights and suggests opportunities for future work, followed by the reference list.

2. LITERATURE REVIEW

Current and emerging applications of CNNs and deep learning architecture in farming are discussed here. Image process and machine basic cognitive process approaches were employed to categories plant diseases before more in-depth research was conducted. The following actions are typical for such systems: For starters, a digital camera is used to take a picture. Prior to advanced analysis, an image typically undergoes several preprocessing operations, including enhancement, segmentation, filtering, and color space conversion. Once these steps are completed, the most relevant features are extracted and provided to a classifier for decision-making (H.-A. Joung n.d.). Thus, the effectiveness of the classification process is strongly influenced by the selected preprocessing and feature extraction strategies. Recent literature highlights that models trained on publicly available datasets are capable of achieving progressive performance. In comparison, Convolutional Neural Networks (CNNs) represent a deep supervised architecture that independent learns hierarchical features directly from raw data.

CNN has surpassed the progressive in nearly all important categorization challenges in recent years. Its architecture allows for simultaneous feature extraction and categorization (D. Koundal n.d.). technology such as CNN, AV, SP, etc. Many researchers have continuously constructed Convolutional Neural Networks (CNNs), which are a specialized form of neural networks that has found widespread use in pattern recognition tasks. CNNs were initially motivated by the work of (Hubel and Wiesel) in 1962. For CNNs to be shift-, scale-, and distortion-invariant, their architects incorporate the following three concepts: the local receptive field, the general weight, and the spatial or temporal displacement (A. A. R. H. E. N. Hassan Ibrahim n.d.). A wide variety of CNN models have been designed and applied in the domain of object recognition. There's the LeNet, the AlexNet, the GoogLeNet, and so on. LeNet was the first convolutional neural network (CNN) architecture designed by LeCun et al. (A. A. R. H. E. N. Hassan Ibrahim n.d.) for recognizing handwritten digits. After fully integrating MLP, it features two convolutional layers and two additional sub-layers. Convolutional neural networks (CNN) have been proposed by certain researchers for use in leaf recognition and plant disease categorization. In order to recognize plants from pictures of their leaves, Atabay devised a convolutional neural network design. There are five distinct components to the proposed architecture.

After each pool layer, we use either the Regenerative stimulation functions such as ReLU (Rectified Linear Unit) and ELU (Exponential Linear Unit), combined with the Max-Pooling layer. In total, 32 plant species and 1,907 samples were analyzed in the Flavia (S. Pandiaraj, "Investigation of smart methodologies in Epidermodysplasia Verruciformis detection," n.d.) and Swedish leaf databases using the suggested approach. Images of individual leaves captured against a white background make up the database's contents. All input pictures are grayscale and 160 pixels square. Classification accuracy for the model was 97.24%, while accuracy for both data sets was 99.11%. Comparing the suggested architecture for leaf classification tasks based on CNNs, where cutting-edge and integrated methods are applied for feature construction and classification of leaves reveals promising results. In their investigation of complex systems, Angie K. Reyes et al. (J. Vendrow n.d.) employ deep learning techniques that eliminate the need for manually generated parts.

The convolutional neural network (CNN) was taught using 1.8 million photos from the 2012 ILSVRC dataset, and it employed a finite element technique to transfer recognition learning from the broad domain to the narrow task of detecting vegetation. A dataset consists of photographs of plants or plant parts collected in the wild or under controlled conditions. A 0.486 accuracies were the mean they achieved. Sarada P. Mohanty et al. classified plant illnesses using preexisting CNN architectures as Alex Net and Google Net (S. I. Hossain n.d.). CNN training involved the use of damaged plant images in order to recognize 14 crop species and 26 distinct diseases and healthy plant leaves from a publicly available dataset of photographs collected under controlled conditions. (or absence). Overall, the model was accurate to the tune of 99.35%. However, this model's accuracy plummeted to 31.4% when evaluated on a batch of photographs captured in operating under environmental conditions different from the training dataset. General, the breadth of CNN demonstrates the possibility of plant disease classification. Uuz et al. proposed the use of a Deep Convolutional Neural Network (DCNN) to detect olive peacock spot and aculeus olearius diseases. In total, there are 3,400 samples of olive leaves, divided into three groups: healthy, olive peacock spots, and aculeus olearius illness. DCNN model is utilized as the basis for the VGG16 and VGG19 architectures. By applying the RMSProp optimizer in combination with Stochastic Gradient Descent (SGD), network performance is enhanced. Esgario et al. presented a CNN-based multi-objective network. Purpose: To quantify the biotic stress experienced by coffee leaves. Through the use of computational trials, a more precise and trustworthy detection system was created. With a multi-problem CNN network built on the ResNet50 framework, we were able to attain recorded 86.51% accuracy for severity assessment and 95.24% for biotic stress categorization. Thus, this approach is appropriate for detecting and measuring biotic stress in coffee farms. Escario et al. (D. Chumachenko n.d.) suggested using a Convolutional Neural Network (CNN) that had been pre-trained to identify plant illnesses.

DenseNet201, ResNet101, Google Net, VGG19, Visual Pure Mathematics Group 16 (VGG16), and Alex Net are all examples of pre-trained CNN models. The authors chose four species that were affected by ten diseases: leaf spot, Yellow Vein Mosaic Virus, Citrus Canker, Orange Indian Ant, Brownish Spot, Cercospora Leaf Spot, Two Months, Epilachna Blight, Small Leaf Blight, and Tobacco Mosaic Virus (TMV). Prediction scores and picture classifications were continuously assessed for each disease category. A Bayesian deep learning approach was proposed by Hernández et al. (S. Akbarian n.d.) to identify plant diseases. In terms of classification performance, Bayesian reasoning outperforms state-of-the-art approaches. The posterior density function is evaluated with respect to plant disease identification, and the prediction uncertainty is quantified. Chen et al. first proposed using leveraging DCNN transfer learning in order to identify plant leaf blight. For evaluation purposes, they choose Inception and VGGNet modules that had already been trained on ImageNet. Experiment results showed that 92.00% accuracy was achieved while classifying photos of rice plants. Extraction of high-dimensional features for categorization allows for improved experimental outcomes in terms of both the user's personal data and the public data set. This allowed for a disease categorization confirmation accuracy of 91.83 percent. A CNN-

based approach was recommended by Lee et al. (D. Jacob n.d.) to identify Ginkgo leaf diseases.
The models Inception V3 and VGGNet-16 are used together. Accuracy of 98.44% was attained in controlled laboratory settings, with the VGG model providing an additional 92.19 percent. But as processing time grows longer, data redundancy becomes more of a problem. An accuracy of 96.19 percent was achieved by the proposed categorization model. Table.1 explains the brief summary of literature review.

Table 1. Summary of Related Work

Remark No	Research Methodological analysis	Illness Work	Dataset Collect	Accuracy %
[62]	R-CNN	Virus disease Frogeye leaf spot Bacterial null	synthetic soybean leaf disease image dataset	83.34
[63]	CNN-SVM	Grade 0 Grade 1 Grade 2	Hyper spectral images of a soybean leaf sample	96.67
[64]	HDL-AOA	Beanhalo blightPythium, Rhizoctonia rootrot, Anthracnose	SoybeanLarge DataSet	98.23
[65]	Alex Net and Google Net CNNs	bacterialblight, frogeyeleaf spot (FLS), brown spot,	Maharashtra, India	98.75
[66]	(CNN) model	soybean leaf diseases, soybean pod, soybean flower	ICAR (Indian Institute - Soybean Research)	99.56
[67]	PCR (QPCR) method	Brown spot Bacterial blight	USDA National Institute of Food and Agriculture	98
[68]	PCR	Fusarium root rot, Fusarium sporotrichioides, Amino acid max	root rot were collected from fields in Manitoba, Canada, in 2017	
[69]	spectral analysis and linear discriminant analysis (LDA)	Phakopsora pachyrhizi (Asian soybean rust)	greenhouse State University	93
[70]	CNN model VGG19, Google LeNet, Dense121, Xception Net, LeNet, and ResNet50	Anthracnose, Bacterial Blight, Soybean Mosaic Virus, Phytophora Rot, Rust, and Brown Spot	Image Database of Plant Disease Symptoms	98.14
[71]	CNN	Healthy Leaf Septorial leaf blight	The dataset for training is downloaded	99.32

		Frogeye leaf spot Downy Mildew	fromPlantVillage (Hughes, Salathe, and others 2015)	

3-Deep Transfer Learning Methodology

Artificial Intelligence is rooted in self-learning algorithms. As information is more forthcoming with regard to artifacts, the algorithms continuously evolve (S. Pandiaraj, "Investigation of smart methodologies in Epidermodysplasia Verruciformis detection," in 2021 International Conference on Computer Communication and Informatics (ICCCI). IEEE, n.d.). This method for addressing the dilemma continues to improve and adapt. Having been inspired by the human brain, this is a framework that can have self-learning algorithms used with it. Trying to imitative the way that human nerve cell communicates, Artificial Neural Networks (ANNs) are sets of inter-connected nodes. Data is stored, fed via positive or negative weights (in effect, both depending on interaction), and then output in this neural network. By their multi-layered construction and ability to perceive patterns, ANNs have a great deal of promise. The definition "deep learning" is taken here to identify the process that occurs in ANNs (F. A. Radtke n.d.) is used to depict the type of learning that these networks are capable of.

In order to categories coal root, numerous pests, and mosaic virus attacks, this research introduces a deep transfer learning method. The pre-processing and multiplication method is used to address class inequality in the primary database and to generate a more representative sample. The proposed system performs automated classification of soybean leaf diseases using deep transfer learning involves automatic feature extraction followed by the implementation of the pre-trained MobileNetV2 model. The process flow for the suggested solution is depicted Fig. 1. The model was trained using the Adam optimizer with a learning rate of 0.001. Images were normalized to the range [0,1]. Batch size was set to 16, and training was conducted for 6 epochs

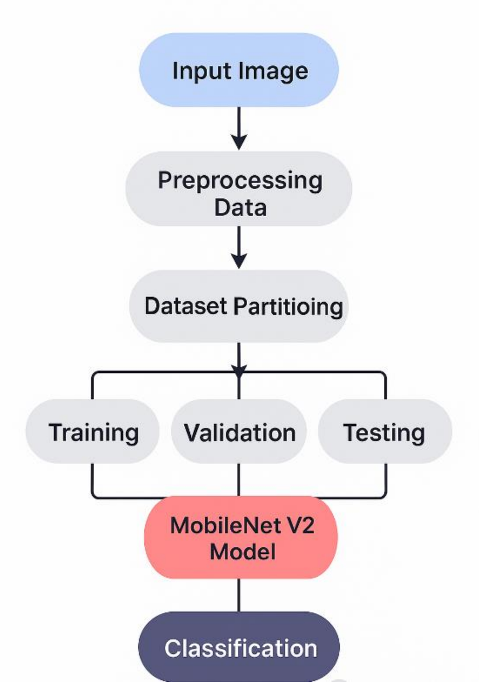

Figure 1. Proposed method flowchart

3.1 Data Preparation

Deep learning model used to perform dataset operations on which tuned for the purpose of accuracy. Soybean lead disease data set was generated by researchers from University of Okara and the University of Agriculture in Faisalabad, Okara. Table.2 shows an across-the-board breakdown of the database.

Table 2. Summary of SDD-2025 dataset

No	Disease	Size
1	Charcoal Root	666
2	Mosaic Virus	666
3	Pest Infestation	668

3.2 SDD 2025 Dataset

Several organizations contributed data from various soybean' fields of different diseases to the SDD-2025. The dataset includes 2000 disease images. We divide all these datasets into three classes: charcoal root, Mosaic virus, and Pest Infestation. The images belonging to different soybean disease categories were randomly selected from the dataset. Fig.2 illustrate the images of disesease are being disucessed in this study.

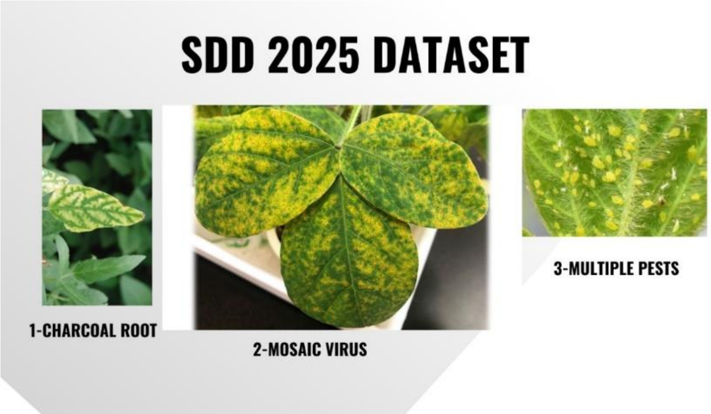

Figure 2. Images of Diseases

3. Image Preprocessing and Data Augmentation

To ensure high consistency of categorization results and strong feature extraction, a comprehensive image preprocessing was performed for all SDD-2025 input images, with multiple data augmentation techniques applied such as rotation of the input images (0.8×-1.2× scale), horizontal flipping, zooming, shearing and brightness adjustments that simulate real-world field variations and enhance the model's ability to generalize, while reducing the risk of overfitting the relatively small dataset of 2000 images. These transformations artificially increase the training set by creating various variations of each image, enabling the MobileNetV2 model to learn invariant features that are essential in defining the subtle disease pattern in the images from the three disease classes, Charcoal Rot, Mosaic Virus and Pest Infestation, under different lighting conditions, angles, and scales found in field photographs relevant to agriculture. In particular, rotation and flipping eliminate orientation differences, zooming and shearing correct scale and perspective differences, and brightness correction corrects inconsistencies in illumination of the field, which, combined, boost training data diversity by 5-10× without needing to collect more images. The augmentation strategy was applied to the training data and not to the validation or testing data, which helps regularize the model, avoids memorization of the training samples and has a high impact on the test accuracy (96.14%) and improves its ability to perform well on a large number of images from the field that the model had not seen during training, which is crucial for the deployment environment of the model that relies on limited resources.

3.4 Image Resizing and Standardization

The original SDD-2025 dataset contains all images at high resolution (6000 × 4000 pixels), which is great for the detail of the diseases being visualized in the images but does create a high computational problem when training deep learning networks. All images were preprocessed to be 256 × 256 pixels to ensure efficient training of the models, as it would also be compatible with the MobileNetV2 architecture that was originally designed to process images of only 224 × 224 pixels. To reduce memory usage, around 98.7% of the images were downsized to 256 × 256 pixels, capturing essential disease features like texture patterns, lesion shapes, and color variations. This resizing has multiple benefits: to increase the computational efficiency for hardware with limited resources, to speed up the convergence of training over the 6 epochs, to provide uniform dimensions to the input across the entire dataset, and to match the input size of 224-256 pixels that is optimized by MobileNetV2. Despite the loss of fine detail, empirical results have shown that the resolution maintained the discriminative information that is useful for correctly classifying Charcoal Rot, Mosaic Virus and Pest Infestation with a reported accuracy of 96.14%, which allows for the deployment of the model in edge devices for real-time monitoring of agricultural environments.

3.5 Dataset Splitting and Evaluation Protocol

To develop a robust model and evaluate its performance with the MobileNetV2 architecture objectively, the SDD-2025 database containing 2000 soybean leaf disease images was systematically divided into three parts, with a ratio of 70:15:15.The SDD-2025 dataset with 2000 pictures of soybean leaf disease was divided into three parts, 1400 for training, 300 for validation, and 300 for testing, in a systematic manner to develop a robust model and unbiased evaluation with the MobileNetV2 architecture. The model parameters were optimized using the training set (70%) and hyperparameters were tuned, early stopped and overfitting was monitored throughout the 6-epoch training process using Adam optimizer with an independent held-out validation set (15%). The testing set (15%) was not seen at all during training, and only used for the final evaluation to estimate the model's performance in the real world in new soybean disease images from a Pakistani agricultural field. It is a stratified split that maintains near-equal class balance across the Charcoal Rot, Mosaic Virus, and Pest Infestation categories (roughly 466/100/100 per category in train/val/test) with little or no biased learning towards the most prevalent categories thus ensuring statistical reliability of the performance measures. Only the training set was augmented in each epoch, whereas data was resized only to 256×256 pixels in the validation and test sets, which better simulate the realistic situation of a smart farming application under resource constraints, thus leading to the quoted accuracy of 96.14% as a reliable performance indicator.

3.6 Architecture of MobileNetV2

A thorough analysis of the MobileNetV2 structure (L. Yu, "Automated melanoma recognition in dermoscopy images via very deep residual networks," n.d.) is under way as part of the ongoing research project to tackle the problem of soybean disease classification. The choice of employing a MobileNetV2 model is stuck between a range of different considerations. The training data set used for a model is very small, which can cause overfitting. A way to mitigate this effect is through the use of a smaller but more efficient network, like MobileNetV2. MobileNetV2 is a design that optimizes the cost associated with errors while maximizing the rate of execution along with memory usage. The low memory consumption is also another attractive feature, aside from the fast execution rate, which increases the feasibility of parameterization and testing. The basic design of MobileNetV2 is based on that of the previous model, designated as MobileNetV1. Dissociable convolution and linear whirl that are deeply dissociable, as well as inverse residuals, will be discussed in the following section since they are two of the most profound concepts that define the MobileNetV2 architecture.

Other very efficient models that employ depth-wise partitioning are Xception and Shuffle Net, introduced. MobileNetV2 (EDWARD ZHANG n.d.) also employs the same depth-wise segmentation employed in MobileNetV1. Deep-resolvable convolution can be exchanged with a normal convolution in two steps. First, we apply a separable convolution individually on each of the feature maps in map-based convolution. The resulting feature map is then acted upon and represented by a second map. Pointwise convolution with a 1 x 1 kernel is computed in parallel across all maps here. Images are acted upon in a standard convolution along their height, width, and channel axes simultaneously, as illustrated in Figure 3. Instead, a separable convolution's initial process is a diligent examination of the height and breadth parameters. Channel dimensions are under its control during the second process, something that is attributed to the factorization aspect of the conventional process.

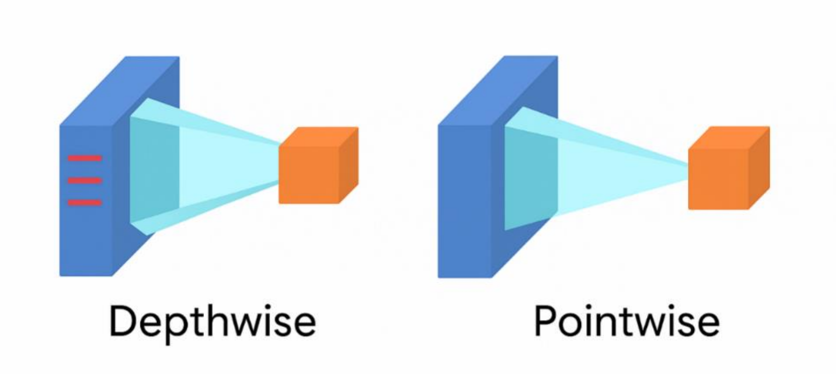

Figure 3. Traditional convolution and depth-wise separable convolution

The building pieces that make up the ResNet network are described and contrasted with inverted residues, using three convolution operators and impedance and residual coupling, both blocks achieve similar results. To convert information from the input domain to the intermediate representation and from the intermediate representation to the output domain. In order to filter the interval representation, as depicted in Figure 4, we apply a 3 x 3 filter (H. Yang and P. A. Bath n.d.) (L. Yu, “Automated melanoma recognition in dermoscopy images via very deep residual networks,” n.d.).

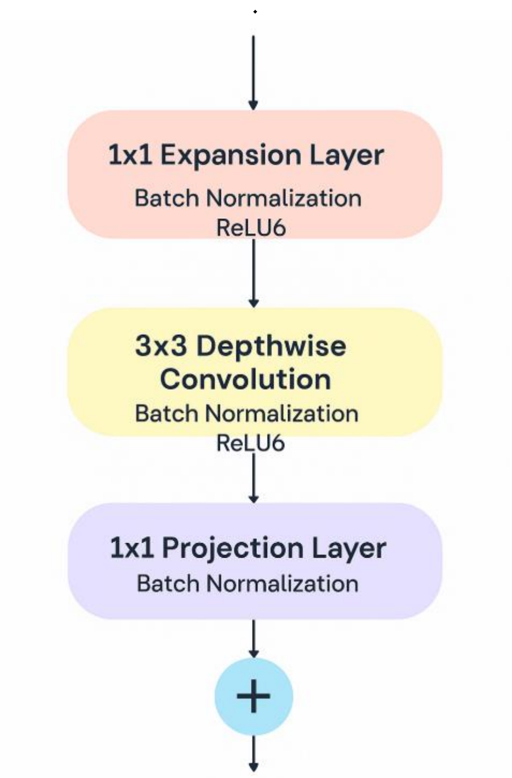

Figure 4. Bottleneck residual block

Intra-block convolution output fewer feature maps than the first and final matter block solutions, while reverse convolution's initial and final mixture yield the fewest feature maps. When comparing MobileNetV2 to ResNet, the residual property between the training and test feature maps (channels) is lower. Layer sizes can vary from modest to enormous when many units are joined in a design. Using MobileNetV2's residuary block settings helps cut down the network's memory footprint.

For each edifice block, MobileNet V2 features an extra depth-based and point-to-point conversion layer. Data with numerous channels can be changed into sensors with a small number of maps using the projection layer, a new improver to MobileNetV2 that wasn't present in V1. (channels). The system's block that stores the results of each individual block. After pass through a 1 x 1 diffusive convolution layer, the number of feature mappings (channels) will rise proportionately to the spreading factor used in the ensuant depth-based convolution. The second major addition to the MobileNetV2 architecture is the concept of residual connectivity. A system's gradient flow is aided by the constitution of residual connections. In MobileNetV2, we use ReLU6 as the activation mathematical function for every tier of the architecture to ensure that our models are consistently robust. The projection layer's output, on the other hand, does not contain the activation function. MobileNetV2's full structure consists of 17 residual block blocks, a 1 x 1 sequential convolutional layer, a global average pool layer, and a classification layer. In Figure 5, we can see the finalized MobileNetV2. The pointwise convolution is a 1x1 convolution, which means that it has a kernel size of 1x1.

In this example, the input mental image x is first passed through two convolutional layers. The convolutional layers extract features from the image and reduce its size by half. The features extracted by the convolutional layers are then passed through a fully connected layer with 128 neurons. The fully connected layer combines the features extracted by the convolutional layers and uses them to classify the image into one of the three soybean disease classes. The class with the highest probability is the class that the image is classified as. Table 3 shows the model and parameters that gave the best results with an accuracy of 96.14 percent.

Table 3. Parameters

Parameters Used	Values of Parameters
Architecture	MobileNetV2
Transfer Learning	ImageNet Pretrained
All diseases shown	Sample images from the SDD-2025 soybean disease dataset
Optimizer	Adam
Learning Rate	0.001
Activation Function	ReLU + Softmax
Loss Function	Categorical Cross Entropy
Batch Size	16

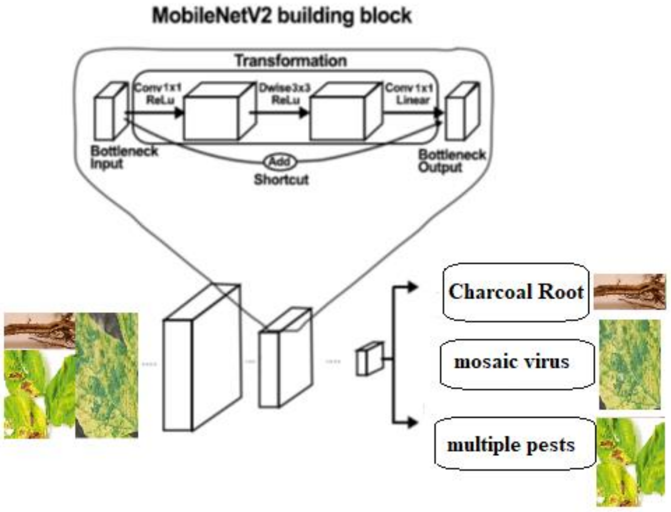

Figure 5. Classifier based on mobilenetv2

4. Results and Discussion

An experiment was run with the parameters provided in Table 3 to assess the efficiency of the newly proposed mobilenetv2 architecture: An Adam optimizer with a categorical cross-entropy is loss function over the course of 6 epochs, 16 batches, and the standard alpha values. The results of this new approach indicated a 96.14 percent success rate

in detecting foliar diseases in soybeans. In terms of soybean leaf blight, the transfer training model achieved a recall of 94% and an accuracy of 97%. The F1 ratings for the three disorders ranged between 94% and 96.14%. In Table 4, we can see that the gap between our test set's accuracy and the accuracy seen in the leaderboard tests for SDD-2025 is not large enough to warrant abandoning the suggested strategy. Due to the lack of readily available ground truth for SDD-2025, the test is embedded in The leaderboard is 2000 photos.

Area under the receiver operating characteristic curve (AUC), a measure developed by the Kaggle organizer. Figure 6 depicts the accuracy and loss at each time point during training and validation. It demonstrates that there is a rapid improvement in training and validation accuracy after the first session, and that this improvement is maintained until the fifth period. However, both training and validation losses quickly decline after the first period, and then level off after four. By applying the data augmentation strategy to the training set, the suggested method improves classification scores on the SDD-2025 database.

Table 4. Classification results and accuracy on sdd-2025 database

Presentation Measure #	Charcoal-grey Rot %	Mosaic Virus %	Multiple Pests %	Average Accuracy %	Leader Accuracy %
Accuracy %	96.14%	96.14%	96.14%	96%	96.14%
Recall	84%	99%	99%	-	
F1-Score	91%	91%	99%	-	
Precision	99%	83%	99%	-	

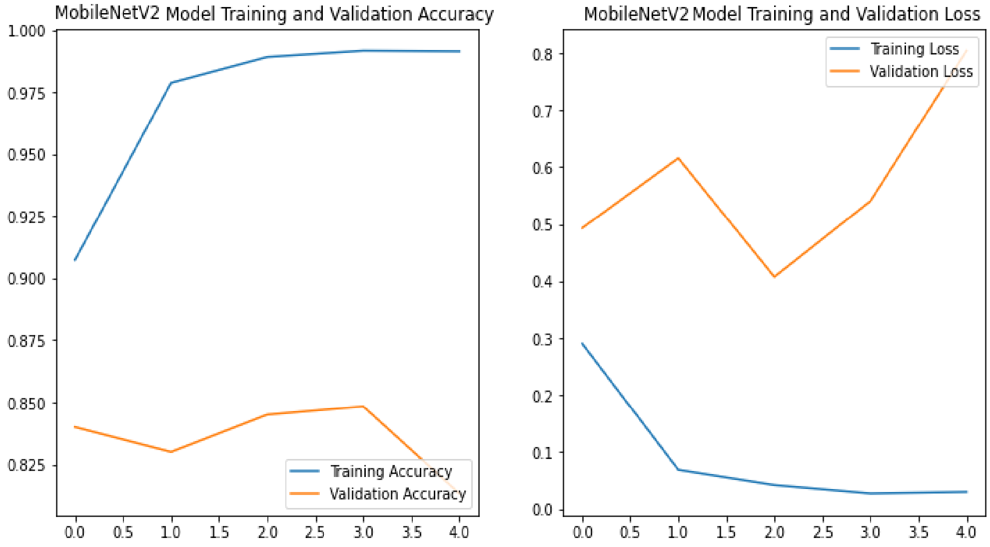

Figure 6 (a)Accuracy graph where (b) loss value graph of proposed model

Accuracy, preciseness, the ROK curve, and model correctness can be computed with the confusion matrix, a almighty ML tool. The disorder matrix was used to graphically assess the precision of the classification. Fig 6. Shows the accuracy gain and loss graph. This indicates that the categorization is more accurate in the case of MobileNetV2 than for the classes that are dark-colored, while the incorrectly predicted samples are the ones that are shown with light-colored representations. True predictions are signified along the diagonal in the confusion matrix, while false predictions are indicated along the other diagonal.

Fig.7 shows the confusion matrix when the feature based method is applied to the SDD-2025 dataset, the recommended MobileNetV2 subject performs better. The overall consistency of the suggested MobileNetV2 system is 96.14%, and the defect is 3.86% also the Iterative improvement of the several model prototype

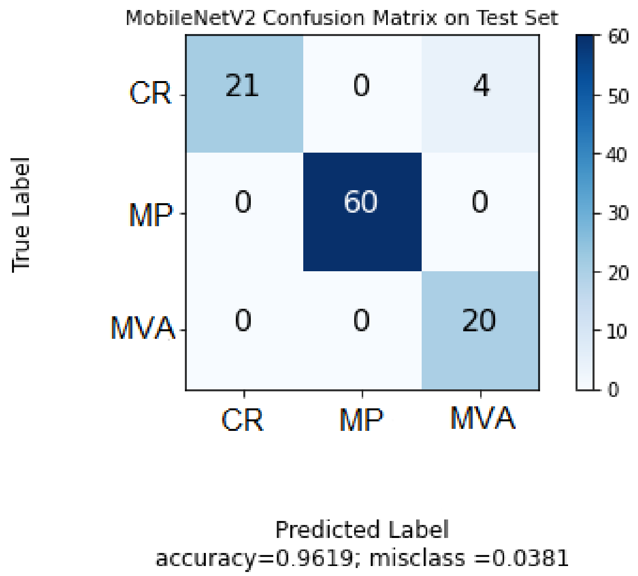

Figure 7. Confusion matrix of MobileNetV2

The MobileNetV2 model has a higher area nether the curve (about 96.14%) and attempted to promote success in the validation and test classes. Fig. 8 shows the ROC slope as a measuring of the efficaciousness of these up-to-date know-how, with black stand for the ROC Curve and red the being of outlier articulation.

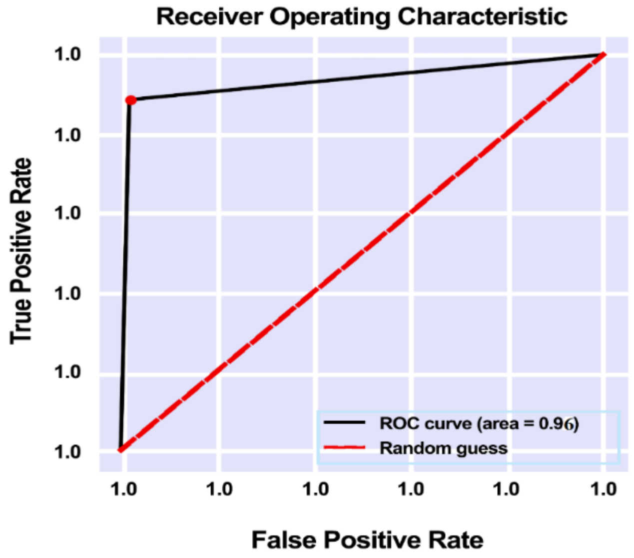

Figure 8. Proposed model ROC Curve on SDD-2025

When the feature activity strategy was applied to the training phase, measurement methods such as preciseness, F1-measure, recall, clarity, and ROC curve show that the proposed method conducted highly well in the SDD-2025 database.

5. Conclusion and Future Work

In this study, a lightweight deep learning approach with MobileNetV2 was introduced for detecting soybean leaf diseases. The proposed model was trained with images of soybeans with three disease categories in the SDD-2025 dataset: Charcoal Rot, Mosaic Virus, and Pest Infestation. The effectiveness of the proposed approach was verified through experiments, where the agricultural disease detection accuracy was successfully obtained as 96.14%, showing that the transfer learning and lightweight CNN networks can be effectively applied to the agricultural disease detection in resource-constrained environments. The developed framework offers an efficient and automated solution for early disease diagnosis of soybean without much human intervention after training. The proposed system can benefit farmers by timely detection of the disease, minimizing crop losses, disease management and soybean productivity, when compared to traditional manual inspection methods. Moreover, the light weight design of the model enables its application in real-time agricultural monitoring and mobile farm assistance system. Although the results are promising, there are a number of limitations. Data used were relatively small and from specific geographical areas; no external validation was done for different environments and soybean varieties. Generalization and prediction power of the model may be influenced by climatic variations, illumination, severity of the disease, and plant species variations. Future research will involve larger and more varied field images, more soybean diseases, and testing the model under field conditions. Additionally, explainable AI techniques, sophisticated data augmentation methods, and the implementation of the framework on edge and mobile devices for real-world smart farming applications could be further enhancements.

REFERENCES:

Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510–4520.

S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Frontiers in Plant Science, vol. 7, p. 1419, 2016.

D. P. Hughes and M. Salathé, “An open access repository of images on plant health to enable the development of mobile disease diagnostics,” arXiv:1511.08060, 2015.

K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

J. Too, S. Yujian, L. Njuki, and Y. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease identification,” Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.

A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.

J. Rashid, I. Khan, G. Ali, S. H. Almotiri, M. A. AlGhamdi, and K. Masood, “Multi-level deep learning model for potato leaf disease recognition,” IEEE Access, vol. 10, 2022.

J. Rashid, I. Khan, G. Ali, S. ur Rehman, F. Alturise, and T. Alkhalifah, “Real-time multiple guava leaf disease detection using hybrid deep learning,” Computers and Electronics in Agriculture, 2021.

S. Sladojevic et al., “Deep neural networks based recognition of plant diseases by leaf image classification,” Computational Intelligence and Neuroscience, 2016.

P. Kaur, S. H. S. Gill, and S. K. Saini, “Plant disease detection using deep learning: A review,” Archives of Computational Methods in Engineering, 2022.

H. Durmuş, E. O. Güneş, and M. Kırcı, “Disease detection on the leaves of the tomato plants by using deep learning,” in 2017 6th Int. Conf. Agro-Geoinformatics, pp. 1–5.

A. Brahimi, K. Boukhalfa, and A. Moussaoui, “Deep learning for tomato diseases: classification and symptoms visualization,” Applied Artificial Intelligence, 2017.

X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, "Identification of maize leaf diseases using improved deep convolutional neural networks," IEEE Access, 2018.

S. G. S. Patel, "Plant disease detection using CNN: A review," Journal of Artificial Intelligence Research, 2020.

Javeed, M. U., Aslam, S. M., Sadiqa, H. A., Raza, A., Iqbal, M. M., & Akram, M. (2025). Phishing website URL detection using a hybrid machine learning approach. Journal of Computing & Biomedical Informatics, 9(01).

Raza, A., Zongxin, S., Qiao, G., Javed, M., Bilal, M., Zuberi, H. H., & Mohsin, M. (2025). Automated classification of humpback whale calls in four regions using convolutional neural networks and multi scale deep feature aggregation (MSDFA). *Measurement*, *255*, 118038.

Raza, A., Javed, M., Fayad, A., & Khan, A. Y. (2023). Advanced deep learning-based predictive modelling for analyzing trends and performance metrics in stock market. *Journal of Accounting and Finance in Emerging Economies*, *9*(3), 277-294.

Nauman, M., Ashraf, S., Javeed, M. U., Aslam, S. M., Farooq, U., & Raza, A. (2025). Deep transfer learning for COVID-19 screening: Benchmarking ResNet50, VGG16, and GoogleNet on chest X-ray images. *International Journal of Advanced Computing & Emerging Technologies*, *1*(2), 69-83.

Raza, A., Javeed, M. U., Javed, M., Aslam, S. M., Saleem, R., Irshad, W., ... & Ejaz, G. (2025). A hybrid machine learning framework for personalized news recommendation. *International Journal of Advanced Computing & Emerging Technologies*, *1*(3), 49-62.

Raza, A., Javeed, M. U., Arshad, M. A., Akram, W., Aslam, S. M., Ejaz, G., ... & Shaukat, A. (2025). Intelligent image gallery system using deep learning for automated fruit and vegetable classification. *International Journal of Advanced Computing & Emerging Technologies*, *1*(3), 1-15.

Javeed, M. U., Aslam, S. M., Aslam, S., Iqbal, M., Farhan, M., Javed, M., & Raza, A. (2025). Unveiling ambivalence in reviews: Using Sentence-BERT and K-NN for airline recommendations. *Technical Journal*, *30*(03), 51-59.

Sankoh, A. P., Raza, A., Parwez, K., Shishah, W., Alharbi, A., Javed, M., & Bilal, M. (2025). Automated Facial Pain Assessment Using Dual-Attention CNN with Medical-Grade Calibration and Reproducibility Framework.

Raza, A., Javed, M., Aslam, S. M., Javeed, M. U., Samual, P., Raheem, S., ... & Arshad, M. A. (2025). AI-powered sentiment analysis for social media opinion mining: A hybrid NLP and machine learning approach. *International Journal of Advanced Computing & Emerging Technologies*, *1*(5), 1-15.

Sankoh, A. P., Raza, A., Parwez, K., Shishah, W., Alharbi, A., Javed, M., & Bilal, M. (2026). Automated Facial Pain Assessment Using Dual-Attention CNN with Clinically Calibrated High-Reliability and Reproducibility Framework. *Biomimetics*, *11*(1), 51.

Raza, A., Qiao, G., Li, L., Javed, M., Rakha, M. A., Bilal, M., ... & Javed, S. (2025, December). A Deep Learning Framework for Imbalanced Bioacoustic Call-Type Classification Using Focal Loss and SMOTE. In *2025 International Conference on Frontiers of Information Technology (FIT)* (pp. 1-6). IEEE.

Parwez, K., Sohail, S. I., Bilal, M., Shishah, W., Raza, A., & Javed, M. (2026). A Trust-Centered Explainable Deep-Learning Framework for Acute Lymphoblastic Leukemia Detection Using Multi-Model Fusion and Interpretability Scoring. *Algorithms*, *19*(2), 162.

Parwez, K., Sohail, S. I., Raza, A., & Zia, M. A. (2026). LeukoXAI-Lite: A reusable explainable AI toolkit for federated leukemia diagnosis with visual explanations and performance analysis. *Software Impacts*, 100813.

Muhammad Usman Javeed, “A DEEP LEARNING APPROACH FOR SECURING IOT SYSTEMS WITH CNN-BASED PREDICTION OF WORST-CASE RESPONSE TIME”, SES, vol. 3, no. 7, pp. 376–385, Jul. 2025.

H. Shakeel et al., “LncRNAs disease: A text mining approach to find the role of lncRNA in aging,” *J. Comput. Biomed. Informatics*, vol. 9, no. 1, 2025. [Online]. Available: https://www.jcbi.org/index.php/Main/article/view/1000

S. Saifullah, M. U. Javeed, A. Ayub, M. M. Zahoor, Z. Raza, S. M. Aslam, et al., “Impacts of cloudburst events on biodiversity in Pakistan’s northern areas and development of a machine learning model for cloudburst prediction,” Int. J. Adv. Comput. Emerg. Technol., vol. 1, no. 4, pp. 44–63, Oct. 2025, doi: 10.64796/y82wy489.

“An Evaluation of Machine Learning Classifiers for Nominal Weather Humidity Prediction,” *Int. J. Adv. Comput. Emerg. Technol.*, vol. 1, no. 4, pp. 1–21, 2025, doi: 10.64796/r3m9r804.

M. U. Javeed, M. Nauman, S. Aslam, S. M. Aslam, M. M. Zahoor, Z. Raza, *et al.*, “Deep learning in hematology: Automated counting of blood cells using YOLOv5 object detection,” *Int. J. Adv. Comput. Emerg. Technol.*, vol. 1, no. 3, pp. 16–34, 2025. [Online]. Available: https://ijacet.com/index.php/IJACET/article/view/21.

M. U. Javeed, M. Akram, S. M. Aslam, S. Hussain, M. M. Nazakat, and M. Nauman, “Predicting customer loyalty from e-commerce reviews using aspect-based sentiment analysis and ANN,” Asian Bull. Big Data Manag., vol. 5, no. 3, pp. 49–61, 2025, doi: 10.62019/3akt6733.

“An ANN Based Intelligent System for Measuring Customer Loyalty”, KJCIS, vol. 8, no. 1, Jul. 2025, doi: 10.51153/kjcis.v8i1.250.

Muhammad Usman Javeed, Sammar Fatima, Mirza Mumtaz Zahoor, Muhammad Azhar, Zeeshan Raza, Shafqat Maria Aslam, & Muhammad Nauman. (2025). MindMate: An Emotion-Aware Generative AI System for Personalized Mental Health Support. Journal of Computing & Biomedical Informatics. Retrieved from https://jcbi.org/index.php/Main/article/view/1167

M. Aleem, M. U. Javeed, A. Shaheen, W. Ullah, S. M. Aslam, and H. M. J. Haider, “A Machine Learning-Based Framework for Pre-Deployment Prediction of Mobile Application Success,” Int. J. Adv. Comput. Emerg. Technol., vol. 1, no. 5, pp. 28–54, Nov. 2025. [Online]. Available: https://doi.org/10.64796/sqxa6s02.

“Optimized Deep Convolutional Neural Network for Robust Occluded Facial Expression Recognition,” *Asian Bull. Big Data Manag.*, vol. 5, no. 3, pp. 62–80, 2025. doi: 10.62019/dpfhnf43.

Maryam Zaffar, “A NOVEL FEATURE SELECTION METHOD FOR PREDICTION OF FACTORS AFFECTING ANXIETY IN HEALTH CARE WORKERS DURING PANDEMIC”, SES, vol. 4, no. 2, pp. 331–353, Feb. 2026.

“Conversational AI in Healthcare: An Exploration of Large Language Models for Patient Support and Communication”, IJACET, vol. 1, no. 5, pp. 65–79, Nov. 2025, doi: 10.64796/2yv51w38.