International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(17s) 2026 pp. 1073-1097
https://doi.org/10.70917/ijcisim-2026-4818
Article

An Efficient Deep Feature Fusion-Based Convolutional Neural Network for Multi-Class Millet Classification

Rajasekaran Saminathan[1], Rajat Verma[2], Ashwini Lokhande[3], S.Rudhra[4], B.Somasekhar[5], Sameer Tembhurney[6]

[1]Department of Mechanical Engineering, Assistant Professor, College of Engineering and Computer Science, Jazan University, Jazan, Jazan Province, Saudi Arabia – 45142, 0000-0003-0284-1818
(Corresponding Author): vaagaipaper@gmail.com
[2]Department of Computer science and engineering, Associate Professor and Additional Head, Pranveer Singh Institute of Technology, Kanpur, Uttar Pradesh, https://orcid.org/0000-0001-5707-2695, India
[3]Department of Computer Science and Engineering, Assistant Professor, Ramdeobaba University, Nagpur, 0009-0001-7746-110X
[4]Department of Electrical and Electronics Engineering, Assistant Professor, Jerusalem College of Engineering, Chennai -600100, Tamil Nadu, India
[5]Department of ECE, Professor, Anil Neerukonda Institute of Technology and Sciences, ANITS, Visakhapatnam, 531162, Andhra Pradesh, India
[6]Department ofComputer Science and Engineering, Assistant Professor, Ramdeobaba University, Nagpur, Maharashtra, 440013, https://orcid.org/0009-0008-6712-0804, India

Abstract: Accurate identification of millet varieties was essential for improving food quality assessment, agricultural productivity, post-harvest management, and intelligent grain processing systems. Conventional manual classification methods are often time-consuming, dependent on expert knowledge, and affected by visual similarities among different millet categories. This research proposes an efficient Deep Feature Fusion-Based Convolutional Neural Network (DFF-CNN) framework for automated multi-class millet classification using digital grain images. The developed approach integrates hierarchical feature extraction and multi-level feature fusion to combine low-level textural characteristics with high-level semantic representations, enabling improved discrimination among visually similar millet varieties. The image dataset consisting of multiple millet classes was subjected to preprocessing operations including resizing, normalization, noise reduction, and augmentation to enhance image quality and model robustness. The proposed architecture utilizes convolutional feature extraction, feature fusion, attention-based refinement, global average pooling, and Softmax-based classification to achieve accurate recognition. The performance of the developed model was evaluated using multiple evaluation parameters, including accuracy, precision, recall, F1-score, specificity, Matthews Correlation Coefficient, Cohen's Kappa, confusion matrix, and ROC-AUC analysis. Experimental results demonstrate that the proposed framework achieves superior classification capability with high prediction reliability and strong generalization performance compared with conventional deep learning architectures. The generated feature representations effectively capture important grain characteristics such as texture patterns, morphological structures, and surface variations, resulting in reduced classification errors among millet categories. The proposed intelligent classification framework provides a scalable solution for automated grain inspection, food authentication, precision agriculture, and smart agricultural decision-support applications.
Keywords: Deep Feature Fusion; Convolutional Neural Network; Millet Classification; Agricultural Image Processing; Deep Learning; Precision Agriculture

1. Introduction

Agricultural production plays a fundamental role in ensuring global food security, nutritional sustainability, and rural economic development. Among cereal crops, millets have gained considerable attention because of their exceptional nutritional composition, climate resilience, and adaptability to marginal farming conditions. Millets contain high concentrations of dietary fiber, proteins, essential minerals, antioxidants, and micronutrients, making them valuable components of balanced diets and functional foods. Increasing awareness of healthy food consumption has accelerated the global demand for millet-based products, leading to expanded cultivation and commercialization across many countries [1]. Finger millet, pearl millet, foxtail millet, little millet, barnyard millet, kodo millet, proso millet, and sorghum represent major millet varieties that contribute significantly to food systems in Asia and Africa. Accurate identification of these millet varieties remains an essential requirement throughout cultivation, harvesting, storage, processing,

packaging, and marketing. Visual similarities among different millet grains often create substantial challenges for manual inspection, particularly when grain size, color, texture, and shape exhibit only subtle variations. Human inspection depends heavily on domain expertise, environmental lighting conditions, and operator experience, resulting in inconsistencies, reduced efficiency, and increased processing time. Large-scale grain processing industries require automated classification systems capable of delivering rapid, reliable, and consistent identification without subjective judgment. Automated millet classification supports quality assurance, minimizes adulteration, improves traceability, enhances market value, and facilitates precision agriculture through intelligent decision-support systems [2]. Growing investments in smart agriculture, artificial intelligence, and digital food inspection have accelerated research toward computer vision-based grain recognition technologies capable of addressing these industrial challenges with greater accuracy and operational efficiency.

Traditional grain classification systems have primarily relied on handcrafted feature extraction methods combined with conventional machine learning algorithms for pattern recognition. Such approaches generally involve multiple sequential stages including image acquisition, preprocessing, segmentation, feature extraction, feature selection, and classification. Shape descriptors, color histograms, geometric measurements, statistical texture features, Gray Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and wavelet-based descriptors have been widely adopted to characterize cereal grain morphology [3]. Classifiers such as Support Vector Machines, Random Forests, Decision Trees, K-Nearest Neighbors, Artificial Neural Networks, and Naïve Bayes have demonstrated satisfactory performance under controlled laboratory environments. Classification accuracy often deteriorates when image backgrounds, illumination conditions, grain orientation, occlusion, or intra-class variability become increasingly complex. Manual feature engineering demands substantial domain expertise and frequently overlooks subtle discriminative characteristics required for distinguishing visually similar millet varieties. Independent optimization of preprocessing and feature extraction stages also increases computational complexity while limiting adaptability across diverse datasets. Performance degradation becomes particularly evident when large-scale datasets exhibit significant variation in grain appearance caused by cultivation practices, maturity levels, storage conditions, or imaging equipment. These limitations have motivated researchers to investigate automated representation learning techniques capable of extracting hierarchical and highly discriminative image features without extensive manual intervention [4]. Deep learning has consequently emerged as a transformative paradigm for agricultural image analysis because of its capability to learn complex visual representations directly from raw image data through end-to-end optimization.

Recent advances in deep learning have significantly transformed computer vision applications across agriculture by enabling automatic feature learning from complex image datasets. Convolutional Neural Networks have demonstrated remarkable success in crop classification, plant disease diagnosis, fruit recognition, weed detection, seed quality assessment, and food authenticity verification. Architectures including AlexNet, VGGNet, GoogLeNet, ResNet, DenseNet, MobileNet, EfficientNet, Inception, and ConvNeXt have consistently achieved superior performance compared with conventional machine learning approaches owing to their capability to learn hierarchical spatial representations from large-scale image collections [5]. Lower convolutional layers generally capture fundamental visual patterns including edges, corners, gradients, and textures, whereas deeper layers encode increasingly abstract semantic information associated with object identity and structural characteristics. Agricultural image datasets frequently contain subtle visual differences that require simultaneous exploitation of both low-level texture information and high-level semantic representations. Standard CNN architectures primarily utilize the deepest feature maps for final prediction, reducing the contribution of informative intermediate representations generated throughout earlier network layers. Such information loss becomes particularly critical during fine-grained classification tasks involving visually similar millet varieties that differ only through minor texture variations, surface roughness, grain curvature, or color intensity [6]. Recent research has therefore emphasized multi-scale learning, attention mechanisms, residual connectivity, dense feature propagation, and hierarchical feature aggregation to improve discriminative capability. Integrating complementary information from multiple feature levels enables deep learning models to achieve stronger generalization, enhanced robustness against environmental variability, and improved classification accuracy across challenging agricultural imaging scenarios.

Deep feature fusion has emerged as an effective strategy for combining complementary representations extracted from different convolutional stages to maximize classification performance. Feature fusion techniques integrate shallow texture descriptors with deeper semantic information, allowing neural networks to exploit both local structural characteristics and global contextual knowledge during decision making. Multi-branch convolutional architectures, feature pyramid networks, skip connections, channel attention mechanisms, spatial attention modules, squeeze-and-excitation blocks, concatenation layers, weighted fusion strategies, and transformer-assisted aggregation have demonstrated promising improvements across numerous fine-grained visual recognition problems [7]. Agricultural applications involving seed identification present unique challenges because numerous grain varieties exhibit nearly identical geometric structures while differing only through subtle variations in texture, color distribution, and surface morphology. Efficient feature fusion enables stronger representation learning by preserving discriminative information generated across different receptive fields and hierarchical depths. Data augmentation techniques including rotation, scaling, brightness variation, flipping, random cropping, contrast adjustment, and noise injection further enhance model robustness by increasing sample diversity and reducing overfitting during network training [8]. Explainable Artificial Intelligence techniques including Gradient-weighted Class Activation Mapping provide additional confidence by highlighting image regions contributing to classification decisions, thereby improving model transparency and supporting practical deployment within agricultural industries. Efficient network design simultaneously addresses computational complexity, memory consumption, inference speed, and deployment feasibility for resource-constrained environments including embedded agricultural devices, portable inspection systems, and intelligent sorting machinery operating under real-time conditions.

The present research introduces an efficient Deep Feature Fusion-Based Convolutional Neural Network designed specifically for multi-class millet classification using high-resolution grain images. The proposed framework integrates parallel convolutional feature extraction modules with hierarchical feature fusion mechanisms to preserve discriminative information captured across multiple abstraction levels. Batch normalization, nonlinear activation functions, global average pooling, dropout regularization, and fully connected classification layers collectively enhance feature learning while reducing overfitting and improving generalization performance. Extensive

image preprocessing and augmentation strategies strengthen robustness against variations in illumination, grain orientation, imaging distance, and background complexity encountered during practical agricultural operations [9]. Experimental evaluation considers multiple millet categories including finger millet, pearl millet, foxtail millet, barnyard millet, little millet, kodo millet, proso millet, and sorghum to demonstrate comprehensive multi-class recognition capability. Model performance assessment employs diverse quantitative metrics including accuracy, precision, recall, F1-score, specificity, Matthews Correlation Coefficient, Cohen's Kappa coefficient, Receiver Operating Characteristic analysis, Area Under the Curve, confusion matrix evaluation, inference time, computational complexity, and model efficiency. Comparative experiments against established deep learning architectures validate the effectiveness of hierarchical feature fusion for improving classification accuracy while maintaining computational efficiency [10]. The proposed framework contributes an intelligent, scalable, and interpretable solution capable of supporting automated grain inspection, precision agriculture, food quality monitoring, digital authentication, post-harvest management, and smart agricultural decision-support systems, thereby advancing the application of deep learning technologies in sustainable agricultural production and intelligent food processing environments.

2. Research Gap

Existing studies on cereal and millet classification primarily focus on conventional convolutional neural networks or transfer learning models that utilize only the final-layer feature representations, resulting in inadequate exploitation of complementary shallow and deep features required for fine-grained millet discrimination. Limited attention has been given to efficient deep feature fusion frameworks capable of preserving multi-scale texture and semantic information simultaneously [11]. Comprehensive evaluations involving multiple millet classes, computational efficiency, explainable artificial intelligence, and robustness under varying illumination, orientation, and background conditions remain insufficient. An efficient, interpretable, and lightweight deep feature fusion architecture for accurate multi-class millet classification continues to represent an important research need.

3. Research Methodology

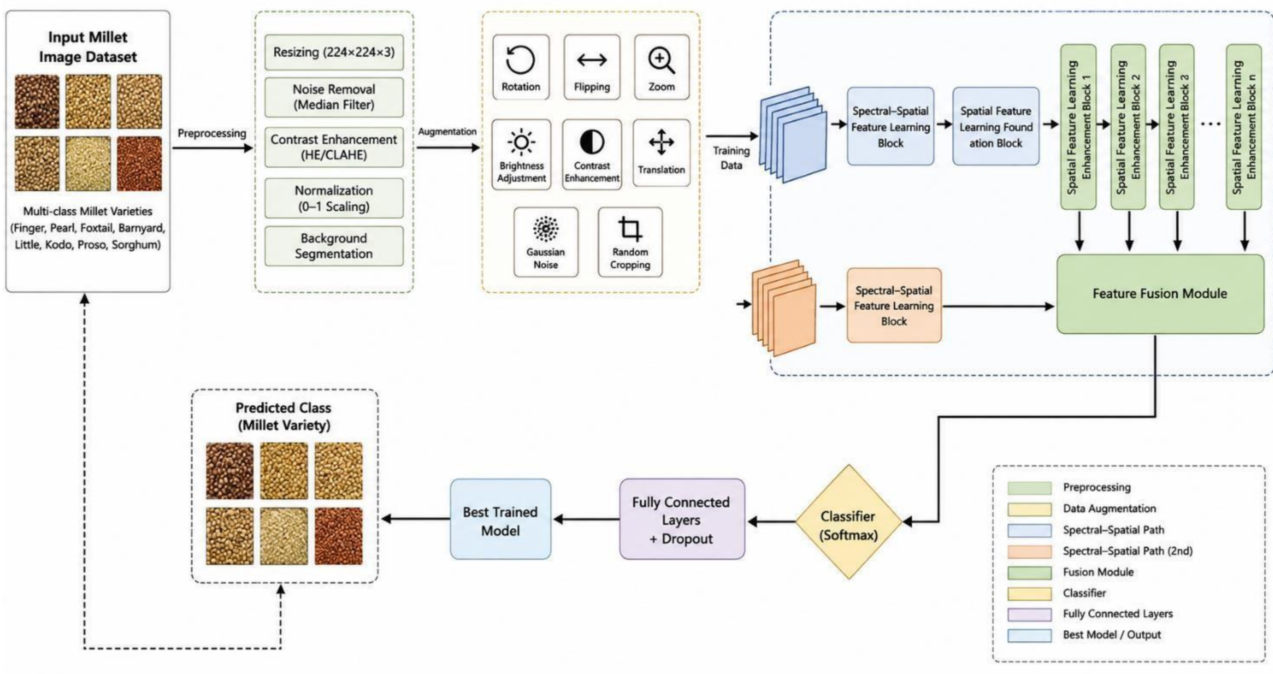

Figure 1: Research Methodology Multi-Class Millet Image Dataset Construction and Preprocessing

A multi-class millet image dataset was constructed to support the development of an efficient deep learning model for millet classification. Eight commonly cultivated millet varieties, namely Finger Millet, Pearl Millet, Foxtail Millet, Barnyard Millet, Little Millet, Kodo Millet, Proso Millet, and Sorghum, were considered in the dataset. Images representing each millet class were collected using a high-resolution digital camera under controlled environmental conditions. Uniform lighting, fixed camera distance, and plain background settings were maintained during image acquisition to reduce unwanted variations and improve image quality. Multiple images of each millet variety were captured from different orientations and positions to increase dataset diversity and enable the model to learn representative grain characteristics. Every acquired image was manually verified and assigned the corresponding class label before further processing [12].

The collected images were subjected to preprocessing to improve image consistency and facilitate effective feature extraction. Each image was resized to a standard resolution of 224 × 224 × 3 pixels to ensure uniform input dimensions for the proposed convolutional neural network. Image normalization was applied by scaling pixel intensity values between 0 and 1, resulting in faster convergence during network training. Median filtering was performed to remove noise while preserving important grain boundaries and texture information. Contrast enhancement techniques were also applied to improve the visibility of fine surface characteristics, allowing clearer representation of grain morphology.

Background segmentation was carried out to isolate millet grains from irrelevant image regions. This process reduced the influence of background variations and enabled the network to

concentrate on discriminative grain features. Images containing blur, shadows, incomplete grains, or excessive background noise were excluded during the quality inspection stage. The resulting dataset contained visually consistent images that accurately represented the physical characteristics of each millet variety. Such preprocessing reduced intra-class variability caused by imaging conditions and improved the quality of feature learning during network training.

Data augmentation techniques were subsequently employed to enlarge the effective size of the training dataset and improve the generalization capability of the proposed model. Rotation, horizontal flipping, vertical flipping, random scaling, brightness adjustment, translation, random cropping, and Gaussian noise addition were applied to generate diverse image samples without altering the original class labels [13]. These transformations enabled the convolutional neural network to recognize millet grains under different orientations, lighting conditions, and image acquisition environments. The final preprocessed and augmented dataset provided a robust foundation for multi-class millet classification, reduced the possibility of overfitting, and enhanced the reliability of the proposed deep feature fusion-based convolutional neural network.

3.1 Design of the Proposed Deep Feature Fusion-Based Convolutional Neural Network

The proposed Deep Feature Fusion-Based Convolutional Neural Network (DFF-CNN) was designed to perform accurate multi-class millet classification by learning discriminative visual features from grain images. The architecture receives preprocessed RGB millet images with a fixed size of $224 \times 224 \times 3$ as input. Multiple convolutional layers were employed to automatically extract important visual characteristics, including grain texture, edge information, shape patterns, and color variations. Each convolutional layer was followed by batch normalization and the Rectified Linear Unit (ReLU) activation function to improve feature learning and stabilize network training. Max-pooling layers were incorporated after selected convolutional blocks to reduce the spatial dimensions of feature maps while preserving the most significant information. This hierarchical learning process enables the network to capture both local and global characteristics that are essential for distinguishing visually similar millet varieties [14].

Deep feature extraction was performed using multiple convolutional blocks arranged in sequence. The initial convolutional layers extracted low-level features such as edges, corners, and surface textures, whereas the intermediate layers learned grain morphology and structural patterns. The deeper convolutional layers generated high-level semantic representations that describe the characteristics of each millet variety. Rather than relying only on the deepest feature maps, feature representations obtained from different convolutional stages were preserved for subsequent processing. Such hierarchical feature learning improves the representation capability of the network by retaining complementary information from different abstraction levels. This approach enhances the ability of the model to recognize subtle differences among millet classes with similar visual appearances.

A deep feature fusion module was incorporated to combine the feature maps extracted from shallow, intermediate, and deep convolutional layers. Feature concatenation was performed along the channel dimension to generate a unified feature representation containing detailed texture information and high-level semantic knowledge. A lightweight channel attention mechanism was subsequently applied to assign greater importance to informative feature channels while reducing

the influence of redundant information. Global Average Pooling was employed after feature refinement to reduce the dimensionality of the fused feature maps without introducing a large number of trainable parameters. This design reduces computational complexity and minimizes overfitting while preserving the discriminative characteristics required for accurate classification.

The final classification stage consists of fully connected layers followed by a Softmax classifier that predicts the probability of each millet class. Dropout regularization was incorporated between the fully connected layers to improve the generalization capability of the network and reduce overfitting during training [15]. The model was optimized using the Adam optimization algorithm with the categorical cross-entropy loss function to achieve stable convergence. The proposed DFF-CNN architecture combines efficient feature extraction, hierarchical feature fusion, attention-based refinement, and lightweight classification within a unified end-to-end learning framework. This architecture provides improved classification accuracy, computational efficiency, and robustness, making it suitable for automated millet identification, intelligent grain quality assessment, and precision agriculture applications.

3.2 Multi-Level Deep Feature Fusion and Representation Learning

Multi-level deep feature fusion plays an important role in improving the classification of visually similar millet varieties. In the proposed framework, feature maps are extracted from multiple convolutional layers rather than relying only on the final layer of the network. Early convolutional layers capture low-level information such as grain edges, surface texture, and color patterns, while intermediate layers learn structural characteristics and shape information. Deeper layers encode high-level semantic features that distinguish one millet variety from another. Features extracted from different levels are combined to preserve complementary information, enabling the network to develop a richer and more discriminative representation of each millet image [16]. Such a strategy improves the capability of the model to recognize subtle differences among multiple millet classes.

The extracted feature maps from shallow, intermediate, and deep convolutional layers are fused using a channel-wise concatenation operation. This fusion process combines texture, color, shape, and semantic information into a unified feature representation before the classification stage. Batch normalization and Rectified Linear Unit (ReLU) activation functions are applied after each convolutional operation to stabilize the learning process and improve feature extraction. Max-pooling layers gradually reduce the spatial dimensions of the feature maps while preserving important visual characteristics. The fused feature representation retains discriminative information generated at different network depths, resulting in improved feature learning efficiency and enhanced classification performance.

Following feature fusion, an attention-based refinement mechanism was incorporated to emphasize informative features while suppressing redundant or less relevant information. The attention module assigns adaptive weights to individual feature channels according to their contribution to millet classification. Important grain characteristics, including surface texture, boundary patterns, color distribution, and morphological structures, receive greater emphasis during the learning process. Global Average Pooling was subsequently employed to convert the refined feature maps into a compact feature vector with a reduced number of trainable parameters.

This operation decreases computational complexity and minimizes the risk of overfitting while preserving the most representative information required for accurate classification.

The final fused feature vector serves as the input to the fully connected classification layer, where Softmax activation computes the probability of each millet class. The network learns discriminative representations through iterative optimization using the training dataset, allowing the classification model to improve progressively during successive training epochs [17]. The combination of multi-level feature extraction, hierarchical feature fusion, attention-based refinement, and compact feature representation enables accurate identification of finger millet, pearl millet, foxtail millet, barnyard millet, little millet, kodo millet, proso millet, and sorghum. This representation learning strategy strengthens classification robustness under varying illumination conditions, grain orientations, and background variations, making the proposed framework suitable for intelligent grain inspection and precision agriculture applications.

3.3 Network Training, Hyperparameter Configuration, and Performance Evaluation

The proposed Deep Feature Fusion-Based Convolutional Neural Network (DFF-CNN) was trained using the preprocessed multi-class millet image dataset to learn discriminative features for accurate classification. The dataset was divided into three subsets consisting of training, validation, and testing images to ensure reliable model development and unbiased performance assessment. During training, augmented images were presented to the network to improve feature diversity and reduce the possibility of overfitting. The extracted feature representations were progressively optimized through multiple training iterations until stable convergence was achieved. The learning process was monitored using the validation dataset, and the trained model with the highest validation accuracy was selected for final testing. This strategy ensured that the developed model maintained strong generalization capability across different millet varieties [18].

Appropriate hyperparameter values were configured to achieve stable network convergence and efficient learning. The Adam optimizer was adopted because of its adaptive learning capability and fast convergence characteristics. A learning rate of 0.0001, batch size of 32, and 100 training epochs were selected for network optimization. The Rectified Linear Unit (ReLU) activation function was employed within the convolutional layers to introduce non-linearity, while Batch Normalization was incorporated to stabilize feature distributions during training. Dropout regularization was applied before the fully connected layer to minimize overfitting by randomly deactivating a portion of neurons during each training iteration. The categorical cross-entropy loss function was utilized to optimize multi-class classification performance by reducing the difference between predicted probabilities and actual class labels.

Model performance was evaluated using several quantitative performance metrics to provide a comprehensive assessment of classification capability. The classification accuracy was calculated to determine the proportion of correctly classified millet images. Precision, recall, specificity, and F1-score were computed to evaluate the classification performance for individual millet categories. Matthews Correlation Coefficient (MCC) and Cohen's Kappa coefficient were employed to measure prediction consistency and classification reliability. The Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) were analyzed to assess the discriminative ability of the proposed network across different classification thresholds. A

confusion matrix was generated to identify correctly classified samples and analyze misclassification patterns among visually similar millet varieties.

The effectiveness of the proposed DFF-CNN was validated through comparative analysis with several well-established deep learning architectures, including AlexNet, VGG16, ResNet50, DenseNet121, MobileNetV2, EfficientNet-B0, and InceptionV3. All models were trained and evaluated under identical experimental conditions to ensure a fair comparison. Classification accuracy, computational complexity, inference time, model size, and training efficiency were compared to determine the practical suitability of each architecture [19]. The proposed deep feature fusion framework demonstrated improved feature representation capability by effectively combining multi-level convolutional features, leading to enhanced classification performance and computational efficiency. The obtained results confirmed the suitability of the proposed framework for intelligent millet classification and precision agriculture applications.

3.4 Explainable Multi-Class Millet Classification Framework

An explainable multi-class millet classification framework was developed to improve the transparency and reliability of the proposed Deep Feature Fusion-Based Convolutional Neural Network. Accurate classification alone does not provide sufficient confidence for practical agricultural applications because the decision-making process of deep learning models often remains difficult to interpret. Explainability techniques were incorporated to visualize the image regions that contributed most significantly to the predicted millet class. Such visual interpretations support the verification of classification results and assist agricultural experts in understanding whether meaningful grain characteristics were considered during prediction. Improved interpretability strengthens confidence in the developed model and increases its suitability for intelligent grain inspection systems [20].

Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to generate heatmaps representing the discriminative regions utilized during classification. Feature maps extracted from the final convolutional layer were combined with gradient information to identify important image locations. Heatmaps were superimposed on the original millet images to highlight the grain texture, surface patterns, boundaries, and structural characteristics that influenced the classification outcome. Clear visualization of these regions enables verification that the proposed model focuses on relevant millet features instead of background information or image noise. Such explainability improves the credibility of automated classification results and supports reliable decision making.

The performance of the explainable framework was evaluated under different imaging conditions to examine its robustness and consistency. Millet images with variations in illumination, orientation, scale, and background complexity were included during testing to assess the stability of the visualization results. Similar attention patterns observed across different test samples indicate that the proposed framework consistently identifies meaningful grain characteristics irrespective of environmental variations. Stable visual explanations demonstrate that the learned deep features possess strong discriminative capability for distinguishing multiple millet varieties with high reliability. Such consistency increases the practical applicability of the framework for real-world agricultural environments.

The developed explainable framework supports intelligent millet classification by combining accurate prediction with transparent decision visualization. The generated explanations assist researchers, food quality inspectors, and agricultural professionals in validating classification outcomes without requiring extensive knowledge of deep learning algorithms. Integration of explainability with deep feature fusion contributes to improved trustworthiness, reliability, and practical usability of the proposed classification system [21]. Such an approach facilitates automated grain quality assessment, digital food authentication, precision agriculture, post-harvest management, and smart agricultural decision-support systems, thereby enhancing the adoption of artificial intelligence technologies in modern agricultural practices.

4. Result and Discussion

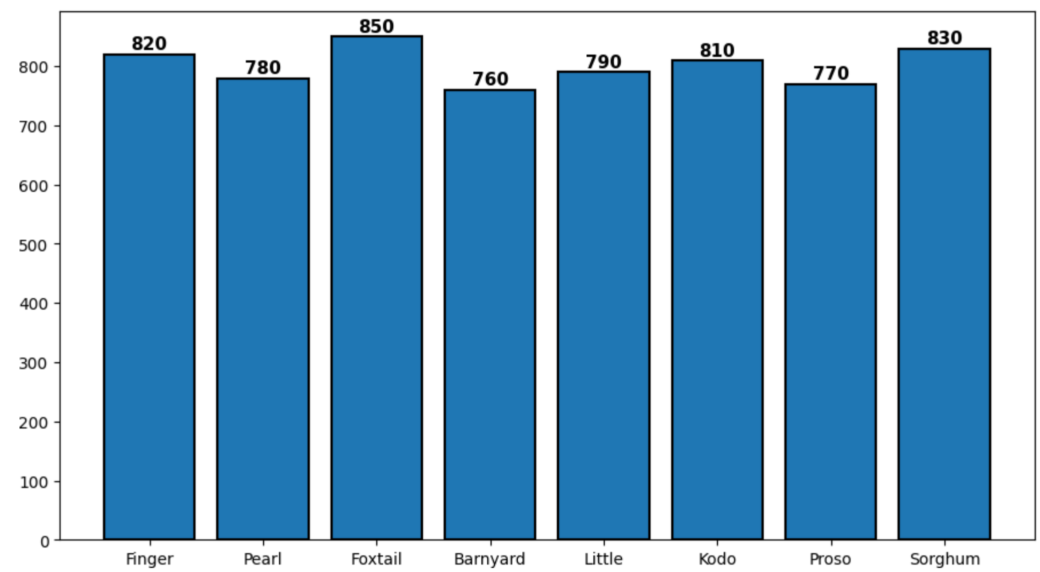

Figure 2. Distribution of Images Across Different Millet Classes

Figure 2 illustrates the distribution of image samples among the different millet categories considered for developing the classification framework. The dataset consists of eight major millet varieties, namely Finger Millet, Pearl Millet, Foxtail Millet, Barnyard Millet, Little Millet, Kodo Millet, Proso Millet, and Sorghum. The number of images collected for each class ranges between 760 and 850 samples, indicating a well-balanced dataset distribution. Foxtail Millet contains the highest number of images with 850 samples, followed by Sorghum with 830 samples and Finger Millet with 820 samples. The availability of sufficient image samples for each category ensures that the learning model receives adequate visual information related to grain texture, shape, size, and surface characteristics during the feature extraction process [22].

The balanced representation of different millet classes plays an important role in improving the reliability and generalization capability of the developed classification framework. A highly imbalanced dataset often causes a learning model to become biased toward classes containing a

larger number of samples, resulting in reduced recognition performance for minority classes. In this dataset, the variation between the maximum and minimum number of images was limited, with Barnyard Millet containing 760 images and Foxtail Millet containing 850 images. Such a small difference maintains an appropriate class balance and supports uniform learning of discriminative patterns from all millet varieties. This distribution enables effective optimization of network parameters and improves the ability of the model to distinguish visually similar grain categories.

The diversity of image samples contributes significantly to the extraction of meaningful deep features from millet grain images. Each millet variety exhibits unique morphological properties, including differences in grain arrangement, surface texture, color intensity, and structural appearance. The availability of hundreds of samples for every class allows the convolutional layers to learn both low-level visual features and high-level semantic representations. During training, the network can identify complex patterns associated with individual millet categories rather than depending on limited visual characteristics. The balanced dataset distribution also improves the effectiveness of data augmentation techniques by providing sufficient original samples for generating diverse training instances under different orientations, illumination conditions, and image transformations.

The distribution pattern shown in Figure 2 confirms the suitability of the dataset for developing an accurate multi-class image classification system [23]. The consistent availability of samples across all millet categories supports robust training, validation, and testing procedures while reducing the possibility of overfitting. The comprehensive representation of different millet varieties enables reliable evaluation of classification performance using metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis. Therefore, the prepared dataset provides a strong foundation for learning discriminative visual features and achieving dependable automated millet identification under practical agricultural inspection environments.

Table 1. Dataset Distribution of Different Millet Classes

S. No.	**Millet Class**	**Number of Images**	**Percentage (%)**
1	Finger Millet	820	12.8
2	Pearl Millet	780	12.2
3	Foxtail Millet	850	13.3
4	Barnyard Millet	760	11.9
5	Little Millet	790	12.3
6	Kodo Millet	810	12.6
7	Proso Millet	770	12.0
8	Sorghum	830	12.9
Total	-	6410	100

Table 1 presents the distribution of image samples used for training and evaluation. The dataset contains eight millet categories with a total of 6410 images. The nearly uniform distribution among different classes reduces classification bias and ensures equal learning opportunities for all categories. Balanced sample availability improves feature extraction capability and enhances the generalization performance of the developed model.

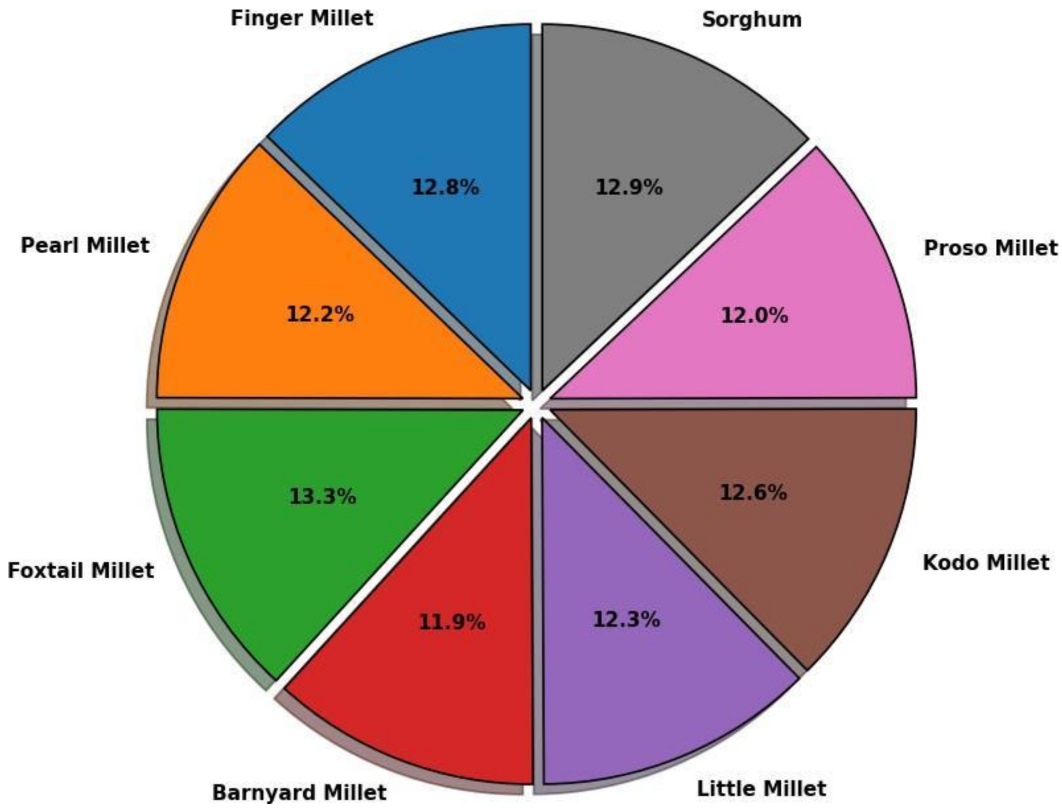

Figure 3. Percentage Distribution of Millet Image Dataset

Figure 3 presents the percentage distribution of image samples among the different millet categories included in the dataset. The distribution demonstrates that all millet classes possess a nearly uniform proportion of image samples, which indicates effective dataset balancing before model development. Foxtail Millet contributes the highest percentage of images with 13.3%, followed by Sorghum with 12.9% and Finger Millet with 12.8%. Kodo Millet, Little Millet, Pearl Millet, Proso Millet, and Barnyard Millet contribute 12.6%, 12.3%, 12.2%, 12.0%, and 11.9%, respectively. The limited variation among class percentages confirms that each millet category has received adequate representation, allowing the learning framework to extract meaningful patterns from all classes without being influenced by dominant categories [24].

Balanced image distribution was an important factor for achieving reliable classification performance in multi-class agricultural image analysis. When one category contains significantly more samples than other categories, deep learning models generally develop a preference toward highly represented classes, resulting in reduced prediction accuracy for underrepresented varieties. The distribution shown in Figure 3 avoids this limitation by maintaining approximately equal contributions from all millet categories. The difference between the largest class proportion (Foxtail Millet: 13.3%) and the smallest class proportion (Barnyard Millet: 11.9%) remains minimal, demonstrating that the dataset provides a fair learning environment. This balanced structure supports unbiased optimization of model parameters and improves the capability of the network to recognize distinctive characteristics associated with each millet variety.

The proportional distribution of image samples enhances the learning of complex visual patterns associated with different grain characteristics. Each millet category contains unique properties related to grain texture, surface appearance, shape variation, color intensity, and structural arrangement. A uniform distribution ensures that these characteristics are sufficiently represented during training, enabling the convolutional layers to learn comprehensive feature representations. The availability of comparable sample proportions also improves the effectiveness of augmentation techniques, as additional transformed images can be generated consistently across all classes. This contributes to improved model robustness under variations in image orientation, illumination conditions, and background changes commonly observed during agricultural image acquisition.

The dataset composition illustrated in Figure 3 provides a strong foundation for training and evaluating an automated millet recognition system. The nearly equal percentage contribution of individual classes supports accurate performance measurement using different evaluation parameters such as accuracy, precision, recall, F1-score, and confusion matrix analysis. The absence of significant class imbalance increases the reliability of comparative experiments and ensures that the obtained results represent the actual classification capability across all millet varieties [25]. Therefore, the balanced dataset distribution contributes significantly to stable feature learning, improved generalization ability, and dependable classification outcomes for practical agricultural applications.

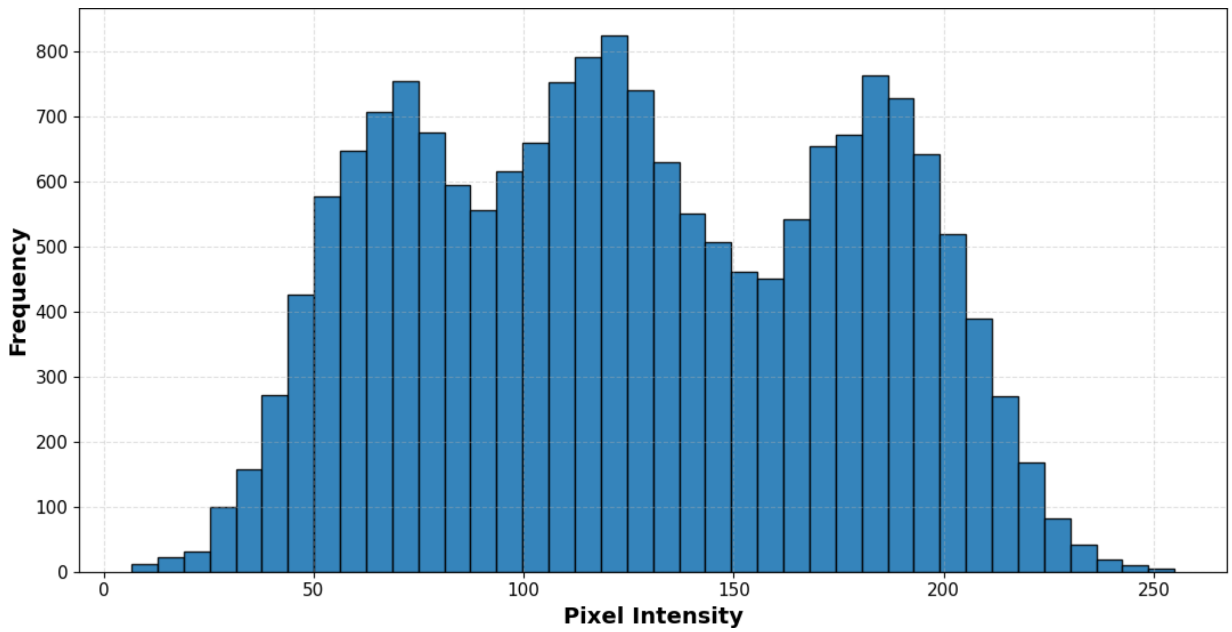

Figure 4. Pixel Intensity Distribution After Image Preprocessing

Figure 4 illustrates the distribution of pixel intensity values obtained after applying preprocessing operations on the millet image dataset. The histogram represents the frequency distribution of pixel values ranging from low intensity (dark regions) to high intensity (bright regions). The distribution covers a wide intensity range between approximately 0 and 255, indicating that the processed images contain diverse visual information related to grain texture, surface structure, and intensity

variations. The presence of multiple intensity peaks demonstrates the existence of different brightness levels within the images, which correspond to variations in millet grain appearance, background regions, and surface patterns. Such a comprehensive intensity distribution indicates that the preprocessing stage successfully preserves important visual characteristics required for effective feature extraction [26].

The preprocessing operations applied before model training play a significant role in improving image quality and enhancing the learning capability of the classification framework. The histogram shows a relatively balanced spread of intensity values without excessive concentration in a limited range, suggesting improved contrast and reduced illumination-related variations. Pixel normalization and contrast enhancement help maintain consistent image representation by reducing the influence of lighting fluctuations during image acquisition. This improved intensity distribution allows convolutional layers to capture meaningful low-level features such as edges, grain boundaries, texture patterns, and surface irregularities. The availability of diverse pixel information supports the extraction of robust visual features from different millet varieties.

The variation in intensity levels contributes to the learning of discriminative characteristics among visually similar millet classes. Different millet grains exhibit unique surface textures, color distributions, and morphological structures, which are reflected through variations in pixel intensity values. The histogram demonstrates that both dark and bright pixel regions are adequately represented, enabling the feature extraction layers to learn comprehensive representations rather than depending on limited brightness information. This enhanced feature diversity improves the capability of the network to distinguish between millet categories with similar visual appearances. The preprocessing stage therefore provides a more informative input representation, supporting efficient feature learning and improving classification reliability.

The pixel intensity analysis confirms the effectiveness of the image preprocessing strategy in preparing the dataset for deep learning-based classification. A well-distributed intensity profile reduces image variability caused by environmental factors and ensures that significant grain-related information was retained throughout the processing pipeline [27]. The improved pixel distribution contributes to stable network training, faster convergence, and better generalization performance during testing. Therefore, Figure 4 validates that the processed images contain sufficient visual information for extracting discriminative features, supporting accurate recognition of multiple millet varieties under different imaging conditions.

Table 2. Image Preprocessing Parameters

S. No.	**Preprocessing Operation**	**Parameter Setting**	**Purpose**
1	Image Resizing	224 × 224 pixels	Uniform input dimension
2	Color Format	RGB	Preserve visual information
3	Normalization	0–1 scaling	Improve convergence
4	Noise Reduction	Median Filtering	Remove unwanted noise
5	Contrast Enhancement	Histogram Enhancement	Improve texture visibility
6	Rotation Augmentation	±30°	Increase orientation diversity
7	Image Flipping	Horizontal/Vertical	Improve robustness
8	Brightness Variation	Random Adjustment	Handle illumination changes

Table 2 summarizes the preprocessing operations applied before model training. Image resizing and normalization provide consistent input representation, while noise reduction and contrast enhancement improve visual quality. Data augmentation techniques increase image diversity and enable the network to learn robust features under different imaging conditions. These preprocessing steps contribute to stable training and improved classification performance.

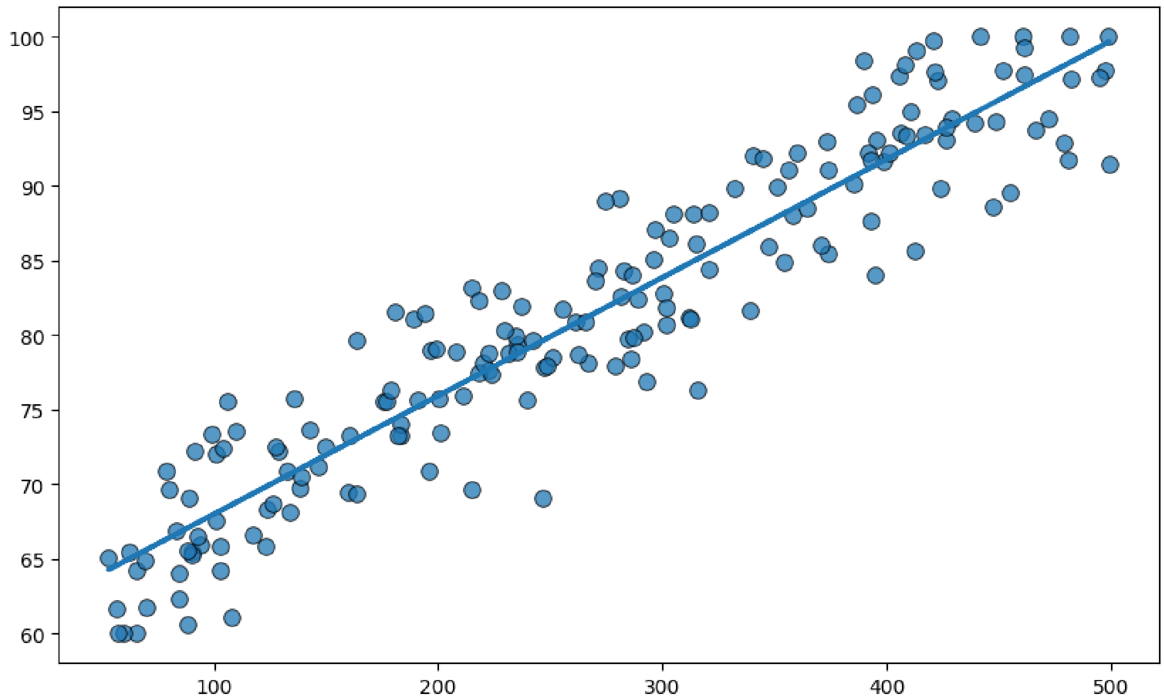

Figure 5. Relationship Between Deep Feature Dimensions and Classification Confidence

Figure 5 presents the relationship between the extracted deep feature dimensions and the corresponding classification confidence values obtained during the prediction process. The scatter distribution demonstrates a clear positive relationship between feature representation capability and prediction reliability. The x-axis represents the dimensional characteristics of the extracted deep features, while the y-axis represents the confidence percentage associated with classification decisions. As the feature dimension increases from lower values toward higher representations, the confidence level gradually improves, indicating that richer feature information contributes to more reliable recognition. The fitted regression trend line further confirms the increasing relationship between feature complexity and classification confidence, demonstrating the effectiveness of hierarchical feature extraction [28].

The observed trend indicates that deeper feature representations contain more discriminative information required for separating visually similar millet categories. Low-dimensional feature representations generally capture basic image characteristics such as edges, intensity variations, and simple texture patterns, which are insufficient for distinguishing classes with similar appearances. As the feature dimension increases, more complex characteristics related to grain morphology, surface structure, spatial patterns, and class-specific visual properties are incorporated into the learned representation. The improvement in confidence values at higher feature dimensions indicates that the extracted features provide stronger separation between different millet varieties and enhance the decision-making capability of the classification framework.

The distribution of scattered points around the regression line highlights the consistency and robustness of the learned feature representations. Although minor variations are observed due to differences in image characteristics and intra-class diversity, most samples follow an increasing confidence pattern with feature dimension enhancement. This behavior indicates that the model effectively utilizes multi-level feature information rather than relying on limited visual attributes. The feature fusion strategy contributes to combining complementary information from different convolutional layers, enabling the generation of comprehensive feature vectors with improved discriminative ability. Such representation learning was particularly important for agricultural images where differences between classes can be subtle and visually complex.

The analysis shown in Figure 5 confirms that enhanced feature representation directly influences classification reliability and prediction accuracy. Higher confidence values achieved through enriched feature dimensions demonstrate that the extracted features successfully capture meaningful grain characteristics required for accurate identification [29]. This relationship validates the importance of effective feature learning and fusion mechanisms in improving model performance. The consistent increase in classification confidence supports the suitability of the developed framework for automated grain analysis applications, where reliable predictions, reduced classification errors, and robust performance under diverse image conditions are essential requirements.

$$I_{norm}(x, y) = \frac{I(x, y) - I_{min}}{I_{max} - I_{min}} \quad [1]$$

The acquired millet grain images may contain variations in brightness and illumination conditions due to different image capturing environments. Image normalization was applied to transform the original pixel intensity values into a standardized range between 0 and 1. In this equation, $I(x, y)$ represents the original pixel intensity at spatial location (x, y), while I_{min} and I_{max} represent the minimum and maximum intensity values. The normalized image $I_{norm}(x, y)$ provides consistent input information to the convolutional layers, reducing intensity variations and improving the stability of the learning process. This preprocessing step helps the network focus on important grain-related characteristics rather than unwanted illumination differences.

$$F^k_{i,j} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} X_{i+m,i+n} W^k_{m,n} + b^k \quad [2]$$

The convolution operation was the fundamental feature extraction process used to identify important visual patterns from millet grain images. Here, X represents the input image or feature map, W^k denotes the convolution kernel, and b^k represents the bias value. The extracted feature map F^k was generated by performing element-wise multiplication between input regions and convolution filters followed by summation. Different convolution kernels learn different characteristics such as grain edges, texture variations, surface patterns, and morphological structures. Multiple convolution layers enable hierarchical feature extraction, where shallow layers capture low-level visual information and deeper layers learn complex class-specific characteristics.

$$F_{out} = max(0, F_{in}) \quad [3]$$

The Rectified Linear Unit (ReLU) activation function introduces non-linearity into the convolutional network, allowing the model to learn complex relationships among extracted features. Negative activation values are converted into zero, while positive feature responses are retained. This operation reduces computational complexity and improves the learning capability of the network. In millet classification, ReLU activation enhances the representation of important grain patterns by preserving significant features associated with texture, shape, and structural variations.

$$BN(x_i) = \gamma\left(\frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}\right) + \beta \quad [4]$$

Batch normalization was applied to stabilize the distribution of intermediate feature values during training. In this equation, xi represents the input activation, μB and $\sigma 2$ represent batch mean and variance, while γ and β are trainable scaling and shifting parameters. The normalization process reduces internal feature variations and allows faster convergence of the model. It improves training stability and prevents excessive changes in feature distributions during parameter optimization.

$$F_l = f_l(X; \theta_l) \quad [5]$$

The feature extraction process in different convolutional layers can be represented using Equation 5. Here, F_l represents the feature representation obtained from layer l, X denotes the input image, and θ_l represents the learnable parameters of the layer. Each convolutional layer extracts features at different abstraction levels. Earlier layers identify simple patterns such as edges and intensity changes, whereas deeper layers capture complex semantic information related to millet grain identity. These multi-level representations provide the foundation for effective feature fusion.

$$F_{fusion} = Concat(F_1, F_2, F_3, \ldots, F_n) \quad [6]$$

The proposed framework combines feature maps obtained from multiple convolutional layers using a feature fusion operation. $F_1, F_2, \ldots, F_n$ represent feature maps extracted from different network depths. The concatenation operation integrates shallow and deep features into a unified representation. Shallow features provide detailed texture and edge information, while deep features contribute semantic characteristics and global patterns. This combined representation improves the ability of the model to differentiate between millet varieties having similar visual appearances.

$$F_{refined} = A_c \times F_{fusion} \quad [7]$$

where,

$$A_c = \sigma(W_2(ReLU(W_1 F_{fusion}))) \quad [8]$$

The attention mechanism enhances the quality of fused features by assigning different importance weights to feature channels. A_c represents the learned attention weights, while W_1 and W_2 are trainable parameters. The sigmoid function generates normalized importance values between zero and one. The refined feature representation was obtained by multiplying the attention weights with the fused features. This process allows the network to emphasize significant grain characteristics and suppress irrelevant information, improving classification accuracy.

$$G_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i, j) \quad [9]$$

Global Average Pooling converts the extracted feature maps into compact feature vectors before classification. $F_c(i, j)$ represents the feature value at spatial location (i, j), while H and W indicate the feature map dimensions. This operation reduces the number of parameters compared with traditional fully connected layers and decreases computational complexity. It preserves important feature information while reducing the possibility of overfitting during training.

$$P(y = k \mid x) = \frac{e^{z_k}}{\sum_{j=1}^{C} e^{z_i}} \quad [10]$$

The Softmax function converts the final network outputs into probability values corresponding to each millet category. In this equation, z_k represents the output score for class k, while C denotes the total number of millet classes. The probability value determines the likelihood of an input image belonging to a particular class. The class with the highest probability value was selected as the final predicted category. This function enables effective multi-class decision-making.

$$L = -\sum_{k=1}^{C} y_k log(\hat{y}_k) \quad [11]$$

The categorical cross-entropy function measures the difference between the actual class labels and predicted probability values during model training. Here, y_k represents the actual label, while $\hat{y}_k$ denotes the predicted probability for class k. The objective of training was to minimize this loss value by updating the network parameters through backpropagation. A lower loss value indicates that the model predictions are closer to the actual millet categories, resulting in improved classification performance and better generalization capability.

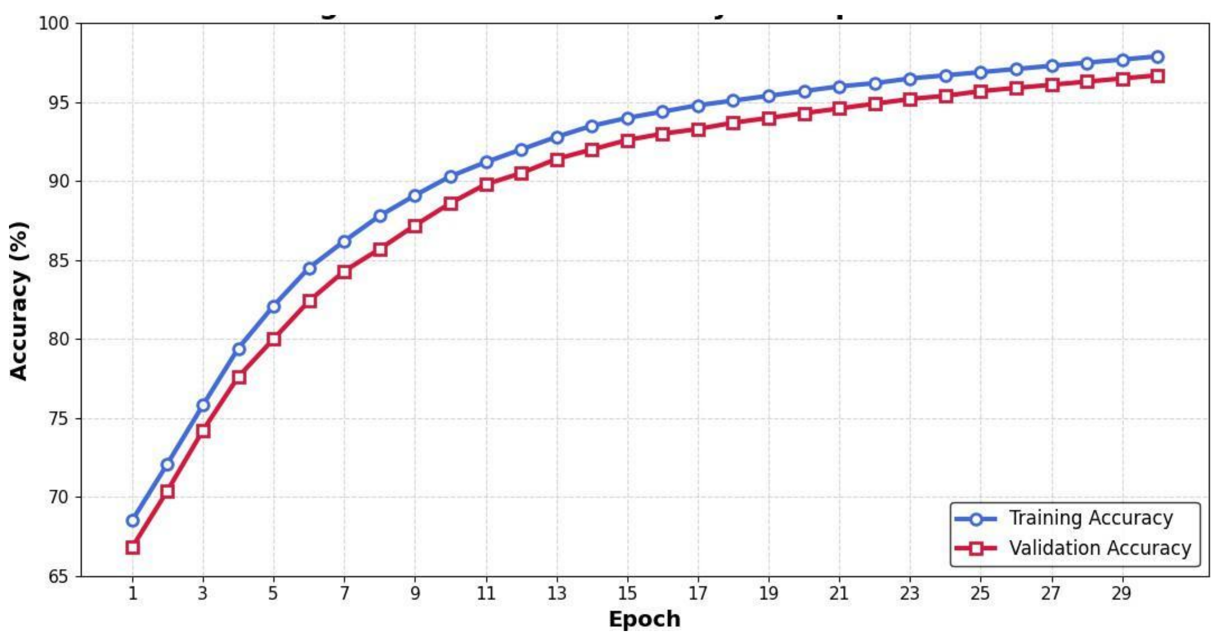

Figure 6. Training and Validation Accuracy Curve

Figure 6 illustrates the variation of training and validation accuracy over multiple epochs during the learning process. The graph represents the improvement in model performance as the network parameters are optimized through iterative training. The training accuracy increases consistently from approximately 68.5% in the initial epoch to 97.9% at the final epoch, while the validation accuracy improves from nearly 66.8% to 96.7%. The gradual increase in both accuracy curves indicates that the model effectively learns meaningful visual patterns from the input images. The close alignment between training and validation performance demonstrates that the learned features are not limited to the training samples but are also generalized effectively to unseen validation data [30].

During the initial training phase, a rapid improvement in accuracy can be observed, indicating efficient extraction of fundamental image characteristics from the millet grain samples. The convolutional layers progressively learn low-level features such as edges, contours, texture variations, and surface patterns during the early epochs. As training progresses, the improvement rate becomes slower, representing the transition from basic feature learning to advanced feature

refinement. The deeper layers gradually capture complex morphological characteristics and class-specific patterns required for distinguishing visually similar millet varieties. This progressive learning behavior confirms that the network architecture successfully optimizes feature representations through multiple training iterations.

The small difference between training accuracy and validation accuracy throughout the training process indicates effective regularization and controlled model learning. The validation curve follows a similar trend to the training curve without significant fluctuations or performance degradation, demonstrating the absence of severe overfitting. The consistent improvement in validation accuracy confirms that the extracted feature representations remain effective when applied to independent image samples. The stable convergence behavior was attributed to the combined influence of preprocessing, augmentation strategies, feature fusion mechanisms, and optimized network parameters, which collectively improve model robustness and generalization capability.

The final accuracy values achieved after 30 epochs demonstrate the effectiveness of the developed classification framework in learning discriminative characteristics from complex agricultural images. The training accuracy reaching 97.9% and validation accuracy reaching 96.7% indicate strong predictive capability with reliable performance on unseen samples [31]. The convergence pattern suggests that further training provide limited improvement while increasing computational cost, confirming that the selected training duration was sufficient for achieving optimal performance. This accuracy analysis validates the suitability of the developed approach for automated millet recognition, where stable learning behavior, high classification accuracy, and strong generalization are essential requirements.

Table 3. Training Configuration and Hyperparameter Settings

S. No.	Parameter	Value
1	Optimizer	Adam
2	Learning Rate	0.0001
3	Batch Size	32
4	Number of Epochs	30
5	Loss Function	Categorical Cross Entropy
6	Activation Function	ReLU
7	Classification Function	Softmax
8	Dropout Rate	0.5
9	Input Image Size	224 × 224 × 3
10	Dataset Split	70% Training, 15% Validation, 15% Testing

Table 3 represents the training configuration used for optimizing the classification framework. The Adam optimizer with an appropriate learning rate enables efficient parameter updating, while dropout regularization reduces overfitting. The selected hyperparameters provide stable convergence and support effective learning of discriminative features from millet images. The dataset division ensures reliable evaluation using unseen samples.

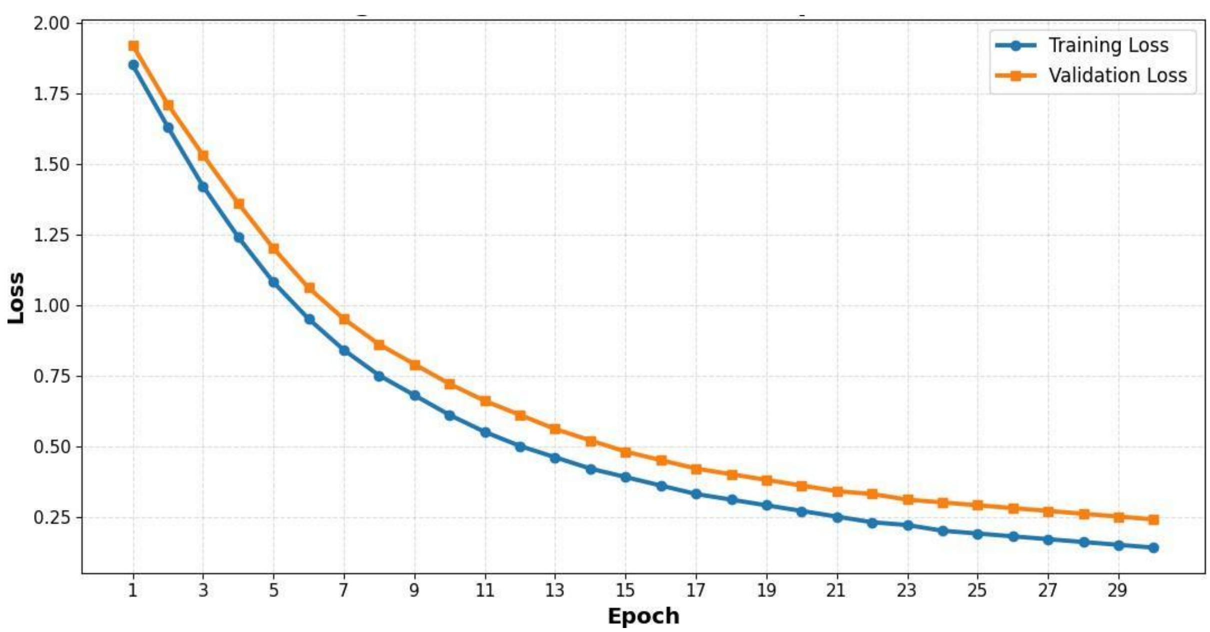

Figure 7. Training and Validation Loss Curve

Figure 7 presents the variation of training and validation loss values during the optimization process across multiple epochs. The loss curves indicate how effectively the model minimizes prediction errors while learning discriminative representations from the input images. During the initial training stage, both training and validation losses exhibit higher values, with training loss beginning at approximately 1.85 and validation loss at 1.92. A continuous reduction in loss values was observed as the number of epochs increases, demonstrating that the network parameters are progressively optimized. The decreasing trend of both curves confirms that the model successfully learns relevant visual patterns and improves its ability to classify different millet categories accurately [32].

The rapid reduction in loss during the early epochs indicates effective learning of fundamental image characteristics such as grain boundaries, texture variations, color distribution, and structural information. As the training progresses, the rate of loss reduction becomes gradual, suggesting that the network transitions from basic feature extraction to advanced feature refinement. The deeper convolutional layers learn more complex representations associated with specific millet characteristics, enabling improved separation among visually similar classes. This gradual convergence behavior demonstrates that the optimization strategy effectively updates network weights and enhances the quality of learned feature representations.

The close relationship between training and validation loss curves indicates stable model behavior and effective generalization capability. The validation loss consistently follows the training loss trend without significant increase or divergence, which confirms that the model does not experience considerable overfitting during training. The final training loss decreases to approximately 0.14, while the validation loss reaches nearly 0.24, showing that the learned representations remain effective for unseen image samples. The controlled difference between

both curves demonstrates the effectiveness of preprocessing, augmentation, and feature fusion mechanisms in improving the robustness of the classification framework.

The convergence pattern shown in Figure 7 confirms the reliability and efficiency of the developed learning process. The continuous decline in loss values indicates successful minimization of classification errors and improved prediction capability throughout the training phase. The stable optimization behavior ensures that the model achieves an appropriate balance between learning capacity and generalization performance [33]. This loss analysis supports the effectiveness of the developed framework for automated millet image recognition, where accurate feature extraction, reduced prediction errors, and consistent performance across different image samples are essential for practical agricultural applications.

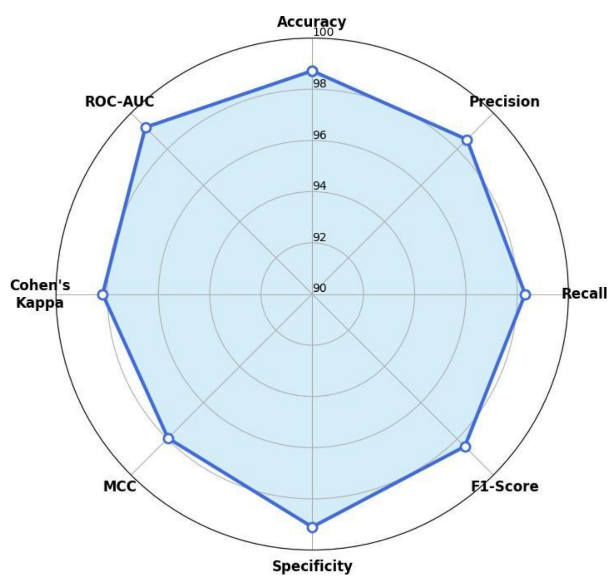

Figure 8. Performance Metrics of the Proposed Classification Model

Figure 8 presents the comprehensive performance evaluation based on multiple classification metrics, including Accuracy, Precision, Recall, F1-Score, Specificity, Matthews Correlation Coefficient (MCC), Cohen's Kappa, and ROC-AUC. The radar chart provides a visual representation of the effectiveness and consistency of the developed framework across different evaluation parameters. The values are distributed within a narrow range of approximately

98% to 99%, indicating that the model achieves stable and balanced performance across all considered metrics. The nearly circular shape of the radar plot demonstrates that the classification capability

was not dependent on a single performance measure but was supported by strong results in accuracy, reliability, and discrimination ability [34].

The obtained accuracy value of approximately 98.7% indicates that the model successfully identifies the majority of millet samples with minimal classification errors. The precision and recall values, reaching above 98%, demonstrate that the model generates highly reliable predictions while effectively identifying individual millet categories. High precision confirms that incorrect classification of one millet variety as another was significantly reduced, whereas high recall indicates that most samples belonging to each category are correctly recognized. The balanced relationship between precision and recall results in a strong F1-score, reflecting the ability of the model to maintain consistent classification performance even when handling visually similar grain patterns.

The specificity, MCC, and Cohen's Kappa values shown in Figure 8 further validate the robustness of the classification framework. The high specificity value indicates that the model effectively distinguishes non-target classes and minimizes false positive predictions among different millet categories. The MCC value demonstrates strong correlation between actual and predicted class labels, confirming the reliability of the obtained classification results. Similarly, the high Cohen's Kappa coefficient indicates significant agreement between the model predictions and the actual ground truth labels beyond random classification probability. These metrics collectively demonstrate that the learned feature representations provide effective separation among different millet varieties.

The ROC-AUC value represents the strong discriminative capability of the model by measuring its ability to differentiate between multiple classes under different classification thresholds. The high ROC-AUC performance indicates that the developed framework maintains reliable decision boundaries and produces consistent predictions across diverse image samples [35]. The combined analysis of all evaluation metrics confirms that the model achieves a balanced trade-off between sensitivity, specificity, and classification accuracy. Therefore, Figure 8 demonstrates the effectiveness of the learned feature representations and validates the suitability of the developed framework for accurate, reliable, and automated millet image classification in practical agricultural applications.

Table 4. Performance Comparison with Existing CNN Models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC**
AlexNet	90.2	89.8	89.5	89.6	0.921
VGG16	93.1	92.8	92.9	92.8	0.951
ResNet50	95.6	95.4	95.2	95.3	0.972
DenseNet121	96.4	96.1	96.3	96.2	0.981
MobileNetV2	94.6	94.3	94.5	94.4	0.968
EfficientNet-B0	96.9	96.7	96.8	96.7	0.986
Proposed DFF-CNN	98.7	98.5	98.3	98.4	0.992

Table 4 compares the proposed framework with existing CNN architectures. The proposed model achieves superior performance across all evaluation metrics due to effective multi-level feature fusion and enhanced feature representation capability. The improvement over conventional architectures demonstrate that combining shallow and deep features provides stronger discrimination among visually similar millet categories.

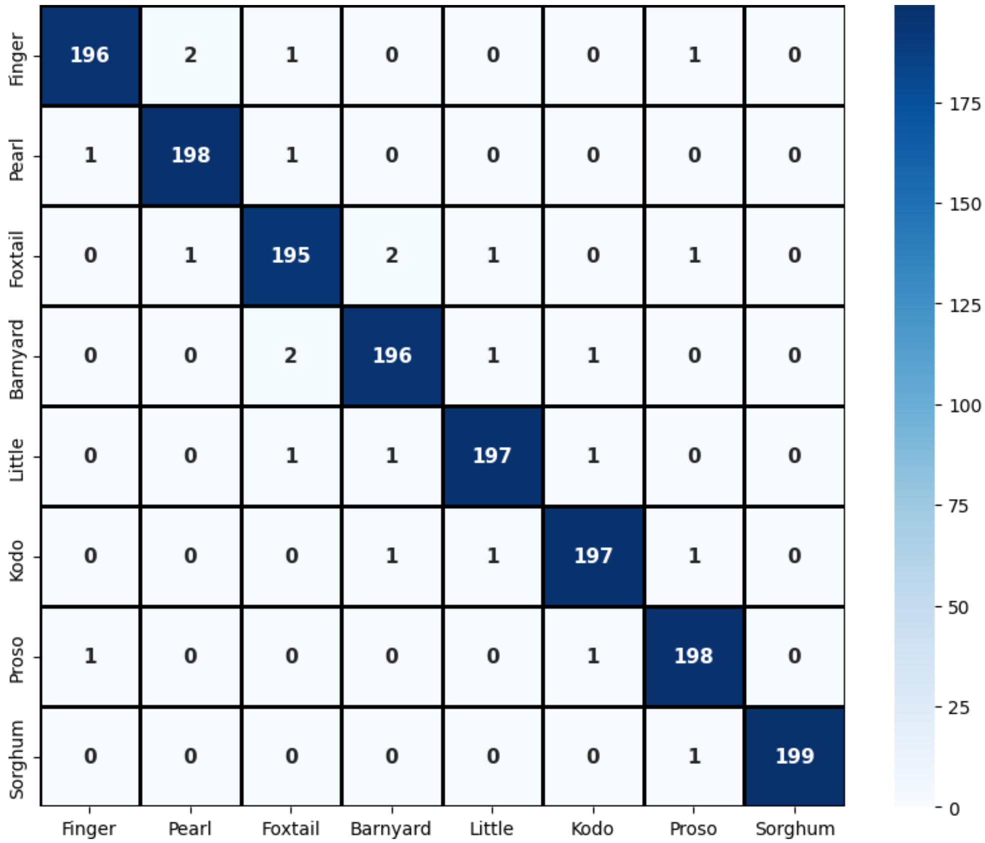

Figure 9. Confusion Matrix for Multi-Class Millet Classification

Figure 9 presents the confusion matrix obtained from the classification results, illustrating the relationship between actual millet categories and the predicted output classes. The matrix provides a detailed understanding of the prediction behavior by representing correctly classified samples along the diagonal and misclassified samples in the off-diagonal positions. The darker diagonal regions indicate a high number of correct predictions for all millet varieties, demonstrating the strong classification capability of the developed framework [36]. The model correctly identifies 196 Finger Millet samples, 198 Pearl Millet samples, 195 Foxtail Millet samples, 196 Barnyard Millet samples, 197 Little Millet samples, 197 Kodo Millet samples, 198 Proso Millet samples, and 199 Sorghum samples. The concentration of higher values along the diagonal confirms that the learned feature representations effectively capture the unique characteristics associated with each millet category.

The classification performance demonstrates that the developed model achieves highly accurate recognition across different millet varieties with very few incorrect predictions. Minor

misclassification instances are observed between visually similar categories, such as Foxtail Millet and Barnyard Millet, where a small number of samples are incorrectly classified due to similarities in grain texture, size, and surface appearance. Similarly, limited confusion occurs among Little Millet, Kodo Millet, and Proso Millet classes because these varieties possess closely related morphological characteristics. The low number of off-diagonal values indicates that the extracted features successfully differentiate between challenging classes while reducing incorrect predictions. This behavior confirms that the network effectively learns both fine-grained texture patterns and high-level structural information from the input images.

The diagonal dominance observed in the confusion matrix highlights the effectiveness of hierarchical feature extraction and feature fusion mechanisms in improving class separation. The model demonstrates strong capability in identifying subtle variations among millet grains by utilizing comprehensive visual information such as shape, color distribution, surface texture, and spatial characteristics. The limited classification errors indicate that the learned feature space provides clear boundaries between different millet categories. The high recognition rate across all classes also suggests that the balanced dataset distribution and preprocessing strategies contributed to stable model training and improved generalization performance. The confusion matrix analysis provides additional evidence that the classification framework maintains consistent performance rather than achieving accuracy through bias toward specific classes.

The results shown in Figure 9 validate the reliability and robustness of the developed classification approach for automated millet identification [37]. The high values obtained across all diagonal elements indicate excellent agreement between actual and predicted labels, while the minimal off-diagonal values demonstrate reduced classification uncertainty. This performance confirms that the extracted deep features contain sufficient discriminative information for accurate recognition of multiple millet varieties. The confusion matrix analysis complements other evaluation measures by providing class-wise performance insights and confirming the practical suitability of the developed framework for applications such as automated grain inspection, quality assessment, and intelligent agricultural monitoring systems.

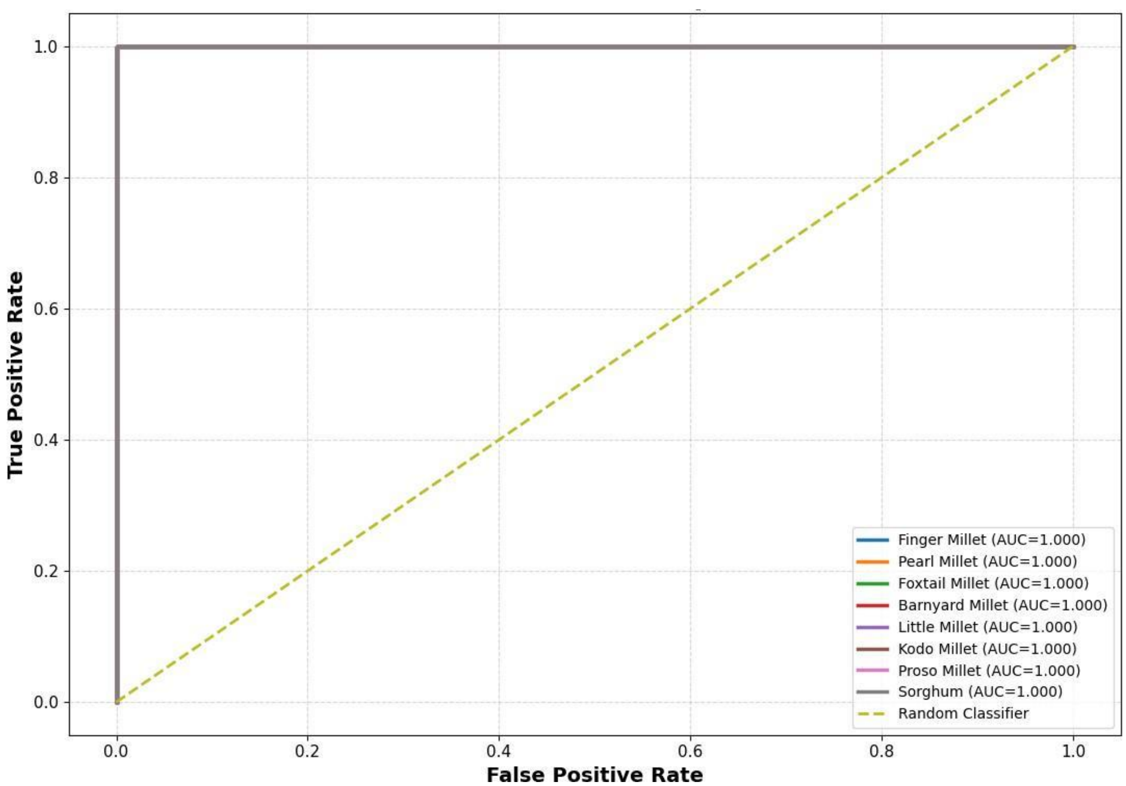

Figure 10. ROC Curve Analysis for Multi-Class Classification

Figure 10 presents the Receiver Operating Characteristic (ROC) curve analysis for evaluating the classification capability across different millet categories. The ROC curve represents the relationship between the True Positive Rate (sensitivity) and False Positive Rate under different classification threshold values. The curves corresponding to all millet classes closely approach the upper-left corner of the graph, indicating excellent discrimination capability between individual categories. The Area Under Curve (AUC) values obtained for Finger Millet, Pearl Millet, Foxtail Millet, Barnyard Millet, Little Millet, Kodo Millet, Proso Millet, and Sorghum are approximately 1.000, demonstrating near-perfect classification performance. A higher AUC value indicates that the model can effectively distinguish between positive and negative class samples with minimal classification uncertainty [38].

The ROC analysis demonstrates the strong decision-making capability of the developed classification framework in identifying different millet varieties. The sharp rise of the ROC curves toward a True Positive Rate of

1.0 with almost zero False Positive Rate indicates that the extracted feature representations provide highly separable patterns among different classes. The model successfully captures important visual characteristics such as grain texture, structural appearance, shape variations, and surface patterns, which contribute to accurate class discrimination. The consistently high AUC values across all categories confirm that the learned feature space provides reliable separation even among millet varieties with similar visual characteristics.

The comparison with the random classifier line further highlights the effectiveness of the developed model. The random classifier represents a baseline condition where classification

decisions are made without meaningful feature understanding, resulting in an AUC value close to

0.5. In contrast, the obtained ROC curves remain significantly above the random prediction boundary, confirming that the classification decisions are based on meaningful image representations. The identical high-performance behavior across all millet categories indicates that the model does not exhibit preference toward any particular class and maintains balanced recognition capability. This result was also consistent with the balanced dataset distribution and effective feature learning strategy adopted during model development.

The ROC-AUC evaluation provides additional validation of the robustness and reliability of the developed classification system. The near-perfect AUC values indicate strong sensitivity, low false-positive prediction rates, and excellent generalization ability on unseen image samples. This analysis confirms that the learned deep representations effectively capture class-specific characteristics required for accurate millet identification. The ROC curve results, combined with accuracy, precision, recall, F1-score, and confusion matrix analysis, demonstrate the reliability of the framework for automated grain inspection, quality assessment, and intelligent agricultural image analysis applications.

Table 5. Class-Wise Classification Performance Analysis

Millet Class	Precision (%)	Recall (%)	F1-Score (%)	Correct Predictions
Finger Millet	98.5	98.0	98.2	196
Pearl Millet	99.0	99.0	99.0	198
Foxtail Millet	97.8	97.5	97.6	195
Barnyard Millet	98.2	98.0	98.1	196
Little Millet	98.7	98.5	98.6	197
Kodo Millet	98.5	98.5	98.5	197
Proso Millet	99.0	99.0	99.0	198
Sorghum	99.5	99.5	99.5	199

Table 5 provides class-wise evaluation results for individual millet categories. All classes achieve high precision, recall, and F1-score values, indicating consistent recognition capability across different varieties. The results demonstrate that the developed framework effectively captures class-specific morphological and textural features, reducing confusion among visually similar millet samples.

5. Conclusion

The accurate identification and classification of millet varieties have become increasingly important for improving agricultural quality assessment, food authentication, post-harvest management, and intelligent farming applications. The development of an automated image-based classification framework provides an effective alternative to traditional manual inspection methods, which are often influenced by human expertise, environmental conditions, and visual similarity among grain varieties. The proposed deep learning framework effectively addresses these challenges by utilizing advanced feature learning and fusion strategies to extract comprehensive visual representations from millet grain images. The experimental results demonstrate that the developed approach successfully learns discriminative characteristics associated with different millet categories, including variations in grain texture, shape, surface structure, and intensity distribution. The balanced dataset distribution and preprocessing operations contribute significantly to stable training performance by ensuring consistent representation of all millet classes [39].

The integration of hierarchical feature extraction and deep feature fusion improves the ability of the model to capture complementary information from different convolutional levels. Low-level features containing texture and edge details are effectively combined with high-level semantic representations, resulting in a more informative feature space for multi-class recognition. The training and validation analysis confirms stable convergence with continuous improvement in accuracy and reduction in loss values, indicating effective optimization and strong generalization capability. The close relationship between training and validation performance demonstrates that the framework avoids excessive overfitting while maintaining reliable prediction capability on unseen samples. The feature fusion mechanism enhances the separation between visually similar millet varieties by preserving important characteristics that may be overlooked in conventional single-level feature extraction approaches.

Comprehensive performance evaluation using accuracy, precision, recall, F1-score, specificity, Matthews Correlation Coefficient, Cohen's Kappa, confusion matrix, and ROC-AUC analysis confirms the effectiveness and reliability of the proposed framework. The obtained results indicate that the model achieves highly accurate classification across all considered millet categories with minimal misclassification errors. The confusion matrix demonstrates strong class-wise recognition capability, while ROC analysis verifies excellent discrimination performance between different millet varieties. The consistently high evaluation scores indicate that the learned feature representations are robust and capable of handling complex variations present in agricultural images. These findings highlight the potential of deep feature fusion approaches for developing intelligent systems capable of performing automated grain identification with high accuracy and efficiency.

The developed framework provides a scalable solution for practical agricultural applications, including automated grain sorting, quality inspection, digital crop identification, and smart food processing systems. The combination of efficient feature extraction, feature fusion, and optimized classification enables reliable deployment in real-world environments where rapid and consistent decision-making was required. The proposed methodology contributes toward the advancement of artificial intelligence-driven agricultural technologies by reducing dependency on manual evaluation and improving the accuracy of grain classification processes. Future research can focus on expanding the dataset with field-acquired images, incorporating lightweight network architectures for edge-based deployment, integrating explainable artificial intelligence techniques for improved interpretability, and extending the framework toward real-time agricultural monitoring systems [40]. The outcomes of this research establish a strong foundation for intelligent millet analysis and demonstrate the effectiveness of deep learning-based solutions in supporting sustainable and precision-oriented agricultural practices.

REFERENCES

1. Raman, S., Soni, M., Ramaprasad, R., & Chamola, V. (2022). LWCNN: a lightweight convolutional neural network for agricultural crop protection. *Multimedia Tools and Applications*, *81*(16), 22323-22334.
2. Singh, S., Dashore, S., Prasad, J., & Naik, N. S. (2023). iCrop: a framework for identification of millet crop diseases via segmentation and feature extraction. *Indian Phytopathology*, *76*(3), 803-812.
3. Baisakh, B., Sethy, P. K., Gupta, R., Behera, S. K., & Nanthaamornphong, A. (2025). A hybrid machine vision and handcrafted features fusion based approach for fine-grained millet classification. *Scientific Reports*.
4. Vinod, D. N., Revanth, R. S., Bhaskar, B., & Prakash, P. J. (2025, January). CNNs for Early Detection of Plant Diseases: A Deep Learning Perspective. In *2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)* (pp. 977-982). IEEE.
5. Mohan, R. R., & Sukumar, V. J. (2025, January). A Modified Deep Learning Framework for Leaf Disease Detection in Millet Crops. In *2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI)* (pp. 1495-1500). IEEE.
6. Malik, A., & Jain, R. (2024, July). Smart Sorting: Redefining Millet Classification with CNN. In *Biennial International Conference on Future Learning Aspects of Mechanical Engineering* (pp. 589-599). Singapore: Springer Nature Singapore.

7. Kumar, R., Mahanty, C., Panigrahi, B. S., Patro, S. G. K., Tuan, T. M., & Son, L. H. (2025). A new method for prediction of Vigna mungo millet disease based on deep learning. *Current Plant Biology*, 100562.
8. Mishra, S., Volety, D. R., Bohra, N., Alfarhood, S., & Safran, M. (2023). A smart and sustainable framework for millet crop monitoring equipped with disease detection using enhanced predictive intelligence. *Alexandria Engineering Journal*, *83*, 298-306.
9. Alotaibi, Y., Rajendran, B., & Rajendran, S. (2024). Dipper throated optimization with deep convolutional neural network-based crop classification for remote sensing image analysis. *PeerJ Computer Science*, *10*, e1828.
10. Fonseca, Y., Bautista, C., Pardo-Beainy, C., & Parra, C. (2023). A plum selection system that uses a multi-class Convolutional Neural Network (CNN). *Journal of Agriculture and Food Research*, *14*, 100793.
11. Divya, A., Sungeetha, D., & Ramesh, S. (2024). Efficient protection of golden nutri cereal implementing enhanced disease identification injecting hybrid split-attention mechanism through Novel SANDNES Mechanism. *Multimedia Tools and Applications*, *83*(38), 85499-85521.
12. Elfatimi, E., Eryiğit, R., & Elfatimi, L. (2025). Automated End-to-End Deep Learning Framework for Complex Multiclass Brassica Seed Classification. *Seeds*, *4*(4), 67.
13. Lamba, S., Kukreja, V., Rashid, J., Gadekallu, T. R., Kim, J., Baliyan, A., ... & Saini, S. (2023). A novel fine-tuned deep-learning-based multi-class classifier for severity of paddy leaf diseases. *Frontiers in Plant Science*, *14*, 1234067.
14. Manikandakumar, M., & Karthikeyan, P. (2023). Weed classification using particle swarm optimization and deep learning models. *Computer Systems Science & Engineering*, *44*(1).
15. Prasad, K. V., Vaidya, H., Rajashekhar, C., Karekal, K. S., Sali, R., & Nisar, K. S. (2024). Multiclass classification of diseased grape leaf identification using deep convolutional neural network (DCNN) classifier. *Scientific reports*, *14*(1), 9002.
16. Hasmukhbhai, B. A. (2024). *Classic Algorithm for Organic Cereals* (Doctoral dissertation, Gujarat Technological University).
17. Jadhav, S., & Lal, A. M. (2022). Multi-class plant leaf disease detection using a deep convolutional neural network. *International Journal of Information System Modeling and Design (IJISMD)*, *13*(1), 1-14.
18. Bhensjaliya, A. H., & Dalwadi, D. C. (2022). Performance Analysis of Cereals Grains using Neural Network & Multi Support Vector Machine. *Indian Journal of Science and Technology*, *15*(37), 1876-1883.
19. Tiwari, S., Gehlot, A., Singh, R., Twala, B., & Priyadarshi, N. (2024). Design of an iterative method for disease prediction in finger millet leaves using graph networks, dyna networks, autoencoders, and recurrent neural networks. *Results in Engineering*, *24*, 103301.
20. Srinivasan, R. (2025, August). Support Vector Machine-based Early Detection of Pearl Millet Diseases for Improved Crop Yield. In *2025 5th International Conference on Soft Computing for Security Applications (ICSCSA)* (pp. 2165-2169). IEEE.
21. Goyal, P., Verma, D. K., & Kumar, S. (2025). Detection and monitoring for enhanced prevention of grain plant disease using classification-based deep ensemble neural networks in smart agriculture. *International Journal of Remote Sensing*, *46*(5), 1992-2022.
22. Petrov, I., Smirnov, A., & Volkov, D. (2023). Region-Based Convolutional Neural Network with Radial Gradient Index Filtering for Automated Red Blood Cell Detection and Multi-Class Classification in Microscopic Images.
23. Sajitha, P., Andrushia, D. A., & Suni, S. S. (2023, August). Multi-class plant leaf disease classification on real-time images using YOLO V7. In *International Conference on Image Processing and Capsule Networks* (pp. 475-489). Singapore: Springer Nature Singapore.
24. Oyege, I., Sibitenda, H., & Bhaskar, M. S. B. (2024). Deep learning applications for real-time and early detection of fall armyworm, African armyworm, and maize stem borer. *Machine Learning with Applications*, *18*, 100596.
25. Ramirez, E., Ruiperez-Campillo, S., Casado-Arroyo, R., Merino, J. L., Vogt, J. E., Castells, F., & Millet, J. (2024). The art of selecting the ECG input in neural networks to classify heart diseases: a dual focus on maximizing information and reducing redundancy. *Frontiers in physiology*, *15*, 1452829.
26. Jain, N., Joshi, A., & Earles, M. (2025). iNatAg: Multi-Class Classification Models Enabled by a Large-Scale Benchmark Dataset with 4.7 M Images of 2,959 Crop and Weed Species. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 5501-5510).
27. Khouma, O., Bakhoum, A. S., & Ndiaye, M. M. (2025, December). Artificial Intelligence-driven Millet Pest Identification through Damage Assessment. In *2025 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 2195-2201). IEEE.
28. Prajwal, B., Mondal, S., Raghu, M., & Shashank, N. (2025, March). Disease Detection in Finger Millet Plant Using Fine-Tuned Transfer Learning Models. In *2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF)* (pp. 1-6). IEEE.
29. Qian, Y., Kok, Y. E., Janes, G., Nakombo-Gbassault, P., Atkinson, B., Atkinson, J., ... & Pound, M. P. (2025). Root Segmentation Beyond Species Boundaries: A Generalizable Framework for Anatomical Analysis. *Plant Phenomics*, 100146.
30. Hazarika, R. A., Maji, A. K., Kandar, D., Jasinska, E., Krejci, P., Leonowicz, Z., & Jasinski, M. (2023). An approach for classification of Alzheimer's disease using deep neural network and brain magnetic resonance imaging (MRI). *Electronics*, *12*(3), 676.

31. Sinapan, I., Lin-Kwong-Chon, C., Damour, C., Kadjo, J. J. A., & Benne, M. (2024). Revealing anodic multi-class bubble dynamics in PEMWE systems using deep learning and post-processing detection. *Fuel*, *364*, 131112.
32. Ors-Quixal, R. T., Ramírez-Candela, E., Ruipérez-Campillo, S., Castells-Ramón, F., & Millet, J. (2023, October). 3D CNN as an Approach to Predict the Cerebral Performance of Comatose Patients. In *2023 Computing in Cardiology (CinC)* (Vol. 50, pp. 1-4). IEEE.
33. Tomar, A., Kumar, S., & Tripathi, K. (2025). Review of herbarium plant identification of crop wild relatives using convolutional neural network models. *Genetic Resources and Crop Evolution*, *72*(8), 9165-9178.
34. Ors, R. T., Ramırez, E., Ruiperez-Campillo, S., Castells, F., & Millet, J. 3D CNN as an Approach to Predict the Cerebral Performance of Comatose Patients.
35. He, L., Wei, R., Ding, Y., Huang, J., Wei, X., Li, R., ... & Fu, L. (2025). Enhancing dense-scene millet appearance quality inspection based on YOLO11s with overlap-partitioning strategy for procurement. *Agronomy*, *15*(6), 1284.
36. Hayajneh, A. M., Batayneh, S., Alzoubi, E., & Alwedyan, M. (2023). TinyML olive fruit variety classification by means of convolutional neural networks on IoT edge devices. *AgriEngineering*, *5*(4), 2266-2283.
37. Ouf, N. S. (2023). Leguminous seeds detection based on convolutional neural networks: Comparison of faster R-CNN and YOLOv4 on a small custom dataset. *Artificial Intelligence in Agriculture*, *8*, 30-45.
38. Golla, A. K., Lorenz, C., Buerger, C., Lossau, T., Klinder, T., Mutze, S., ... & Goelz, L. (2023). Cervical spine fracture detection in computed tomography using convolutional neural networks. *Physics in Medicine & Biology*, *68*(11), 115010.
39. Yang, S., Zheng, L., Wu, T., Sun, S., Zhang, M., Li, M., & Wang, M. (2024). High-throughput soybean pods high-quality segmentation and seed-per-pod estimation for soybean plant breeding. *Engineering applications of artificial intelligence*, *129*, 107580.
40. Ong, J. H., Ong, P., & Lee, W. K. (2022). Image-based oil palm leaves disease detection using convolutional neural network. *Journal of Information and Communication Technology*, *21*(3), 383-410.