doi.org/10.54473/IJTRET.2026.10406
International Journal of Trendy Research in Engineering and Technology
Volume 10 Issue 4 August 2026
ISSN No. 2582-0958

COLOR SPACE FUSION IN DEEP LEARNING MODELS FOR BIOMETRIC FACE IDENTIFICATION AND PLANT DISEASE CLASSIFICATION

[1]Diler N Abdulqader, [2]Pawan Shivan Othman
[1]Departmentof Computer and Communication Engineering, College of Engineering, Nawroz University
[2]Department of Computer Science, College of Science, Nawroz University, Duhok, Iraq
diler.abdulqader@nawroz.edu.krd[1], pawan.othman@nawroz.edu.krd[2]

Received 01 August 2026 Received in revised form 03 August 2026 Accepted 04 August 2026
Available Online 06 August 2026

ABSTRACT

Visual classification is a challenging computer vision problem, with issues of lighting variation and color distribution, as well as feature representation. It has been applied in different sectors, such as biometric identification of faces and diagnosis of plant diseases.. Although effective feature learning is necessary for improving classification accuracy, some available approaches depend on unique color-space representations and fail to gather complementary visual data across domains. This is a problem since it is essential for improving classification accuracy. This study proposes a unified deep learning architecture that makes use of multi-color-space feature fusion in order to improve classification efficiency across a wide range of visual tasks. In order to construct a parallel feature-extraction framework that incorporates RGB, HSV, and YCbCr representations, the suggested approach makes use of ResNet-18 branches. To the best of our knowledge, this combination offers a succinct and domain-agnostic technique for multi-domain visual categorization. It does this by simultaneously accumulating luminance and chrominance characteristics, which strengthens the representation of discriminative features. For the purposes of biometric identification and illness classification, a dataset consisting of photographs of face features and plant leaves that was available to the public was employed. Following the completion of the typical preprocessing of the photographs, deep feature embeddings were retrieved from each color-space branch and then fused together through the usage of concatenation method. A number of performance criteria, such as accuracy, sensitivity, and specificity, were utilized in order to assess three different categorization situations. Single color space baselines, dual space combinations, and total fusion were the possibilities for these scenarios. The results showed that the classification was better when multi-color representation (MC) was used. The lowest accuracy was obtained for face recognition with 52.43% accuracy while the maximum accuracy was obtained for plant disease classification with 94.50% accuracy. For a number of statistical studies, scenarios, it is possible to have different color representations and categorize the images correctly. The proposed approach is scalable, generalized and is computationally efficient. This approach may be used to enhance CNN classification systems in all kinds of imaging applications.

Keywords: Color Space Fusion, Deep Learning, Plant Disease Classification, Face Identification, ResNet-18,Feature-Level Fusion

1. INTRODUCTION

Visual categorization has become a challenging problem in computer vision because of the variations in lighting conditions, color distribution, texture patterns and feature representation in the visual processing of several applications such as biometric systems for face recognition and plant disease diagnosis [2],[5]. In unfettered imaging environments, effective feature extraction is crucial in order to get the maximum accuracy in classification. Current methods tend to rely on a single colorspace representation, but do not necessarily provide sufficient visual information to accurately discriminate between different datasets [4, 7]. To address these limitations, state-of-the-art deep learning frameworks have been devised. Such frameworks combine a number of feature representations in order to improve generalizability and robustness. Convolutional Neural Networks (CNNs) and other deep learning techniques have been shown to have impressive performance in visual recognition problems recently [1,6,8]. This is because CNNs can learn the hierarchical representation of features directly from the raw image information. Although the models with CNNs for biometric facial recognition system have come a long way, they are still prone to various problems such as postures, lighting conditions, obstructions, and age [13, 9, 10].The subtle difference in leaf texture, color, lesion shape and size in the classification of plant diseases is a significant difficulty, especially with varying conditions of acquisition and environmental disturbances [3].

The difficulties highlight the need for more unique and robust feature representations that can transfer more knowledge to different domains. Moreover, the commonly used RGB color space used in CNN-based systems for traditional color representation does not provide a satisfactory separation of luminance and chrominance data. This leads to models that depend only on RGB input signals losing important discriminative features when the ambient light varies or when class-specific features are mostly encoded in the saturation or color intensity [14, 15]. The other color spaces like YCbCr and HSV offer better organized representation by decomposing the color components as hue, saturation, and luminance, thus improving the interpretability and robustness of the color space [16], [17]. Many works have revealed that the performance of classification can be improved through an image change in color spaces. Many methods, however, are still based on one particular representation, or are still based on performing a transformation without benefiting from the complimentary features [18, 19]. More recent studies also have considered the integration of multiple feature representations via the use of fusion mechanisms. The focus of this research investigation has been the context for deep learning frameworks. Although these techniques have proven to be effective in specific areas, they tend to focus on one activity only and might not be as flexible in the application of various situations. This is even while they have shown some promising results in some areas. This leaves a gap in the current literature for coherent frameworks that can, whilst being compact, calculate at reasonable speed, be able to input a number of color space representations and do so in a computational efficient manner. Moreover, Feature Level Fusion as a general technique to boost CNN categorisation remains largely unexplored [20, 21].Previous research has explored the efficiency of color space conversion or multi-input architectures, however, a comprehensive study and comparison of the efficiency of concurrent multi-branch fusion of RGB, HSV, and YCbCr across fundamentally different classification domains has not yet been fully explored. This gap is clearly marked as a methodological gap. Unlike many color models that use only a single chromatic space, the model is capable of capturing complementary chromatic and brightness data at the same time, by processing different color spaces in parallel. Moreover, feature concatenation at the feature level can be used to create more complex and discriminative embeddings. This approach is expected to improve the resistance of classification, as it reduces the amount of information lost due to the representations being done in one space, whilst enhancing the variety of characteristics. Based on this result, the following research proposes a multi-branch deep learning system with parallel ResNet-18 [8] architectures. This system features RGB color space, HSV color space and YCbCr color space. The feature embeddings of each branch are fused at the feature level and then classified by a fully connected SoftMax layer. To the best of our knowledge this work is one of the earliest studies to consider color-space fusion as a representation strategy, which is transferable to biometric and agricultural visual classification tasks. The method is designed to enhance the distinguishing ability while maintaining the simplicity of the structure and the efficiency of calculation, suitable for the deployment of real-world applications. This work makes the following contributions:

1. A unified multi-branch architecture with RGB, HSV, and YCbCr representations improves feature learning.
2. An extensive assessment contrasting single-space and fused-space setups across various datasets and performance indicators.
3. An application of color-space fusion in the fields of biometric face identification and plant disease classification is demonstrated through a validation that spans multiple domains.

2. RELATED WORK

Many of the computational techniques have been proposed for image classification problems such as biometric face identification and plant disease recognition and involve various combinations of visual, textural and colour features. The focus of early research in the context of class differentiation was on the use of carefully handcrafted spatial-domain features such as geometric face descriptors, texture statistics and colour histograms [1, 6]. Such methodologies were adopted in implementing the idea which later developed into a quantitative approach in order to make meaningful distinctions in making classification. However, their inability to handle the feature interactions that require non-linear processing and sensitivity to lighting conditions and changes in stances and scales was significant drawbacks. Two spatial descriptors which have been widely used for capturing structural and textural changes in plant leaves and facial images are Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG) [6], [8]. Moreover, statistical representations based on the intensity distribution of pixels have been used to represent discriminative visual patterns.

While these methods work to a certain degree, they only take advantage of the intensity information in spatial space, neglecting the chromatic changes that occur when the differences between classes are encoded primarily in color changes (e.g. symptoms of plant diseases, light-sensitive facial features).

Thus, color-based representations have been gaining more focus as an alternative to purely spatial descriptors. Typically, the RGB color space is used; however, the RGB color space is not explicit about separating the luminance information from the chrominance information. In the realm of visual classification pursuits, such as biometric face identification and plant diseases identification techniques, many computational methods have been proposed, emphasizing different aspects of the visual, textural and color aspects In the early stages of class differentiation research, the main method used was careful application of spatial-domain data, which had been carefully handcrafted, included geometric face descriptors, texture statistics and colour histograms [1, 7]. The methodologies allowed the idea that visual pattern information could be used quantitatively and not just qualitatively, and that this quantitative aspect would be able to offer a rich discriminative element in the classification task to come. However, their success rate for the most complex nonlinear feature interactions, and the sensitivity to variations in illumination and stance/class was a major disadvantage. Though they are very accurate, they might rely on a large set of annotated data and be demanding in terms of computing resources, and could also suffer from some interpretation problems. This limitation applies to when the data supply is limited in real-life applications, or where there is a high demand for computational performance. In search of the solutions to these difficulties, various works have been proposed to adapt some hybrid or multi-branch architectures, although many of these studies do not extend their validity to other application areas..

With the development of traditional method and deep-learning method, it is found that there is still a great lack in the unified feature-level fusion strategy for complementary color representation. While the study of color transformation and multi-input architectures has been studied separately, there is a big gap in the literature that combines both concepts of RGB, HSV, and YCbCr representations into a unified and simplified framework, and evaluates their performance in various fields. This limitation limits their ability to take maximum advantage of chromatic diversity and reduces their ability to maintain high quality with varying imaging conditions. Based on these findings, this research proposes a multi-branch deep neural network for fusing RGB, HSV and YCbCr color spaces in the feature level. The proposed technique aims to improve the extraction of distinguishing characteristics and to improve the classification effectiveness, both in biometric and agricultural fields, by utilizing complementary luma and chroma information.

2.1 Comparative review of existing color-space and deep learning approaches

Methods are either domain-specific or utilize single color-space representations, restricting their application to alternative datasets. The proposed deep learning architecture enhances discrimination and computational efficiency by integrating RGB, HSV, and YCbCr feature-level color representations.

Table 1. **Review:** Plant disease categorization and face recognition: Review

Research/ Year	Data	Features Used	Classification Method	Accuracy Rate	Limitations
Huang et al., 2007 [13]	Facial images	Raw images (LFW dataset)	—	—	Unconstrained data, no feature extraction
Hughes & Salathé, 2015 [12]	Leaf images	PlantVillage dataset	—	—	Controlled environment bias
Schroff et al., 2015 [24]	Facial images	Deep embeddings (FaceNet)	CNN	~99%	Sensitive to pose and illumination
Bianco et al., 2015 [19]	General images	RGB-based features	CNN	Moderate	Illumination sensitivity
Mohanty et al., 2016 [11]	Leaf images	RGB-based CNN features	CNN	~99%	Poor real-world generalization

Masi et al., 2018 [10]	Facial images	Transfer learning features	CNN	High	Requires large-scale training data
Chen et al., 2019 [20]	General images	Multi-feature fusion	CNN	Improved	Domain-specific tuning
Liu et al., 2020 [21]	General images	Fusion representations	CNN	High	Not validated across domains
Shah et al., 2021 [23]	Facial images	Ensemble-based features	Ensemble	Improved	Increased computational complexity
Eleyan et al., 2022 [22]	Facial images	Statistical local descriptors	Ensemble	Improved	Limited deep feature representation
Dai et al., 2023 [24]	Leaf images	Attention + augmentation	CNN	High	Sensitive to color variations
Proposed Method	Face + Leaf images	RGB + HSV + YCbCr (fusion)	CNN (ResNet-18)	52.43% / 94.50%	Lightweight, domain-agnostic

2.2 Conclusion of gaps in this literature

Early articles used RGB lacking fully leveraging complementary data in several color spaces. Many methods neglect crucial hue changes for visual categorization, especially under different lighting conditions. Most articles have not studied feature-level fusion approaches that use several color spaces to improve discrimination. Another issue is that recommended models' generalizability is not usually assessed. Many models are designed for plant disease classification or facial recognition but no other tasks. This restricts these techniques' generalizability and integration. Deep-learning architectures have demonstrated excellent performance, but they require substantially annotated datasets and may be resource-intensive. Meanwhile, there hasn't been nearly enough research on how to incorporate various color representations into efficient and compact designs. There is a dearth of systematic comparison with single-space baselines, and existing fusion-based methods are frequently sophisticated and domain-specific. Due to these restrictions, a unified, computationally efficient framework is required for visual classification tasks that involve different types of data and that can make use of complementary color-space information without sacrificing generalizability.

To address these shortcomings, we provide a multi-branch color-space fusion system that combines RGB, HSV, and YCbCr representations to boost classification accuracy and feature diversity in biometric and agricultural applications.

3. MATERIALS

Two data sets of images were taken from known benchmark sources and shared with the public and used in this work. The datasets used were Labeled Faces in the Wild (LFW) dataset for face identification and PlantVillage dataset for plant disease classification. You can search for "tagged" photos of humans and plants in the databases. The images are several classes that overlap both domains as shown in Figure 1. The datasets were selected because they are widely used in recent studies on computer vision and deep learning [13] and accepted by the community. All photographs are real world photos taken in a range of different and uncontrolled situations. This provides a realistic ground to validate the resilience of the colour space mixture framework developed.

(a) (b)

Figure 1 Examples from the datasets. (a) provides a sample facial image from the LFW dataset, (b) contains plant leaf image from the PlantVillage dataset.

The datasets are illustrated in Figure 1 by a number of examples.
Subfigure (a) displays an example image from the LFW dataset showing a face, and subfigure (b) displays an example image from the PlantVillage dataset showing a plant leaf. Two instances of face pictures may be seen in these subfigures. According to the descriptions of the datasets, the photos were preprocessed and curated to ensure consistency and usability for study and are sourced from various fields and internet locations. The datasets are not fully annotated with detailed metadata, including environmental conditions and full demographic information, but rather they are composed of varying poses, lighting, background and texture. This constraint is acknowledged and discussed later in the study as constraints of the study. Because only publicly-disseminated data was used under research-permissive license, we were able to address issues of ethics. In addition, the data collection did not involve collecting any more information, and the identifiable information was limited to the facial information which could be seen. All experimental approaches followed ethical rules for secondary data exploitation, with conformance of anonymization and non-identifiability rules [16]. To ensure each picture was processed on the same basis, it was first converted into RGB, YCbCr, and HSV color spaces and then resized to 224 × 224 pixels before feature extraction.

4. COLOR-SPACE FEATURE EXTRACTION AND FUSION FRAMEWORK

Before feature extraction, all images were compressed to 224 × 224 pixels and converted to RGB, HSV, and YCbCr color schemes. Color representation is a crucial part of visual recognition tasks because separate color spaces hold complementary information. Despite the fact that it does not explicitly distinguish between luminance and chrominance components, the RGB color space is sensitive to changes in illumination [14], [18]. This is because it preserves the original pixel intensities. In comparison, the HSV color space expresses color data in terms of value, hue and saturation. This allows for a more precise separation of chromatic components from variations of intensity [15], [16].

4.1 Color-space transformation

In a similar vein, the YCbCr color space distinguishes luminance (Y) from chrominance contents (Cb and Cr), which has been shown to improve resilience in activities affected by manipulations in color and light distortion [17].
When an input picture $I(x,y)$ is supplied, the following three transformed representations are produced:

$$I_c = f_c(I), \quad c \in \{RGB, HSV, YCbCr\}$$

where the transformation operator for each color space is written as $fc(\cdot)$. These changes provide three representations that work well together: RGB keeps the information about the shape and texture. HSV is a method that separates hue and saturation, which makes it easier to tell colors apart. The separation of luminance and chrominance that YCbCr offers enhances lighting invariance.

This approach of multi-representation allows you to obtain more complete feature distributions than approaches based on the use of a single space. These changes enable the model to have complementary features that are not discernable from any one of the color-space representations. In past research, the use of multi-color representation was found to greatly enhance classification accuracy and feature diversity in plant disease detection, and other applications such as face recognition[21].

4.1.1 RGB color representation

The most common representation used in deep learning applications and computer vision is RGB, and is used the most frequently and with the highest frequency. The original appearance of the image is maintained and explicit intensity levels for red, green and blue are encoded. RGB is especially sensitive towards changes in illumination and might not be able to record discriminative chromatic features even if there's changes in illumination [15]. This is despite its simplicity and the fact that it can be used with CNN architectures. The colour-space conversions used in this study are shown in figure 2. It highlights the visual properties associated with each of the RGB, HSV and YCbCr color spaces, and the difference between them. Then, the changes of color spaces are demonstrated in (a) RGB picture, (b) HSV representation, and (c) YCbCr representation as shown in Figure 2.

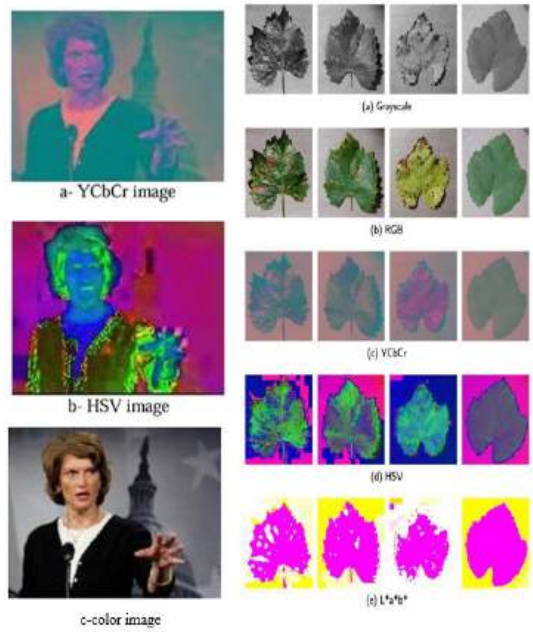


FIGURE 2. Color-space transformations: (a) YCbCr representation, (b) HSV representation, (c) RGB image

4.1.2 HSV color space

HSV is the color space that is broken up into three components: hue, saturation and value. This space is for the separation of colour information. To better tell the difference between color-based features, this style is used. This is particularly helpful if differences in classes are indicated by hue or saturation differences. There are several studies that demonstrate the benefit of the HSV traits when dealing with medical and agricultural images for classification [16].

$$H = \arccos\left(\frac{1}{2}\cdot\frac{(R-G)+(R-B)}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right)$$

HSV is especially suitable for recording colour variations which are not influenced by variations in light, because of its ability to separate chromatic components from intensity more efficiently.

4.1.3 YCbCr color representation

The YCbCr color space is useful for light changes, particularly.YCbCR color space is useful for light changes, especially. YCbCr is useful for extracting features in different lighting situations since it separates brightness from the color. This feature has been extensively used in applications like video processing systems and facial recognition [17], [18].

$$Y = 0.299R + 0.587G + 0.114B$$

$$Cb = B - Y, \quad Cr = R - Y$$

4.2 CNN architectures for deep feature extraction

Convolutional neural networks (CNNs) and other deep learning models have established higher performance in visual classification tests due to their learning ability to hierarchical feature representations [2-7]. For the purpose of feature extraction, ResNet-18 is used as the reference architecture in this study. Through the implementation of residual connections, ResNet is able to solve the issue of fading gradients and make it possible to train larger networks without compromising their performance or efficiency [8].

$$F_{RGB}, \quad F_{HSV}, \quad F_{YCbCr}$$

Each color-space representation is processed individually by a ResNet-18 branch that is specifically designated for this purpose. The feature embeddings that were derived from the three branches are denoted by the letters FRGB, FHSV, and FYCbCr. These initials are utilized throughout the text. From the input image, they (the feature vectors) extract complimentary elements to use in their analysis.

This category contains characteristics that are not connected to lighting, chromatic changes, or structural patterns in any way. It is clear from looking at Figure 3 that the ResNet-18 design takes use of a hierarchical feature extraction process.

The utilization of this method brings attention to the fact that feature representations go through incremental alterations as the network depth grows. Some examples of low-level patterns are edges, gradients, and texture primitives. The earliest convolutional layers are responsible for the majority of the processing of these patterns. These patterns could be able to explain the input picture the majority of the time.

Throughout the course of a deep neural network, successive layers combine this local input into representations that have a greater degree of abstract and semantic value. Because of this, the model is able to recognize patterns that are pertinent to the classification job and that are unique to each class. These hierarchical learning strategies allow the network to gradually modify its feature representations, which in turn increases the network's ability to discriminate between different types of data.

Activation maps are depicted in Figure 3, and they are a representation of the outcomes of convolutional operations that were carried out inside the ResNet-18 architecture. These maps illustrate the process by which spatial information is turned into higher-level feature embeddings by use of a series of nonlinear transformations.

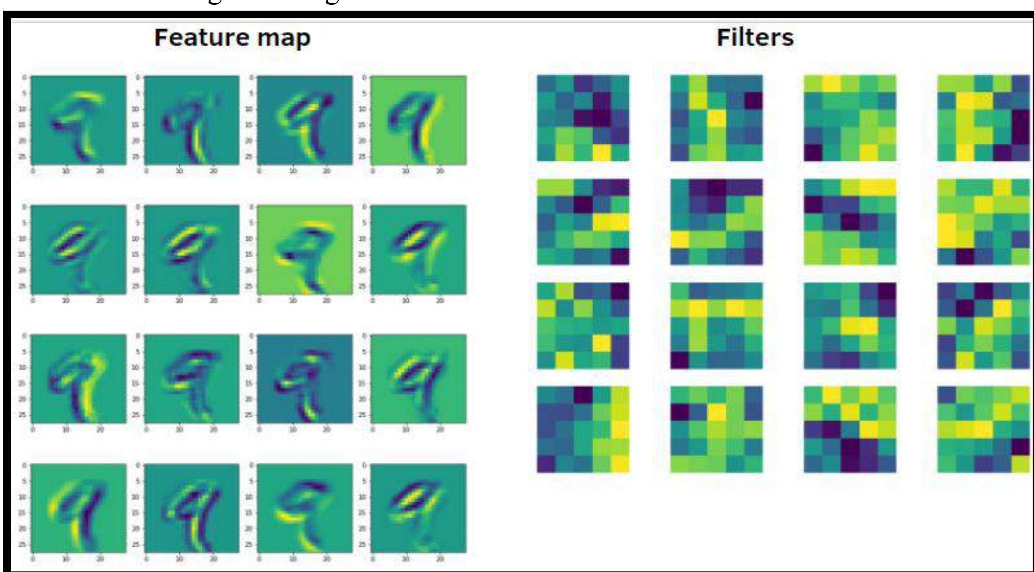


Figure 3 Intermediate Activation Maps

The intermediate activation maps associated with ResNet-18 network are shown in Figure 3. As one moves further and further into the network, the emphasis changes from things like texture and boundaries to more abstract concepts like semantics: as illustrated on these maps.

4.2.1 Feature extraction using resnet-18

The neural network known as ResNet-18 is composed of a number of convolutional layers that are composed of residual blocks. Identity mappings are utilized in order to maintain information between levels. This architecture is therefore well suited for multi-branch designs [8] given that it is effective at learning features and retains computational economy.

$$y = F(x, W) + x$$

The learnt residual function is represented by the expression F(x,W,), and the input feature map by x. This formulation, beyond conveying low level information between layers, allows to propagate gradients efficiently and therefore enhances stability in feature learning.

ISSN No. 2582-0958

4.2.2 Global feature extruction

At the output of each branch, global average pooling is used in order to lower the dimensionality of the feature maps that have been extracted. This technique packs the spatial data into short feature vectors, but preserves the meaningful features that make a difference between one thing and another. On top of that, global pooling helps increase generalization performance and decreases instances of overfitting.

$$F = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} f(i, j)$$

This process not only dimensionality reduces the data but also combines all the spatial information in a single feature vector that retains the discriminative properties of the data. In addition to this, it eliminates completely linked layers that have a high number of parameters, which helps to avoid overfitting.
.

4.3 Proposed method: multi-branch color-space fusion

The use of feature-level fusion is for including complimentary information from a number of different color spaces. On the other hand, feature-level fusion acts on representation at intermediate level, enables the model to simultaneously acquire correlations among diverse spaces of feature. This is in contrast to the fusion of decision-level, which is leading to merging the outputs after the classification is done. Fused feature representation definition: representations, which

$$F_{fusion} = [F_{RGB} \oplus F_{HSV} \oplus F_{YCbCr}]$$

Concatenation is represented by the symbol $\oplus$. This approach creates a higher-dimensional feature vector. This vector preserves each color-space representation while allowing the classifier to handle cross-channel associations. Representationally, concatenation promotes feature variety and class separability in the learned feature space. The concatenation keeps feature expressive and lets the classifier to assign adaptive significance to the subsets of features during the training, unlike fusion based averaging, which sometimes defeat discriminative components. Empirical studies suggest that feature-level integration improves categorization in visual tasks that are complex [20]. Figure 4 shows framework design of the proposed multi-branch fusion. The input image is processed by parallel ResNet-18 branches after being transformed to RGB, HSV, and YCbCr, , then use feature concatenation for fusing before applying classification. The proposed four-branch color-space fusion scheme is presented in Figure 4. Before classification, the input image is transformed into RGB, HSV, and YCbCr representations, processed by parallel ResNet-18 branches, then fused using feature concatenation.

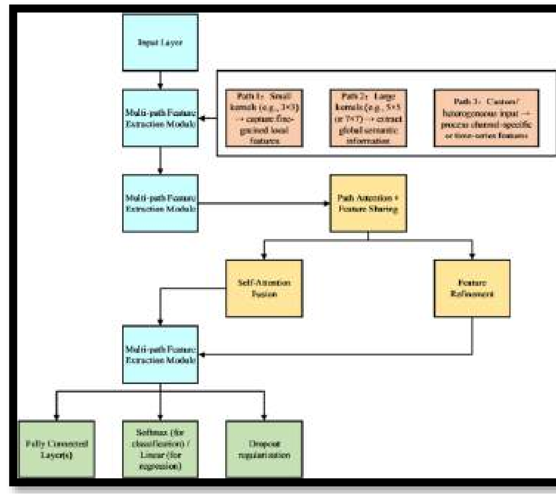


FIGURE 4. A proposed multi-branch color space fusion architecture.

After converting the input image to RGB, HSV and YCbCr space it is passed through parallel ResNet-18 branches, in this stage the classification is performed by concatenating the feature.

4.3.1 The strategy of feature concatination

Concatenation is preferred over averaging/summing because it maintains feature vector dimensionality. Thus, the model is able to learn complex relationships between attributes in various color spaces..

4.3.2 SOFTMAX-Based classification

After a completely linked layer, a SoftMax activation function is used to complete the classification process, which results in the final classification:

$$P(y = k \mid x) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}}$$

The zk is the logit for class k, whereas. This formulation ensures the guaranteed probabilistic output, and facilitates multi-class classification. This formulation guarantees the probabilistic interpretation of the result and enables performing multi-class classification.

International Journal of Trendy Research in Engineering and Technology
Volume 10 Issue 4 August 2026
ISSN No. 2582-0958

5. FEATURE NORMALIZATION AND CLASSIFICATION

To ensure uniformity and to compare with all the extracted feature representations, the feature vectors extracted from the multi-branch color-space fusion architecture were passed thru min-max normalization. This is important to correct inconsistencies in scale of features created from different color-space branches. This ensures that nobody's particular quality take precedence over the learning process. In mathematical terms, the normalizing process refers to the following:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

where x is the original value of the feature and $xmin$ and $xmax$ are the minimal and maximum values in the feature vector, respectively. Each feature number is changed by this transformation so that they are all between 0 and 1. Because of this, the model is trained faster, and the numbers are more accurate.

It is very important to normalize in order to get a good mix between the RGB, HSV, and YCbCr representations. Normalization makes sure that all color spaces create the same fused feature spaces, even if the features they create have different ranges because of changes in how statistics work. Deep learning design has a decision layer that is built on SoftMax after normalization. This layer sorts the data into groups. The SoftMax method turns the raw network outputs into probability distributions across the class names.
This lets you sort things into more than one category.

This is what the SoftMax function looks like when is defined:

$$P(y = k \mid x) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}}$$

The number K is the number of courses, and zk is the logit for class k.The following conditions must be met for the chance of the output:

$$\sum_{k=1}^{K} P(y = k \mid x) = 1$$

The suggested framework is organized and iterative, as shown in Figure 5. Each stage helps to improve the ability to tell features apart and the performance of classification. Combining multi-color representations with deep feature extraction provides the model with more information. This makes it more robust in a variety of visual conditions.

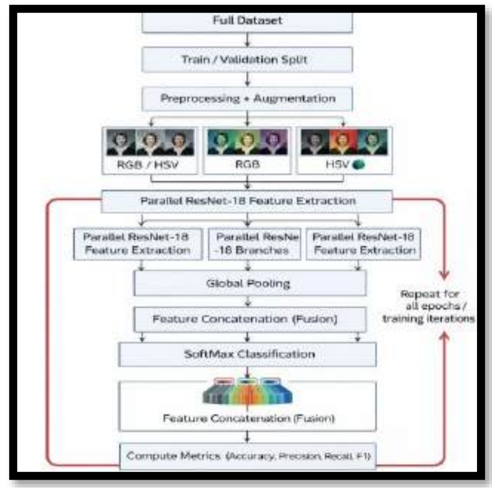


FIGURE 5. The proposed classification architecture is illustrated in a flowchart, showing how to prepare, preprocess and supplement the data, convert the color space (from RGB to HSV to YCbCr), extract features using parallel ResNet-18 branches, pool features globally, fuse features at the feature level, and classify using SoftMax. The method is performed several times throughout training epochs to make sure that performance is measured accurately.

6. EVALUATION METRICS

The effectiveness of the suggested framework was evaluated using the standard classification measures including accuracy, precision, recall (sensitivity), and F1-score. These are obtained from the confusion matrix. TP, TN, FP and FN are true positive, true negative, false positive and false negative respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}}$$

With these metrics, you can get a full picture of the model's success by looking at both its general correctness and its performance in each class.

7. QUANTITATIVE RESULTS.

The suggested model's numeric performance is shown in Table 2 for different color-space representations. This shows how useful feature-level fusion is compared to single-space models. All evaluation measures, such as accuracy, precision, recall, and F1-score, show a constant improvement. This suggests that adding complimentary color information improves the model's ability to tell the difference between things.

Table 2. Table 2. A look at how well single-space models and the fused model work

Dataset	Color Space	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Face Identification	RGB	45.83	11.04	11.54	9.79
Face Identification	HSV	47.62	12.35	13.02	10.21
Face Identification	YCbCr	49.18	13.87	14.33	11.45
Face Identification	Fusion	52.43	15.28	16.85	13.95
Plant Disease	RGB	90.00	89.20	90.50	89.80
Plant Disease	HSV	91.50	90.80	92.00	91.40
Plant Disease	YCbCr	92.30	91.60	92.80	92.20
Plant Disease	Fusion	94.50	93.80	95.00	94.40

The results in Table 2 make it very clear that the suggested feature-level fusion structure works better than individual color-space representations. All evaluation measures, such as accuracy, precision, recall, and F1-score, show a constant improvement. It looks like similar color information helps the model tell the difference.
It can recognize faces 52.43% of the time, which is 6.60% better than RGB (45.83%), 4.81 better than HSV (47.62%), and 3.25 better than YCbCr (49.18%). Since LFW has a big collection and a lot of classes, its absolute accuracy is smaller. In contrast, combining color-space features that are complementary improves the representation of features, even in tasks that focus on structure. More progress has been made in classifying plant illnesses. RGB, HSV, and YCbCr are all not as good as the fusion model, which gets it right 94.50% of the time. This big rise can be explained by the fact that plant diseases usually show up as discolored lesions, chlorosis, and different border lines [24, 28]. The fusion model successfully combines complementary data from the RGB, HSV, and YCbCr color spaces to keep structure patterns, improve chromatic discrimination, and give representations that aren't affected by lighting. Putting these traits together makes the embedding space bigger, which makes it easier to tell the difference between classes and improves classification success [20, 29]. Figure 6 is a comparison accuracy graph that helps to show and support the performance trends in Table 2.

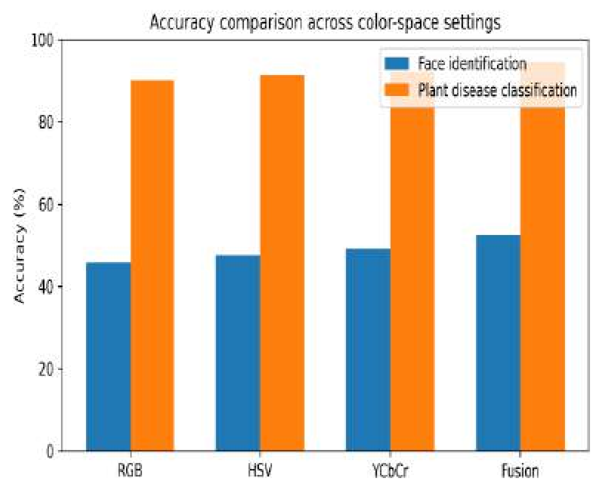


Figure 6 shows a comparison of the accuracy of RGB, HSV, YCbCr, and the combined features across different datasets. The precision of RGB, HSV, YCbCr, and combined forms is shown in Figure 6. It has been shown that mixing color spaces works because the fusion model always does improve. It is getting better at classifying plant diseases when the fusion model and separate color spaces don't agree. This shows that color traits are the most important in putting illnesses into groups. Face identification, on the other hand, has been getting better over time more slowly. This suggests that structure traits are more important than color information in biometric recognition. In any case, the steady rise shows that color-space fusion helps, even in tasks that are driven by structure.

7.1 Comparative analysis and interpretation

The fusion model is superior due to the effective collaboration of color-space representations. Each color space provides a unique interpretation of an image's attributes. The RGB model retains information about spatial arrangements and structural elements. The HSV model emphasizes the subtleties in hue. Moreover, the YCbCr color space distinguishes between luminance and chromaticity. This multi-representation method enables the model to acquire a broader array of characteristics, thereby avoiding the challenges associated with single-space representations. One of the most significant advances has been in the classification of plant diseases - where a major change in colour is given as a symptom. In contrast, the improvement of face identification is not so marked and more influenced by structural and/or geometric properties rather than color properties. However, both the regions are improving, which indicates that the suggested framework could be applicable in other regions as well.

7.2 Behavior training and error analysis

Figure 7 shows the training and validation accuracy curves for the fused model. They indicate that the model is converging steadily without overfitting. The smooth progression of both curves indicates that the model effectively learns generalized features rather than memorizing training samples. To evaluate the training behavior and model stability, the learning curves of the fused model are shown in Figure 7

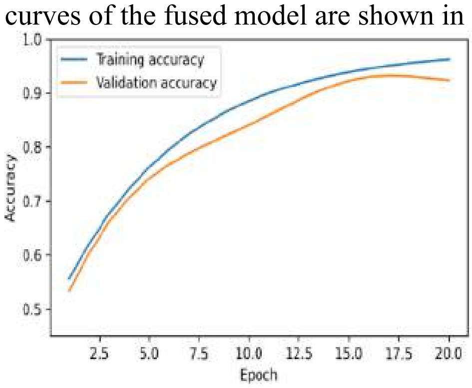
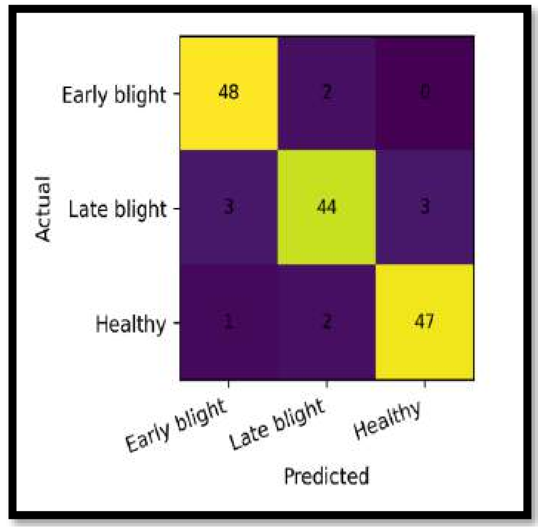


FIGURE 7. Training and validation accuracy curves for the fused model.

In Figure 7, we can observe that the training and validation curves converge continuously. With a small gap between training and validation data and a constant and steady improvement in accuracy, the model is able to generalize well without overfitting. A multi-branch architecture, global average pooling, and feature normalization simplify the model while allowing balanced feature learning across color-space representations, stabilizing it. The classification errors and class-specific performance metrics are detailed within the confusion matrix, as illustrated in Figure 8



.

FIGURE 8. Fused model plant disease categorization confusion matrix.

Figure 8's confusion matrix reveals categorization behavior. Some apparently identical classes are misclassified, but most are correctly identified. The fusion model improves feature representation, but it is still hard to make small distinctions when things change in the actual world. These mistakes are mostly related with classes that have overlapping color and texture features. Feature-level color-space fusion improves classification accuracy and the discriminative structure of the learnt feature space across many evaluation metrics and datasets. Color-space fusion is a representation approach that is both adaptable and dependable for complex visual categorization problems.

8. CONCLUSION

Compact multi-branch color-space fusion system for deep learning-based visual categorization across diverse domains was provided in this article. The proposed model captures complementary chromatic and structural information by processing RGB, HSV, and YCbCr representations using concurrent ResNet-18 branches and merging the embeddings via feature-level concatenation. Fusion typically outperforms single color-space representations in biometric face identification and plant disease classification experiments.

The model outperforms RGB in face identification by 6.60 percent and the best single-space baseline in plant disease classification by 2.20 percent. In the learned embedding space, luminance and chrominance information increase feature diversity and class separability. Since discriminative patterns are strongly linked to color variations, plant disease classification improves considerably, while structural traits modestly enhance face recognition. The framework's continuous progress in both areas proves its generalizability. Features are more dimensional zed and the model becomes more discriminative and robust with feature-level color-space fusion. The transferability of color-aware deep learning architectures for visual recognition across imaging conditions is shown. In subsequent research, attention-guided fusion processes will be investigated, Lab and HLS perceptual color spaces will be included, and the system will be evaluated using more complicated real-world datasets, such as those obtained via unconstrained surveillance and field collected photographs

REFERENCES

[1] M. T. H. Fuad, S. F. Hossain, M. R. Islam, I. Andersson, and M. M. S. Rahman,(2021) "Recent advances in deep learning techniques for face recognition," IEEE Access, 9, 99162-99184, 2021.

[2] I. Adjabi, O. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed,,(2020) "Past, present, and future of face recognition: A review," Electronics, 9(8)1188.

[3] S. M. Shoaib et al., ,(2025) "Leveraging deep learning for plant disease and pest detection: A comprehensive review and future directions," Frontiers in Plant Science, 16, 1538163

[4] A. Zhalgas et al.(2025) "Robust face recognition under challenging conditions: A comprehensive review of deep learning methods and challenges," Applied Sciences, 15(17) 9390, 2025.

[5] B. Amirgaliyev, M. Mussabek, T. Rakhimzhanova, and A. Zhumadillayeva, ,(2025) "A review of machine learning and deep learning methods for person detection, tracking and identification, and face recognition with applications," Sensors, 25(5) 1410.

[6] M. M. Pazouki, O. Toygar, and M. Hosseinzadeh,(2024) "A hybrid approach for color face recognition based on image quality assessment in multiple color spaces," Sakarya Univ. J. Comput. Inf. Sci., 7(1) 1-15.

[7] S. M. H. Abidi et al.,(2026) "Advances in face recognition: A comprehensive review of traditional and deep learning approaches," Electronics, 15(2), 338

[8] R. Soekarta et al.,(2025) "Recent facial image preprocessing techniques: A review," Eng. Proc., 84(1) 2.

[9] E. C. Oerke et al.(2024) "Generative adversarial networks for data augmentation in deep learning-based plant disease identification: A comprehensive review," Computers and Electronics in Agriculture, 212, 108114, 2024.

[10] R. Karthik et al.,(2020) "Attention embedded residual CNN for disease detection in tomato leaves, " Applied Soft Computing, 86, 105933.

[11] S. M. Hassan et al.,(2021) "Identification of plant-leaf diseases using CNN and transfer-learning approach," Electronics, 10(12) 1388,

[12] A. Upadhyay et al.,(2025) "Deep learning and computer vision in plant disease detection: A comprehensive review of techniques, models, and trends in precision agriculture," Artificial Intelligence Review, 58, 92

[13] W. Deng and J. Hu,(2021) "Deep face recognition: A survey," Neurocomputing, 429, 215-244.

[14] J. Zhao et al., (2025) "A review of plant leaf disease identification by deep learning algorithms," Frontiers in Plant Science, 16.1637241

[15] S. Jia et al., (2021) "Discriminant Color Space Learning for Face Recognition in Unconstrained Environments," IEEE Transactions on Information Forensics and Security, 16, 1201–1213.

[16] A. S. Aguiar et al.,(2022) "A color-based vision system for autonomous robotic manipulation in variable lighting environments," Journal of Intelligent & Robotic Systems, 106(2) 42.

[17] M. Gao et al.(2025) "Research progress on color image quality assessment," Journal of Imaging, 11(9)307.

[18] I. Jimenez-Lopez et al., (2025) "Color space comparison of isolated cervix cells for classification," BioMedInformatics, 6(10)261

[19] A. B. C. Author et al., (2024) "Potato leaf disease classification using fusion of multiple color spaces with weighted majority voting on deep learning architectures,"Multimedia Tools and Applications, 83, 14205-4228.

[20] P. T. Chuang et al., (2025) "Deep learning for extraneous material detection in a rice flour processing line," Food Control, 173.111235

[21] M. S. Kiran et al., (2025) "Feature fusion using deep learning algorithms in image classification," Applied Sciences, 15(16) 9053.

[22] A. Eleyan, S. Abdullah, and H. Al-Sewadi, (2023) "Face recognition using ensemble statistical local descriptors," Journal of King Saud University - Computer and Information Sciences, 35(9) 101729.

[23] A. Shah, B. Ali, M. Habib, J. Frnda, I. Ullah, and M. S. Anwar,(2023) "An ensemble face recognition mechanism based on three-way decisions," Journal of King Saud University - Computer and Information Sciences, 35 (4) 196-208.

[24] G. Dai et al., (2023) "PPLC-Net: Neural network-based plant disease identification model supported by weather data augmentation and multi-level attention mechanism," Journal of King Saud University – Computer and Information Sciences, 35(5) 101555.

[25] M. Xiong et al.,(2025) "Leveraging deep learning for plant disease and pest detection," Frontiers in Plant Science, 16, 1538163.

[26] M. S. Krishna et al.,(2025) "Plant leaf disease detection using deep learning: A multi-dataset approach," AgriEngineering, 8(1)4

[27] Z. Salman et al., (2025) "Plant disease classification in the wild using vision transformers and mixture of experts," Frontiers in Plant Science, 16.1522985

[28] L. Xu et al.,(2025) "TSSC: A new deep learning model for accurate pea leaf disease identification," Frontiers in Plant Science, 16.1718758

[29] X. Yang et al.,(2025) "A lightweight and efficient plant disease detection method integrating knowledge distillation and dual-scale weighted convolutions," Algorithms, 18,(7) 433.