

# **2D Image Classification using Hybrid Learning of Handcrafted Features and Deep- Activated Features**

**A  
THESIS**

Presented to the Arden University  
In Fulfillment of the Requirements

For the Degree of  
**MASTER IN SCIENCES**

**IN  
DATA ANALYTICS AND MARKETING**

By  
**SHALLY BANSAL**  
ID: STU94001

*Under the supervision of*  
**DR. YASSER SHOKR**




**DATA ANALYTICS AND MARKETING  
ARDEN UNIVERSITY, BERLIN, GERMANY  
NOVEMBER, 2022**

## **CANDIDATE'S DECLARATION**

I, Shally Bansal, certify that the work in this M.Sc. dissertation was completed at Arden University in Berlin, Germany, under the supervision of Dr. Yaser Shokr. The content of this M.Sc. dissertation has not been submitted for any other degree or diploma.

I confirm that I have accurately recognized, attributed, and referred to the research workers' efforts in the dissertation's text and body. I also acknowledge that I followed all academic truthfulness and authenticity principles in my work and did not falsify or misrepresent any concept, information, or reference. I am aware that any violation of the foregoing will result in institutional disciplinary proceedings.



Date: **28.11.2022**

Place: Berlin

Signature of the candidate

(Shally Bansal)

## **ABSTRACT**

One of computer vision's most popular uses is image classification. This system recognizes the objects in images and gives each one a suitable label. Computer vision learns everything it comes into contact with, just like the human brain, and saves the knowledge for eventual recognition. The features of the image and the class it belongs to are first used to train the image classification system, and then the test images are used to identify the class for the candidate images. A system that has been trained on a large number of images performs better since it has been trained to recognise a range of features.

One of the benchmark datasets for image classification is called Caltech-101. There are 101 object classes in the collection. Caltech-101 is a inspiring multi-class dataset that is used in a variety of research activities.

This dissertation is divided into five chapters. The five chapters of this dissertation are as follows:

We discussed the various stages of the 2D-image classification system in Chapter 1 as well as issues with the image classification system. This chapter also covers the system's various applications. An overview of the literature on the various stages of the image classification system is presented in Chapter 2. This literature review offers crucial background information on image classification. Along with the Caltech-101 image dataset, we have also provided a thorough review of the outcomes obtained by other researchers in relation to image classification systems using a variety of image datasets. A novel feature extraction method using a combination of local hand-crafted features (SIFT and Haralick descriptors) and deep-activated features (VGG19 features) has been presented in Chapter 3. As a method for selecting features, we used the k-means clustering method and locality-preserving projection techniques to accomplish the task. In this work, the classifiers k-NN, MLP, and Random Forest are considered. The adaptive boosting technique is also explored in this study to enhance the classification performance.

A hybrid learning of local features (SIFT and Haralick descriptors) and deep-activated features (VGG19) for image classification systems has been

tested, and the detailed experimental findings based on the proposed approach are described in Chapter 4. Four classification techniques—k-NN, MLP, Random Forest, and Adaptive Boosting are used to classify the images based on their extracted features. Using k-NN, MLP, and Random Forest, respectively, on the hybrid learning of local features and deep features, the work has achieved recognition accuracy of 97.65%, 98.03%, and 99.07%. The work also makes use of adaptive boosting, which has a recognition accuracy rate of 98.67%. k-NN, MLP, Random Forest, and Adaptive Boosting have all been examined for a number of additional performance criteria, including True Positive Rate, Recall Rate, Root Mean Squared Error, and Area Under the Curve.

In Chapter 5, we have presented the conclusions and recommendations of this work.

***Dedicated***  
***to my***  
***beloved family members***  
***who inspired me***  
***all the time***

# Contents

<b>Declaration</b>	<b>i</b>
<b>Abstract</b>	<b>ii-iii</b>
<b>Contents</b>	<b>iv-v</b>
<b>Abbreviations</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>Chapter 1. Introduction</b>	<b>1-11</b>
1.1. Background	1-3
1.2. Rationale	3
1.3. Stages of Image Classification System	4-7
1.4. Uses of Image Classification System	7-8
1.5. Objectives of the work	8
1.6. Research Questions	9
1.7. Major Contributions and Achievements	9
1.8. Dataset Used	9-10
1.9. Dissertation Structure	11
<b>Chapter 2. Review of Literature</b>	<b>12-21</b>
2.1. Techniques used in the work at the different stages of the classification system	12-14
2.1.1. Pre-processing	12-13
2.1.2. Feature Extraction	13
2.1.3. Feature Selection and Dimensionality Reduction Techniques	13-14
2.1.4. Classification	14
2.2. Literature Survey	14-19
2.3. Critical Analysis Based on the Survey	19-20
2.4. Chapter Summary	20-21
<b>Chapter 3. Proposed Methodology</b>	<b>22-34</b>
3.1. Introduction	22

3.2. Feature Extraction Techniques	22-26
3.2.1. Local Features	23-24
3.2.2. Deep Features using VGG19	24-26
3.3. Feature Dimension Reduction Techniques	26
3.3.1. k-means clustering algorithm	26
3.3.2. Locality Preserving Projection	27
3.4. Classification Techniques	27
3.4.1. k-NN	28
3.4.2. MLP	28
3.4.3. Random Forest	29
3.4.4. Adaptive Boosting	29-30
3.5. Proposed Methodology	30-32
3.6. Research Questions	32-33
3.7. Validity and Reliability	33
3.8. Ethics and Bias	33-34
3.9. Limitations	34
3.8. Chapter Summary	34
<b>Chapter 4. Experimental Results and Discussion</b>	<b>35-44</b>
4.1. Dataset	35
4.2. Software and Code Implementation	35-36
4.3. Performance Evaluation Parameters	36-38
4.3. Experimental Results	38-44
4.4. Chapter Summary	44
<b>Chapter 5. Conclusion and Future Directions</b>	<b>45-52</b>
5.1. Introduction	45
5.2. General Conclusion	46-47
5.3. Research question conclusions	47-51
5.4. Recommendations	51
5.5. Errors and limitations	51-52
<b>References</b>	<b>52-59</b>
<b>Appendix</b>	

## ABBREVIATIONS

2D	Two Dimensional
AwA	Animals with Attributes
AUC	Area Under Curve
BRIEF	Binary Robust Independent Elementary Features
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CUB	Caltech-UCSD Birds
CZSL	Conventional Zero Shot Learning
DCNN	Deep Convolutional Neural Network
FAST	Features from Accelerated Segment
FPR	False Positive Rate
GLCM	Grey Level Co-occurrence Matrix
GNN	Gaussian Naïve Bayes
k-NN	k-Nearest Neighbor
LPP	Locality Preserving Projection
ORB	Oriented and Rotation BRIEF
ResNet	Residual Network
RMSE	Root Mean Squared Error
TDP	Task Driven Pooling
TPR	True Positive Rate
VGG	Visual Geometry Group



## List of Figures

Figure No.	Title	Page No.
1.1	Background processing of machine learning system	2
1.2	Basic diagram of 2D-Image classification System	3
1.3.	Image classification system framework	4
1.4. (a)	Original Coloured Image	5
1.4. (b)	Pre-processed using colour to grey-scaled	5
1.4. (c)	Pre-processing using saliency maps	5
1.4. (d)	Pre-processing using saliency maps with threshold	5
1.5.	A few samples of images from Caltech-101 dataset	10
3.1	Architecture of Deep Convolutional Neural Network	25
3.2	Block diagram of the proposed system	32
4.1	Confusion Matrix for multi-class classifier	37
4.2	Area Under Curve (AUC)	38
4.3	Classifier Wise Recognition Accuracy	40
4.4	Classifier Wise Precision Rate	41
4.5	Classifier Wise Recall Rate	40
4.6	Classifier Wise F1-Score	42
4.7	Classifier Wise Area Under Curve	43
4.8	Classifier Wise Root Mean Squared Error	44

## **List of Tables**

<b>Table No.</b>	<b>Title</b>	<b>Page No.</b>
1.1	Description of the Caltech-101 dataset	10
2.1	Image classification results on Caltech datasets using various feature extraction techniques	20
4.1	Feature Extraction Techniques	40
4.2	Classifier Wise Recognition Accuracy	40
4.3	Classifier Wise Precision Rate	41
4.4	Classifier Wise Recall Rate	41
4.5	Classifier Wise F1-Score	42
4.6	Classifier Wise Area Under Curve	43
4.7	Classifier Wise Root Mean Squared Error	43

# Chapter 1

## Introduction

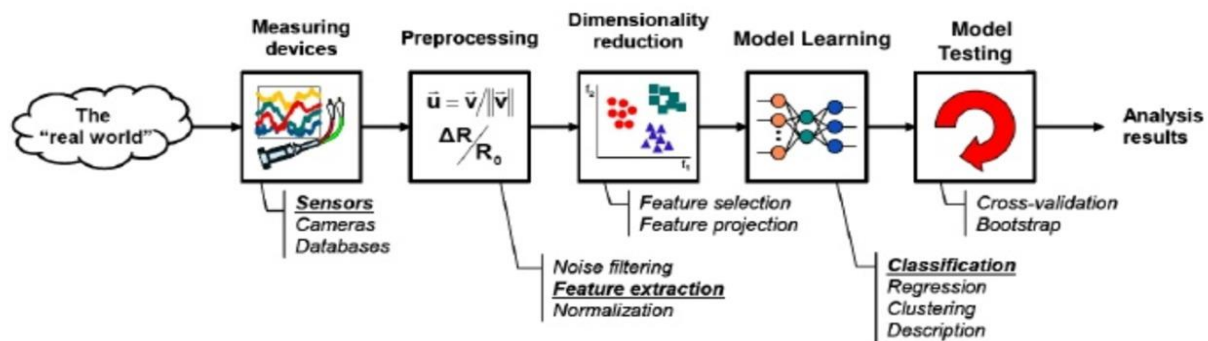
---

Artificial intelligence is becoming important in the lives of digital natives in this era of cutting-edge technology. An artificial intelligence-based system's design is based on learning, processing, and self-correlation. One of the most challenging processes in artificial intelligence is image classification. The image classification system is intended to recognize and classify images and objects that remain in these images. Many researchers have failed to achieve satisfactory results despite numerous attempts due to various issues such as occlusion, lighting effects, geometrical variance, cloudy images, background clutter, processing complexity, etc. In this project, I have presented an efficient system for 2D image classification. As a result of the process of image classification, several issues arise, and the project addresses these by combining traditional machine learning algorithms with cutting-edge VGG19 deep learning methods in order to address the issues.

### 1.1. Background

There are a multitude of real-world applications that can be achieved through the use of the 2D-image classification system, such as retrieval of content-based images, medical imaging, security monitoring, etc., that could benefit from the use of this system. Image classification is now performed using a variety of machine learning and deep learning models. Supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning are the four types of machine learning. In machine learning, the trained system was used to label query images. Supervised learning is used to train models on a labelled dataset. Whereas in semi-supervised learning, some images in the dataset already have labels but the majority of the images do not. Semi-supervised learning assists in the identification of labels for those unfamiliar images. There is no labeled dataset for unsupervised learning. Reinforcement learning trains the dataset using a trial-and-error technique, which assists in the identification of the optimum solution

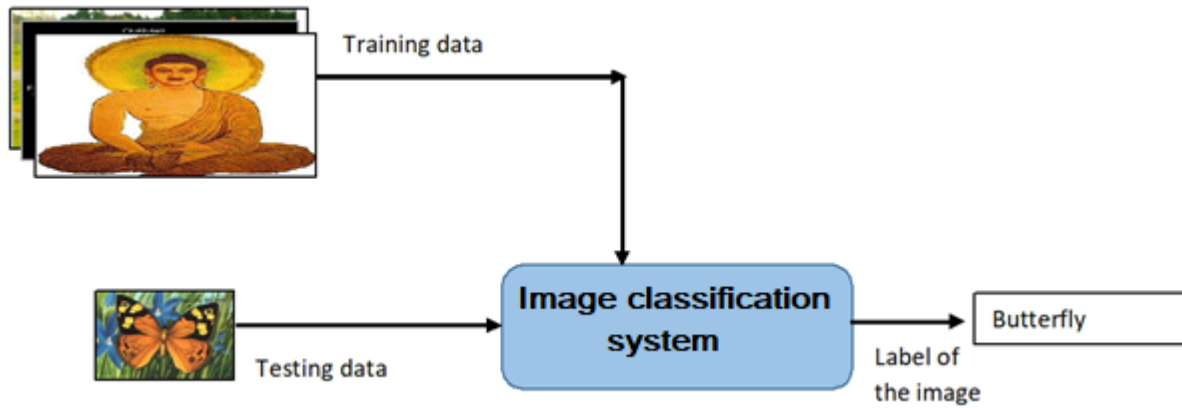
for the input image. I have explained the introduction of the 2D-image classification system in the following sub-section (Fig. 1.1).



**Fig 1.1.** Background processing of machine learning system

### 1.1.1. 2D-IMAGE CLASSIFICATION SYSTEM

The human mind is able to recognize images regardless of their characteristics, but machines are unable to recognize the same images despite the same characteristics. Therefore, a machine that replicates a human brain is required. For example, when a child is in school, they are frequently taught the names of many objects and shapes. He or she can determine additional comparable items that he or she has already learned. In the same manner that a kid learns, the computer must be thoroughly educated using various machine learning algorithms. A database of labels for all images can be created with these algorithms by extracting features from the image and adding the labels to the database. It is essential to store all the characteristics and names of all related objects during the training process. During the testing step, features of the input image are checked by comparing them with the feature database of images. Finally, the system classifies the images based on the extracted features. Quality of image features has a significant impact on the performance of an image classification system. As shown in Figure 1.2, an image may have a number of features including colour, texture, form, and so on, which all contribute to a more accurate classification of the image.



**Fig 1.2.** Basic diagram of 2D-Image classification system

## 1.2. Rationale

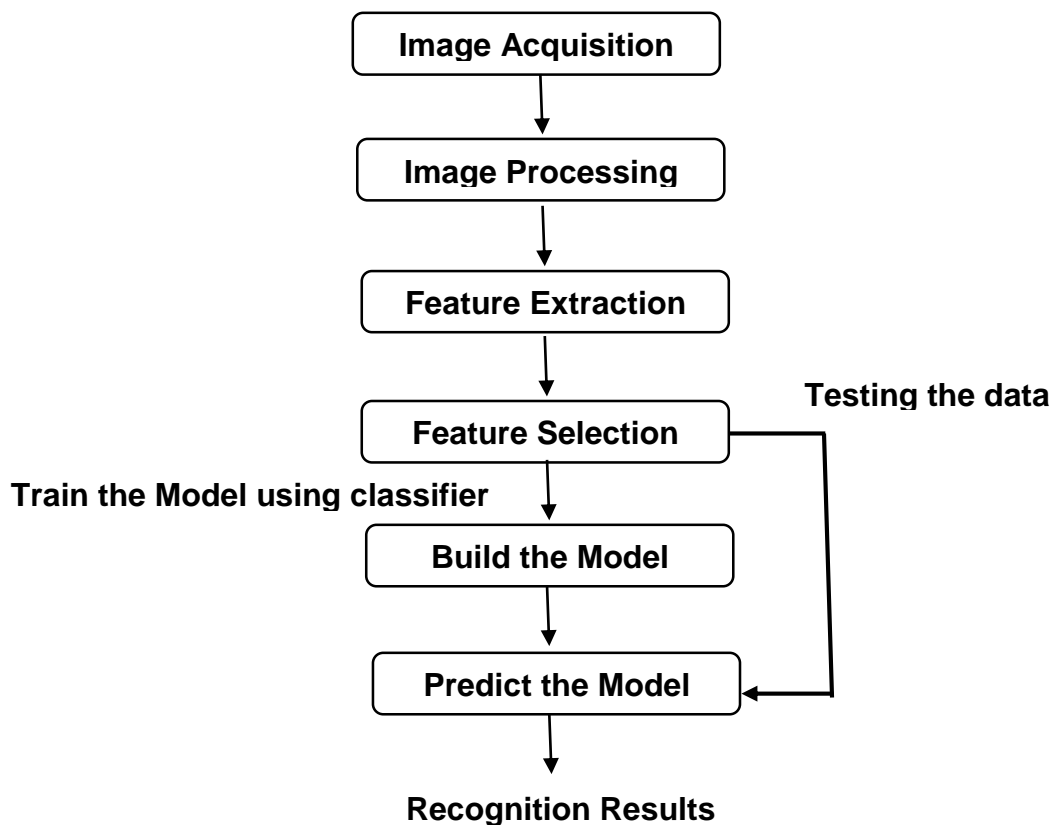
During image classification, various problems occur, such as occlusion, viewport variations, affine translation, changing item counts, and lighting. To resolve these issues, we used a hybrid learning approach that combined handcrafted features (SIFT and Haralick texture) with deep-activated features (VGG19) to improve image classification performance.

SIFT produces better feature extraction results in rotation, scale, and illumination view points, but it is unreliable in foggy images. The Haralick texture descriptor is a statistical texture method that extracts image properties using texture. Haralick *et al.* (1973) devised an approach that uses a grey level co-occurrence matrix to analyse the geographical distribution of grey values, followed by the derivation of local (texture) features. Their approach is not affected by translation or rotation and computes fourteen GLCM features that characterize the basic properties of the image (e.g., smoothness, the directionality of the pattern, image complexity, etc.). Using the VGGNet (Visual Geometry Group) architecture, Simonyan and Zisserman (2014) propose a popular architecture for deep convolutional neural networks in the field of image classification that is based on visual geometry. The model is trained on a huge image dataset, the ImageNet dataset, which contains over 1000 categories and 14 million images. The model has achieved second place in ILSVR-2014. In total, the VGGNet model consists of 19 layers, of which 16 are convolutional layers, which are used to extract feature values, and three layers are fully

connected layers, which are used to categorize features. The VGG19 model is designed by grouping the first 16 layers into 5 groups that are stacked over each other.

### 1.3. Stages of Image Classification System

An image classification system performs a number of operations, such as image acquisition, preprocessing, feature extraction, feature selection, and classification. The flow of these activities is shown in Figure 1.3.



**Fig 1.3.** Image classification system framework

#### 1.3.1. IMAGE ACQUISITION

Image classification begins with the acquisition of the image dataset. The purpose of the present project was to evaluate image classification algorithms on Caltech-101 dataset, one of the benchmark datasets for image classification.

### 1.3.2. PRE-PROCESSING

Preprocessing is the next step in an image classification system that helps to reduce complexity while increasing system accuracy. Images are preprocessed by doing simple tasks such as scaling, converting a colourful image to a grayscale image, etc. Because a few machine learning algorithms perform best on grayscale images, so a translation from colourful to grayscale is required. This type of preprocessing is used by the majority of shape-based recognition algorithms. Next, black and white graphics reduce complexity and take up less memory space.

A saliency map helps in distinctive visual features in an image and creating a meaningful representation of the image. Convolutional neural networks require an image dataset with images of the same size. As a result of the preprocessing phase, the feature extraction phase is able to produce better recognition results, since the visual features of the images are required during this stage (Fig. 1.4).



(a)



(b)



(c)



(d)

**Fig 1.4.** (a) Coloured Image, (b) Pre-processed using colour to grey-scaled, (c) Pre-processing using saliency maps, (d) Pre-processing using saliency maps with threshold.

### **1.3.3. FEATURE EXTRACTION**

The features of an image are the traits that distinguish it from other images. An image's features might include its shape, colour, texture, or spatial layout information. Furthermore, features can be classified as handcrafted or deep-activated features. An image classification system's performance is strongly influenced by the number of features extracted from the image. The handcrafted features might either be local or global. Global features are required to identify the whole image, whereas local features describe the image's primary points. In this project, we explored a prominent pre-trained deep learning feature extraction approach (VGG19) as well as two local feature extraction methods (i.e., SIFT and Haralick descriptor). Because the number of features retrieved from pre-trained convolutional neural networks and SIFT is quite large. So, feature selection algorithms are employed to enhance and accelerate the classification operation. The K-means clustering technique creates a feature vector by picking the most significant characteristics with the help of pre-trained VGG19 and handcrafted features (SIFT and Haralick features).

### **1.3.4. FEATURE SELECTION AND DIMENSIONALITY REDUCTION**

An image includes a great amount of information that takes up a lot of storage space. This phase helps in the detection of crucial information in the large dataset and offers a smaller quantity of data. This feature vector improves the reliability and efficiency of the classification system because vital information is accepted and unnecessary data is eliminated. This solves the over fitting problem. Another advantage of the feature selection and reduction approach is that it requires less space for the feature vector and is fast to compute. So, in this study, k-means clustering was used for feature selection and locality-preserving projection (LPP) was employed for dimensionality reduction.

### **1.3.5. CLASSIFICATION**

The classification stage is the most important part of the image classification system since it allows you to assign a class or label to an image. The reliability of the classification phase is based on the uniqueness of the extracted features of images and the ability of classification to relate an image's features to its class.



Many machine learning and deep learning classification techniques have been developed to produce models that classify images based on their characteristics. We are evaluating the model's performance in terms of image classification based on the quality of predictions made by the classification algorithm based on how well the image classification algorithm performs as a result of these predictions. Classification algorithms such as k-NN, MLP, Random Forest, and adaptive boosting were used to classify images. The model's efficacy is measured using a range of performance evaluation measures.

#### **1.4. Uses of Image classification system**

This section comprises various applications of the image classification system that are discussed as follows:

- **Face Recognition** – In this case, a face is used to identify an object from an image. This method helps in the recognition of a person in an image.
- **Aerial Image Analysis and Counting** – It helps in counting all instances of a specific image.
- **Intelligent Vehicle System** – It enables a computer to operate a vehicle autonomously. An intelligent vehicle system is required to detect the presence of barriers and recognize traffic signs automatically.
- **Pedestrian Recognition** – It helps in the recognition of pedestrians in a cluttered environment.
- **Biometric Recognition** – Images can be used to identify individuals based on their physical characteristics such as the eye, ear, fingerprint, and so on.
- **Surveillance** - An image classification system helps surveillance systems by recognizing a suspect person or vehicle from a still video or image.
- **Security** – Image classification is becoming a critical tool for improving security in areas like airports and government offices. Automated recognition systems might be used to check bags for security purposes at airports. The same is true at banks, where the presence of firearms and other weapons may be easily spotted using image classification technology.

- **Industrial Inspection** – An image classification system makes it simple to identify elements of machinery and monitor their operation for any type of fault or damage.
- **Medical Analysis** – Image classification also helps in the analysis of biological signals. It involves actions such as detecting tumours in MRI images, detecting skin cancer, and showing images acquired by X-rays, MRIs, or other equipment in more detail.
- **Optical Character Recognition (OCR)** – An optical character recognition device assists in character recognition in scanned documents. Any language can be used to write the characters.
- **Assistive Device** – By recognizing various human gestures in an image, computers can communicate in real-time with a human being by recording those gestures within their system.
- Even image classification helps the blind, people with limited eyesight, and the deaf and dumb in their daily lives.
- **Content Based Image Retrieval (CBIR)** – It returns all comparable images from an image database in response to a query rather than words, the query is focused on the image contents.

## 1.5. Objectives of the work

An approach to classifying 2D images using deep-activated features and handcrafted features is presented in this dissertation.

- To empirically study and evaluate a few existing techniques for image classification.
- To design an efficient approach for 2D image classification using handcrafted features and deep-activated features.
- Evaluation and validation of the proposed method through the use of benchmark datasets and performance metrics.

To achieve these goals, various steps of the image classification system was conducted. A hybrid strategy of a few feature extraction approaches was employed during the work

to generate a feature vector for classification purposes. For classification challenges, many classification approaches such as k-NN, MLP, Random Forest, and adaptive boosting methodologies are considered in this work.

## **1.6. Research Questions**

There are a few common research queries related to this research work, and these queries are solved using the work proposed in this dissertation.

- What are the various approaches of image classification?
- Which strategy is best for image classification?
- What are the various datasets available for image classification?
- What are the various problems that are faced in image classification?

## **1.7. Major Contributions and Achievements**

The following are the key contributions of the current work:

1. A thorough review of the various phases of the image classification system has been conducted.
2. For the experimental work, a multi-class and imbalanced dataset of 101 classes was considered.
3. To improve the quality of the input images, a few preprocessing techniques were applied.
4. Various handcrafted feature extraction approaches and a deep-activated features using VGG19 have been explored in order to generate a feature vector for classification purposes.
5. Various multi-class classification algorithms for image classification based on individual feature extraction approaches as well as their combinations have been explored.
6. The outcomes of all algorithms evaluated were assessed using a various performance evaluation measures.

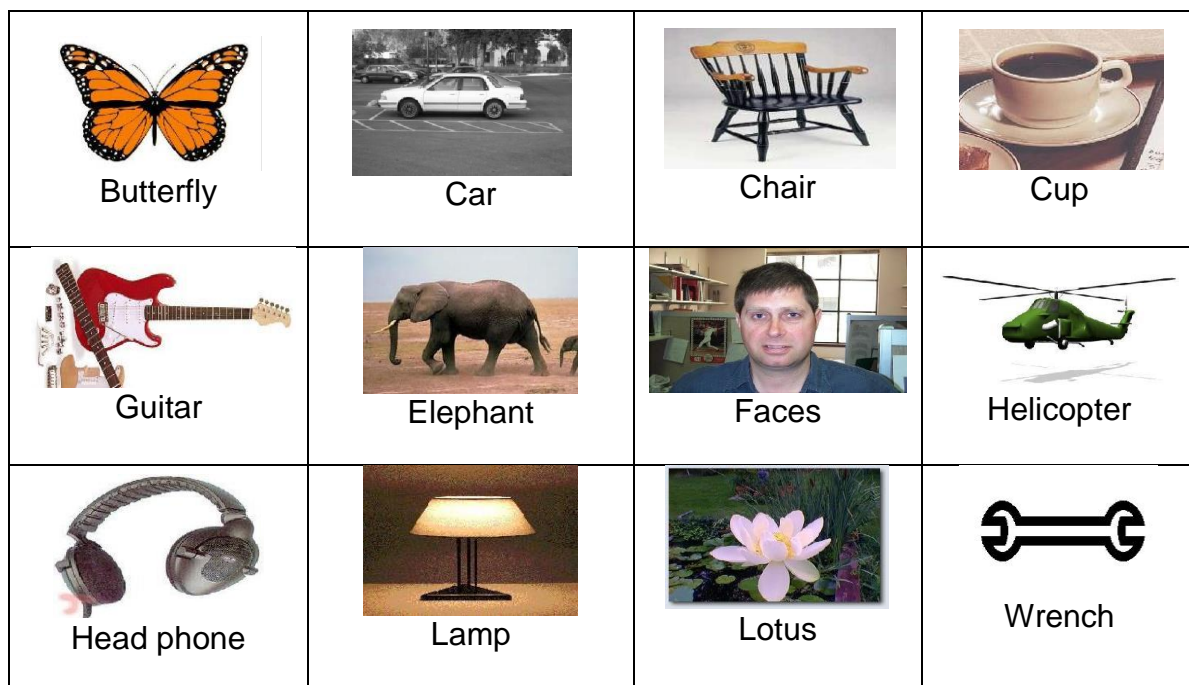
## **1.8. Dataset Used**

To conduct the experiments on Caltech-101 images for the purpose of performing the experiments. There are so many classifications in this dataset that it would be difficult for

a classification system to classify it because it has so many categories. Each class has 40–800 images. The following is the source of this public dataset:

<https://www.tensorflow.org/datasets/catalog/caltech101>.

The dataset includes 101 classes (ant, chair, cup, faces, helicopter, wrench, and so on) as well as 1 backdrop scene. A total of 8677 images were used in the experiment, representing 101 different image classes, in order to be able to examine the experimental results in more detail. The Caltech-101 dataset is one of the most difficult to analyze since it contains a collection of noisy and irregularly shaped images with varying illumination. The number of images in each class ranges from 40 to 800. The images in Figure 1.5 show a few samples of the dataset.



**Fig. 1.5.** A few samples of images from Caltech-101 dataset

**Table 1.1.** Description of the Caltech-101 dataset

Dataset:	Caltech-101 image dataset
Total no of images:	8677
Number of Classes:	101
Number of images per class:	40 – 800
Data Partitioning Strategy:	70:30
Source:	<a href="http://www.vision.caltech.edu/Image_Datasets/Caltech101/">http://www.vision.caltech.edu/Image_Datasets/Caltech101/</a>

## **1.9. Dissertation Structure**

Increasingly, artificial technology is being used to simplify our lives as a result of its ease of use. Some of the applications of AI are image classification, expert systems, face recognition, speech recognition, computer vision, etc. Most of the systems designed using AI are based on learning, reasoning, and self-correlation. The main objective of this dissertation was to present a system that identifies 2D images efficiently. This dissertation examines various challenges encountered during the image classification task and provides an efficient solution to solve all of them. Several approaches have been used to help identify objects and objects in images. The dissertation's structure is given shortly below. In Chapter 1, an overview of the work completed by addressing the various stages of the 2D-Image classification system is provided. In Chapter 2, a study of the literature on the methodologies offered by previous researchers is given. This literature review offers critical baseline knowledge for the image classification system. Chapter 3 offers a suggested technique for the image classification, which uses a hybrid learning of handcrafted features (like SIFT and Haralick texture) and deep-activated features (VGG-19). The experimental study was evaluated using the state-of-the-art classification algorithms, including k-NN, MLP, Random Forest, and adaptive boosting. The comprehensive experimental research work based on the proposed technique is described in Chapter 4. A variety of performance measures were used to assess the performance of the experimental system, including recognition accuracy, precision, recall, F1-Score, and root mean squared error, in order to determine the effectiveness of the experimental system. Finally, chapter 5 ends up the dissertation work by discussing the concluding notes and the future perspectives. This goal has been accomplished because the proposed system successfully classified the images. This experimental attempt employed a hybrid approach that includes various feature extraction algorithms that help in the image classification system. Multiple algorithms were used during the feature extraction and classification processes to successfully complete the task.

## Chapter 2

# Review of Literature

---

A literature survey was conducted in this chapter while examining the various methodologies used to develop the image classification system. This chapter also provides a detailed description of various features and classification techniques for image classification. The literature dealing with each stage of image classification has been presented in Section 2.1. Experimental results using various approaches with different datasets such as PASCAL VOC 2006, Label ME, Corel, Caltech-101, etc. are analyzed in Section 2.2. This chapter also includes an analysis of various image classification algorithms employed on the Caltech-101 dataset. Section 2.3 presents the critical analysis based on the literature survey. Finally, this chapter ends with the chapter summary. This research work enabled us to explore further combinations of different feature extraction approaches to optimize the image classification results.

### **2.1. Techniques used in the work at the different stages of the classification system**

This section includes a brief summary of the various activities employed in this work for image classification, as well as descriptions of the methodologies used in these activities.

#### **2.1.1. Pre-processing**

An image must be pre-processed before it can be used for feature extraction. An image may contain unwanted or noisy data. Pre-processing boosts the efficiency of a classification system's work by reducing unwanted, distorted input and/or emphasizing key characteristics of an image. A coloured image requires a huge amount of memory space to be saved, and computations during feature extraction take longer. Pre-processing is most commonly used to transform colour images into grayscale ones. Images are then converted to binary images, which are then processed into binary format.

Among the preprocessing techniques that can be used to achieve a meaningful representation of images is saliency maps (Montabone and Soto, 2010), which are used

to separate the various visual characteristics in images. In their paper, Chao et al. (2018) describe how image classification can be achieved by bagging features together. They employed a deep convolutional neural network based on salient object recognition to extract the features, and significant data from the image using saliency maps as a pre-processing step.

### **2.1.2. Feature extraction**

A major component of image classification is the feature extraction phase, which extracts the features that distinguish each image from others. The image classification system requires a set of features like colour, texture, edges, corners, and other details to classify a particular image. Various handcrafted approaches or deep learning algorithms can be used to extract image features. Handcrafted features are further classified as local and global, and include edges, corners, and other aspects of the surface of an object such as texture, colour and shape. There are several techniques that can be employed in order to identify local features, such as the Scale Invariant Feature Transform (SIFT), the Haralick Texture Descriptor, etc. The image classification system makes extensive use of local characteristics such as area, perimeter, etc. Global characteristics help in distinguishing an image from its surroundings. However, they are only effective when there is no occlusion or light pollution.

Chien *et al.* (2016) examined the accuracy and processing capability of SIFT, SURF, ORB, and A-Kaze feature descriptors. Descriptors are chosen based on various general invariance criteria, such as rotation, scaling, extracted features, execution time, and memory needed to store the data. There was an in-depth comparison between SIFT, SURF, ORB, BRISK, KAZE, and AKAZE feature descriptors by Tareen and Saleem (2018) that provides a straightforward comparison of the features. To experiment with the methods of classifying images, the researchers used MATLAB and OPENCV in order to work with these approaches.

### **2.1.3. Feature selection and dimensionality reduction techniques**

When creating a feature vector of a size that is appropriate, it may take several algorithms to come up with one that works. Some of these are over-fitting issues—the need for a huge amount of space, a longer execution time, and so on. Various feature selection

techniques can be used to overcome these difficulties. The feature selection method has been presented to reduce the problem of over-fitting data. K-Means clustering and Locality Preserving Project (LPP) techniques were used in this work for dimensionality reduction. The approach employs k-means, which lowers the feature vector's dimensionality and allows for quicker computations.

#### **2.1.4. Classification**

The classification phase is used to classify the features extracted in sub-section 2.1.2. Classification techniques are classified into two types: supervised and unsupervised. In supervised learning, classification methods include Naive Bayes, Bayesian networks, SVM, k-NN, Random Forest, XGBoosting, and others. Unsupervised learning classifiers include fuzzy clustering, hierarchical clustering, etc.

## **2.2. Literature Survey**

Helmer and Lowe (2004) introduced an image classification technique based on part-based modelling. This model solved the occlusion, size, and backdrop clutter issues. A modified nearest-neighbour classifier was used to compute the findings. It produced better results in two classes (vehicles and motorcycles) than existing approaches. After that, cropping was applied to the object in the image, leading to an improvement in accuracy across all four object classes (cars, motorcycles, faces, and airplanes). An efficient codebook was created by Jurie and Triggs (2005) for image classification. The SVM classifier was used to predict the labels for the items based on the 600 code words that were generated. The MIT dataset and the Caltech-101 dataset were used in this experimental study. For unsupervised image classification, probabilistic latent semantic analysis (pLSA) was employed. This model identified an object in an image using a bag of words (Sivic *et al.*, 2005).

Bosch *et al.* (2007) have proposed an image classification framework using the SIFT approach, and its local shape was determined using a histogram of gradients (HOG). A spatial pyramid was built across an area of interest using a spatial pyramid matching technique. For classification, they employed a Random Forest classifier. An image classification framework based on visual words has been developed by Chum and Zisserman (2007). ROI is determined using canny edge detection, SIFT, and appearance



patches based on the Hessian-Laplace operator. They employed a vocabulary of 3000 visual words learned from the Pascal VOC 2006 training set. A grayscale image classification method was developed by Arnow and Bovik (2008). They recommended an addition of global feature to improve the original's accuracy. To detect objects, the Harris corner detector and the SIFT descriptor are used. The outcomes of this research were assessed based on accuracy, and speed. A codebook has been created from the images of the training dataset for the purpose of detecting objects in an image, which are then tagged accordingly as they appear in the image by using the codebook that was created from the training dataset images. SVM performs better than k-NN, for testing conducted on the Caltech-4 cropped dataset. It has been suggested that an efficient method be used to speed up the feature detection process. Wu *et al.* (2012) revealed that the Shi-Tomasi and Harris corner detection approaches reduced execution time by 90% and 70%, respectively. They used this technique to crop an image's non-corners.

The Manhattan Distance technique was employed to determine the similarity metric when comparing images. Roy and Mukherjee (2013) developed the approach for content-based image retrieval by employing a unique feature extraction technique. Ijjina and Mohan (2014) classified images using CNN and a 3D colour histogram as a feature detector. Results from this study, which were tested on the ALOI dataset, had an average accuracy of 97%. Muralidharan (2014) converted a colourful image to a grayscale image to build a feature vector. He then used the recovered edge information to apply Hu's Seven Moment Invariants, the item's centre, and its dimension. Principle Component Analysis was used to derive eigenvalues from the computed feature vector. Xie *et al.* (2015) developed a feature pooling technique for image classification. They compared task-driven pooling (TDP) to other pooling algorithms and found that it produced better, more accurate results. TDP may be applied to either BoW or CNN features. Flower17, Indoor67, and Caltech-101 datasets were used in their experimental study. An analysis of colour and shape features extracted from images was carried out by Diplaros *et al.* (2016)

This was tested on a variety of datasets and shown to be more accurate with rapid recognition. This proposed approach also detected the object when it was obscured by other objects or cluttered. Prajapati *et al.* (2016) proposed the form feature for identifying objects in images. For the purpose of identifying the best matching images in the

database, I was able to compare the computed feature vector of the query image with the recorded feature vectors of the image database. The similarity was calculated using the Euclidean distance measure. Rastegari *et al.* (2016) compared their proposed binarization method to BinaryConnect and BinaryNets. In terms of top-1 accuracy, the proposed approaches beat the other methods on ImageNet by more than 16%. Their proposed solution decreased the network size by 32 times and allowed the recognition system to function on compact devices. It has been suggested that Chao *et al.* (2016) can be used to address the generalized zero-shot learning issue by introducing a conventional zero-shot learning method (CZSL). The authors explored recognizing the item from both visible and unseen images in their study. The experiment was carried out using live datasets from Animals with Attributes (AwA), Caltech-UCSD Birds (CUB), and ImageNet. The authors classified seen and unseen images using stacking classifiers.

Affonso *et al.* (2017) used machine learning to predict the quality of wood boards from images. The results were compared to experimental data obtained using various classification techniques and represented an experimental data comparison. Agarwal *et al.* (2017) conducted a comparison of the SIFT and SURF image classification methods. They tested the performance of both algorithms on various items in various environments. The ORB-BRIEF combination was also found to be the quickest compared to others in terms of detection speed. Shang *et al.* (2017) have proposed a novel feature descriptor, namely, Local Binary Descriptor (LDQBD).

Xie *et al.* (2017) introduced a multi-label classification technique for feature learning. This method includes selection, discrimination, and equalizing features. When these attributes were integrated with CNN, the accuracy outperformed two prominent multi-label classification datasets, VOC2007 and VOC2012. For object detection, Wei *et al.* (2017) employed a cutting-edge framework. This framework, which was based on contour shape descriptors, performed well in crowded images. This was tested on a variety of datasets, including the ETHZ shape courses, the INRIA horses, the Weizmann horses, and two Caltech-101 classes. Shu *et al.* (2018) devised a novel way to classify images using a fine-grained dictionary learning approach that retrieves local properties of the image.

Ahmed *et al.* (2017) proposed a combined regional and texture feature extraction approach for image classification. They extracted shape-based features using statistical moments and image moments. The Local Binary Pattern (LBP) algorithm is used to extract texture features. The size of these combined sets of features is optimized using a principal component analysis algorithm. For classification, they have considered the SVM classification algorithm. The system was tested on Corel-100, Caltech-101, and Caltech-256 datasets and outperformed other methods. Garcia-Gasulla *et al.* (2017) studied the behaviour of features extracted through a deep CNN pre-trained model (VGG19) with an SVM on three datasets: MIT67, flowers102, and CUB200. They analysed the behaviour of individual features in each class and explored the behaviour of intra- and inter-class properties. Xiao *et al.* (2018) compared three pre-trained convolutional neural network models with conventionally constructed feature extraction approaches. They created a breast ultrasound image-based combination technique that employed a mixture of models to extract certain information from an ultrasound image. Zhu *et al.* (2018) used local, global, and deep features to classify high spatial resolution imaging scenes. Liu *et al.* (2018) proposed a bag-of-features approach using SIFT and HOG feature descriptors to classify the objects. The experimental work was conducted on a few categories of the Caltech-101 dataset and the Scene-15 dataset. The K-means clustering algorithm is applied to minimise the size of the feature vector for fast and efficient computations. These form a visual dictionary for the image of the classification process.

Kabbai *et al.* (2019) put forward a novel feature descriptor that included colour and texture information. They examined local and global features of images. They put their proposed work through its paces on a variety of datasets (New-BarkTex, Outex-TC13, Outex-TC14, UIUC sports, MIT scene, Caltech-101, and MIT indoor scene) and compared the classification results to previous handcrafted and deep learning work. Top-SIFTs are a type of SIFT method that was created to reduce the redundant features of an image by including a sparse constraint and spatial distribution. They presented that the method can reduce space and time while also to improve accuracy. Top-SIFTs is a type of SIFT method that was created to reduce the redundant features of an image by including a sparse constraint and spatial distribution. Liu *et al.* (2018) demonstrated that the method can reduce space and time while also improving accuracy. Vo *et al.* (2019) have

introduced a deep learning model for grin recognition that performed well on both balanced and unbalanced datasets. A convolutional neural network was used to extract features from an imbalanced dataset. To train the model in response to the imbalanced data, the authors used high-gradient boosting. A proposed solution for integrated biometric recognition that combines handcrafted and deep features has been proposed by Vo *et al.* (2019). A combination of the suggested approach could be used to collect facial biometrics (facial features, iris features, palm smudges, fingerprints, ear shapes) and iris features (iris features) to detect fraud. Khan *et al.* (2020) introduced a unique human activity detection system that merged handcrafted (shape) and deep learning features for feature extraction and a multi-class support vector machine (SVM) for classification. Many datasets were used in the experiment, including Weizmann, UCF11 (YouTube), UCF19 Sports, IXMAS, and UT-Interaction. Tesfaye and Pelillo (2020) suggested a content-based image retrieval system that combines handcrafted and deep activated features. To determine the similarity between the images in the dataset and the query image, the Euclidean distance technique was used. A feature selection method called k-NN helped to minimize the dimensions of the feature map. Chen *et al.* (2021) present a review of image classification algorithms using CNNs. Convolutional neural networks (CNNs) have dominated the area of image classification since 2012. Typically, the image classification architecture is also used for other visual recognition applications. Convolutional neural network-based image classification methods were the main area of study for Luo (2021). Image classification requires inputting an image and then identifying the image's category using a specified classification algorithm. The image classification effect is superior to conventional machine learning algorithms. Neural networks are a popular and significant research direction in machine learning. An artificial intelligence system can categorize normal and flawed images with over 91% accuracy. To extract distinguishing features from the dataset for classification, a pretrained convolutional neural network based on the PyTorch framework is used (Wu and Zhou, 2021).

Abdellatef *et al.* (2020) presented an integrated biometric recognition method using the fusion of handcrafted and deep features. The proposed approach is used jointly for recognizing a face, iris, palm print, fingerprint, and ear biometric. The features are reduced using principal component analysis (PCA). The proposed system achieved

maximum accuracy for all these recognition systems that were evaluated individually. Khan *et al.* (2020) introduced a novel approach for human activity recognition in which a fusion of handcrafted (shape) and deep features is used for feature extraction. The authors adopted this technique to experiment with the pros and cons of all these feature extraction algorithms. The experiment was tested on various datasets, including Weizmann, UCF11 (YouTube), UCF 19 Sports, IXMAS, and UT-Interaction, and achieved good results with a 70:30 data partitioning strategy. Bansal *et al.* (2021) used the Shi-Tomasi corner detection algorithm to present an efficient technique for 2D object recognition. They used adaptive boosting in combination with a random forest classifier as well as features such as Shi-Tomasi, SIFT, and SURF to achieve a maximum recognition accuracy of 86.4% as a result of the combination of adaptive boosting and a random forest classifier.

### **2.3. Critical Analysis Based on the Survey**

In this chapter, I analyzed various features and classifiers for image classification. The problem of accurately recognizing multiple images despite all the efforts that have been made remains a challenge despite the numerous efforts that have been made. I have analyzed several studies related to image classification and the level of accuracy attained on a certain image dataset. Different combinations of feature extraction techniques can be adopted to improve the recognition accuracy for image classification. During surveying, I observe that a single approach taken individually does not always accurately recognize all images. As a result, the critical and synthesis analysis of the image classification is based on the data surveyed in this chapter. Table 2.1 provides a comprehensive assessment of several research approaches for image classification. Table 2.1 displays the feature extraction approaches and classification algorithms with high accuracy that have been provided by researchers from around the world.

**Table 2.1.** Image classification results on Caltech datasets using various feature extraction techniques

AUTHOR	DATASET	NO. OF CATEGORIES	FEATURE EXTRACTION	CLASSIFICATION TECHNIQUE	ACCURACY
Mahantesh <i>et al.</i> (2015)	Caltech-101 & Caltech-256	101 256	PCA	Neural Networks and four distance measures	Caltech-101: 67% (on 15 training images) 73%(on 30 training images) Caltech-256: 29.6%(on 15 training images) 36%(on 30 training images)
Huang <i>et al.</i> (2017)	Caltech-101	101	SIFT	Classification pipeline	75.7%
Mahmood <i>et al.</i> (2017)	Caltech-101	101	ResNet-152	PCA-SVM classifier	94.7%
Li (2018)	Caltech-101	101	CNN	CNN	75.93%
Rashid <i>et al.</i> (2018)	Caltech-101	101	VGG16, AlexNet and SIFT	Ensemble boosted tree	89.7%
Gupta <i>et al.</i> (2019)	Caltech-101	101	SIFT, ORB	Random Forest	85.4%
Wu and Zhou (2021)	Industrial Components	Not mentioned	Discriminating features	CNN	91.0%
Bansal <i>et al.</i> (2021)	Caltech-101	101	Shi-tomasi	Random Forest	86.4%

## 2.4. Chapter Summary

This chapter provides a comprehensive analysis of various image classification techniques. A large number of image datasets have been reviewed for pre-processing, feature extraction, feature selection, and classification tasks. A few feature extraction methods have been explored and used to extract image features from images. A few feature selection techniques were also presented by various scholars in the past. A study of several image classification strategies provided by various researchers and the accuracy attained on various datasets, including a separate analysis of accuracy on the

Caltech-101 image dataset, has been elaborated. Research on several approaches to the fusion of multiple feature extraction algorithms is also highlighted.

## Chapter 3

# Proposed Methodology

---

### 3.1. Introduction

An image classification methodology is presented in this chapter, which uses a hybrid learning method combining handcrafted and deep-activated features. Image classification systems are commonly used in a variety of real-world applications, such as computer vision, medical imaging, object recognition, and security surveillance systems that use a content-based image retrieval system in order to gain better insight into images. The purpose of the image classification system is to give a label or class to the specified object in the image. In order to accomplish this task, various features of the image are extracted in order to be used as a tool for identifying the appropriate class for the image, regardless of other classes that are present in the image. The proposed classification method for images is based on local handcrafted features (SIFTs and Haralick textures) that are used in conjunction with deep activation features (VGG19) to test its effectiveness. Further, various machine learning classification methods, i.e., k-NN, MLP, Random Forest, and adaptive boosting classifiers, are used to classify the images based on the extracted features. An experiment was conducted using a benchmark dataset called Caltech-101, which was created by the Caltech research team, and consists of a collection of noisy, rotated, and different scaled images that we were able to compare our results with.

### 3.2 Feature Extraction Techniques

In an image, features are the properties that distinguish it from other images. As well as the number of features used in an image classification system, the effectiveness of the system also depends on its size. As more of the features of an image are revealed, it becomes easier for the system to recognize the image. Features of an image can be categorized as handcrafted or deep features. Handcrafted features are further divided into two classes, namely, global features and local features. Histogram Oriented Gradients (HoG), and invariant moments (Hu, Zernike) are the algorithms used for global feature extraction. Local features describe the key points of the image. SIFT, SURF, Local



Binary Pattern (LBP), Binary Robust Invariant Scalable Key-Points (BRISK), and Haralick texture descriptor are some of the examples of local feature extraction algorithms. Deep learning is used to extract the spatial layout of the image based on the features of the image. A number of state-of-the-art deep convolutional neural networks (DCNNs) have been employed for extracting the deep learning features, including AlexNet, VGGNet, ResNet-50, that are considered to be state-of-the-art DCNNs. Local and deep learning features are widely used for image classification. As a result, we considered a hybrid learning of deep and local features for image classification. These local and deep-activated features are discussed in the following sub-sections.

### 3.2.1 Local Features

Local features describe the pattern or key points of an image, such as points, edges, texture, or key point descriptors. Local feature extraction algorithms, namely, SIFT and Haralick texture descriptors, are mentioned in this section as follows:-

#### A. *Scale Invariant Feature Transform (SIFT)*

SIFT stands for Scale Invariant Feature Transform. SIFT algorithm is a local feature extraction algorithm developed by Lowe (2004). This algorithm extracts the local features of the image. This feature extraction method is invariant to scale, rotation, and lighting. It can extract features from low-resolution images. So, it has achieved great popularity for various computer vision applications like recognition of various types of objects, content-based image retrieval, medical imaging, object tracking, robotics, etc. The method works in four phases.

- *Scale-space extrema detection-* Difference-of-Gaussian (DoG) is used to detect the location of the object and identify the potential keypoints that are invariant to orientation and scale.
- *Keypoint localization-* The localized key points are then refined using a threshold to remove low contrast and edges from the images.
- *Orientation assignment-* The second step is followed by assigning various rotations to the image. After completing this step, I will be able to calculate the gradient direction that makes the image rotation insensitive as a result.

- *Key point descriptor*- As a result of transforming these key points into 128-dimensional vectors, you get a feature vector of 128 dimensions. These are the key features that describe the image.

#### B. *Haralick texture descriptor*

Haralick invented a statistical texture descriptor that uses statistical methods to extract the features of an image by analyzing its texture, in order to determine what features it has. The method introduced by Haralick *et al.* (1973) uses a gray level co-occurrence matrix (GLCM) to analyze the spatial distribution of gray values, which is followed by the computation of local (texture) features from each point. The method is invariant to translation and rotation. This algorithm computes fourteen features from GLCM that describe inherent properties of the image (e.g., smoothness, directionality of the pattern, image complexity, etc.). Out of fourteen features, one feature computes unstable results that are not considered in the code. The features obtained from the algorithm are mean, contrast, correlation, homogeneity, energy, entropy, angular second moment (ASM), sum entropy, difference entropy, and four moments (m1, m2, m3, m4). These features reflect the sensitivity of texture, the correlation of the pixels with their neighbouring pixels, the similarity of the pixels, and the uniformity of the grayscale distribution.

### 3.2.2 Deep Features using VGG19

VGGNet (named Visual Geometry Group) is a very popular deep convolutional neural network architecture for image classification that was proposed by Simonyan *et al.* (2014). Datasets with a very large number of images are used to train the model, i.e., the ImageNet dataset, which consists of over 1000 categories and over 14 million images. The model has achieved second place in ILSVR (2014). VGGNet consists of 19 layers, 16 of which are convolutional layers used for feature extraction, and three of which are fully connected layers used for classification. The VGG19 model is designed by grouping the first 16 layers into 5 groups that are stacked over each other. Each group is followed by max-pooling layers, which reduce the size of the feature map constructed from each convolutional group. Then a group of fully connected layers is constructed for classification using the softmax layer. The model has small (3×3) convolutional filters, which enable the model to increase its depth effectiveness. It extracts low-level information about the image, and these features improve the task of image classification's

effectiveness. VGG19 is shown as a diagram in Figure 3.1, which illustrates its architecture.

Layer (type)	Output Shape	Parameters	
input_1 (InputLayer)	(None, 224, 224, 3)	0	
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792	
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928	
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0	
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856	
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584	
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0	
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168	
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080	
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080	
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0	
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160	
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808	
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808	
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0	
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808	
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808	
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808	
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0	
flatten (Flatten)	(None, 25088)	0	
fc1 (Dense)	(None, 4096)	102764544	
fc2 (Dense)	(None, 4096)	16781312	
predictions (Dense)	(None, 1000)	4097000	

Feature Extraction

Classification

**Fig. 3.1.** Architecture of Deep Convolutional Neural Network

The model uses an RGB image of size 224x224 pixels as input and outputs the label of the image. Initially, the image is pre-processed before training the model to extract the key features from it. I have extracted deep features using the pre-trained VGG19 model followed by various machine learning classification algorithms to examine the

performance of the proposed model. The model extracts a total of 250×88 features from an image after the feature extraction phase, which is very large in size.

### **3.3. Feature Dimension Reduction Techniques**

The size of feature vectors created using various feature extraction algorithms might be quite large. This may pose a variety of issues when determining the most suitable label for the image. Some of these are over-fitting issues, the need for a huge amount of space, the longer execution times, etc. Various feature selection and dimensionality reduction strategies can be used to overcome these difficulties. To reduce the likelihood of overfitting, the important features are selected from a feature vector to be used in feature selection. Dimensionality reduction is used to reduce the dimensionality of the feature vector, allowing for faster computations. The following techniques were used in the work: k-means clustering for feature selection and the Locality Preserving Project (LPP) for dimensionality reduction.

#### **3.3.1. k-means Clustering Algorithm**

Using a distance measure, the k-means clustering algorithm (Kanungo *et al.*, 2002) divides a dataset into k groups. Euclidean distance, sometimes called max-min distance, is used to calculate the distance between the centroid of the cluster and the object. There are two outcomes of the k-means clustering algorithm - the centroids of the k clusters and the labels of the training data. In this example,  $x$  represents the data point and  $c_i$  represents the  $i$ th centroid of the k clusters. Murphy-Chutorian and Triesch (2005) introduced an image classification system based on appearance. They extracted information solely at corner locations using a Gabor wavelet response. The k-means clustering technique was used to generate a feature dictionary using the features obtained in the first phase. The images were cropped during the training phase to distinguish between the foreground and background regions. At each foreground point of interest, the Gabor-jet was retrieved and compared to the feature dictionary. In addition, a bottom-up search strategy was used during the recognition phase of the process to compare the Gabor-jet features recovered during the recognition phase of the process with the feature dictionary developed at the beginning of the process. The recommended approach expedites the process for the classification of images.

### **3.3.2. Locality Preserving Projection (LPP)**

He and Niyogi (2004) developed the nonlinear Laplacian Eigenmap's linear approximation. Classification computational costs are reduced by LPP. Shermina (2010) compared LPP with PCA on the AT&T Face Database. Gupta *et al.* (2019) described SIFT and ORB feature extraction algorithms for image classification. To eliminate dimensionality and select features based on their study, the authors used a projection and clustering technique that preserves locality. They compared three-dimension feature vector sizes—8, 16, and 32 dimensions—and revealed that the feature vector of size 8 generated the best results. Finally, they proposed a hybrid strategy for the image classification system that would employ both feature extraction approaches. In this work, I reduced the size of the feature vector using each strategy to eight components with the LPP feature reduction method.

### **3.4. Classification Techniques**

Also, it is important to keep in mind that even though the feature extraction phase is an important part of image classification, it is not the only part of the process, as the quality of the features extracted from the images in the previous phase is also crucial to the efficiency of the classifiers. The features retrieved using the different techniques outlined in sub-section 2.1.2 are then used to categorize the images for the classification task. In order to classify the images, various classification techniques are used. To categorize the images using various classification techniques, a supervised learning approach is employed as part of a supervised learning process. I use supervised learning to identify the proper label for an image based on training data that is matched with the input data. There is no training data accessible in unsupervised learning. In this work, I have considered k-NN, MLP, random forest, and adaptive boosting classifiers to classify the image because these classification models are achieving acceptable results in various domains like content-based image retrieval, document analysis and recognition, computer vision, etc. There are four classification techniques that are briefly discussed in the following sub-headings:

### **3.4.1. k-NN**

The technique determines the appropriate name for the image by analysing the distance between the vectors included in the image and selecting the class of information focuses most close to the testing data of interest by assessing the distance between them. Then, at that point, the class with the most prominent worth of the testing information is returned with a greater portion of the vote. On the off chance that two image classes are provided, the k-NN technique, for example, returns the classification with the shortest distance between every piece of interest in the created image dataset and every piece of interest in the testing classification. In other words, when this algorithm is fed testing data, it assigns the class name that is closest to the data points in the test data. Various distance measures, including Euclidean, Manhattan, Minkowski, and Hamming distances, are used to calculate the distance for the k-NN algorithms. The majority of classifiers employ Euclidean distance. Kim *et al.* (2012) described and compared the k-NN and SVM classifiers. For the recognition challenge, they employed a Caltech-4 cropped image dataset. The Scale Invariant Feature Transforms (SIFT) technique was used to implement the Bag-of-Words (BOW) algorithm. Using the k-means clustering technique, the calculated characteristics were then grouped into comparable groups. Training data was obtained by creating a histogram of these code phrases. This information was then categorized using the k-NN and SVM classifiers. The SVM outperformed the k-NN classifier with an accuracy of 90.56%. Muralidharan (2014) proposed an eigenvalue-based k-nearest neighbour classifier. They used the image's local and global features to classify the 2D images using k-NN classification.

### **3.4.2 MLP**

MLP stands for multi-layer perceptrons, which is a member of the family of artificial neural networks. In the current study, multi-layer perceptrons (MLP) were also used for classification. The MLP-based classifiers were trained using a back propagation learning algorithm with a learning rate of 0.3 and a momentum term of 0.2, that employed a back propagation learning algorithm. In this work, the Weka tool was used for MLP-based classification.

### **3.4.3. Random Forest**

The Random Forest (RF) algorithm is a supervised classification technique (Breimann, 2001). Random forest is an extension of the decision tree classification. Various self-learning decision trees are used by RF (i.e., "forest"). The goal of RF is to reduce the heterogeneity of the two generated data subsets. In contrast to the manual (expert-based) formulation of decision rules, the RF employs self-learning decision trees. It is made up of various decision trees that are generated throughout the dataset's training. The tree's height is increased by the use of several randomized methods. The bulk of decisions generated by k-decision trees are used to forecast the class of an image in this case. Using the random forest algorithm, you can take the average of all predictions that have been made from the decision trees, which allows for the elimination of biases that could be present in the decision trees. It is suitable to work with missing values by preserving good accuracy. It can be used for both regression and classification problems. When compared to a single decision tree, it works well for a large collection of data items.

### **3.4.4. Adaptive Boosting**

AdaBoost is an abbreviation of adaptive boosting, which was proposed by Freund and Schapire (1997). It is an ensemble algorithm that transforms a group of weak learners (classifiers) into strong learners. The base learners are created with the help of a one-dimensional decision tree having one depth, and these decision trees are called decision stumps. Adaboost, or adaptive boosting, is the name for an ensemble boosting classifier. As a result of combining many classifiers, the accuracy of the classifier can be increased. The AdaBoost classifier combines a number of weak classifiers into a single powerful classifier in order to obtain high accuracy. A simple classifier can be any algorithm that accepts weights from the training set. Adaboost should meet two conditions, including that the data samples are representative of real-world observations. It is a machine learning strategy that uses a weighted sum of the output of many classifiers to achieve adaptive boosting results. The major goal of this strategy is to improve recognition accuracy by transforming weak classifiers into strong ones. According to Guo and Zhang (2001), a face recognition technique based on adaptive boosting has been proposed for the recognition of faces. It is known that the majority voting method can be used for

resolving multi-class recognition problems, while AdaBoost is a classifier that is commonly used to discriminate between two classes only. A majority voting system involves combining the results of many classifiers and determining the winner based on a weighted average of the combined results. The purpose of this study was to explore adaptive boosting methodologies as a means of improving classification results for image classification

### **3.5. Proposed Methodology**

This section explains the steps involved in the entire process of classifying images, which can be divided into several parts describing each stage. Using the proposed system, it is possible to classify images faster and more accurately than ever before. Fig. A diagram of the proposed system's architecture is presented in Figure 3.2.

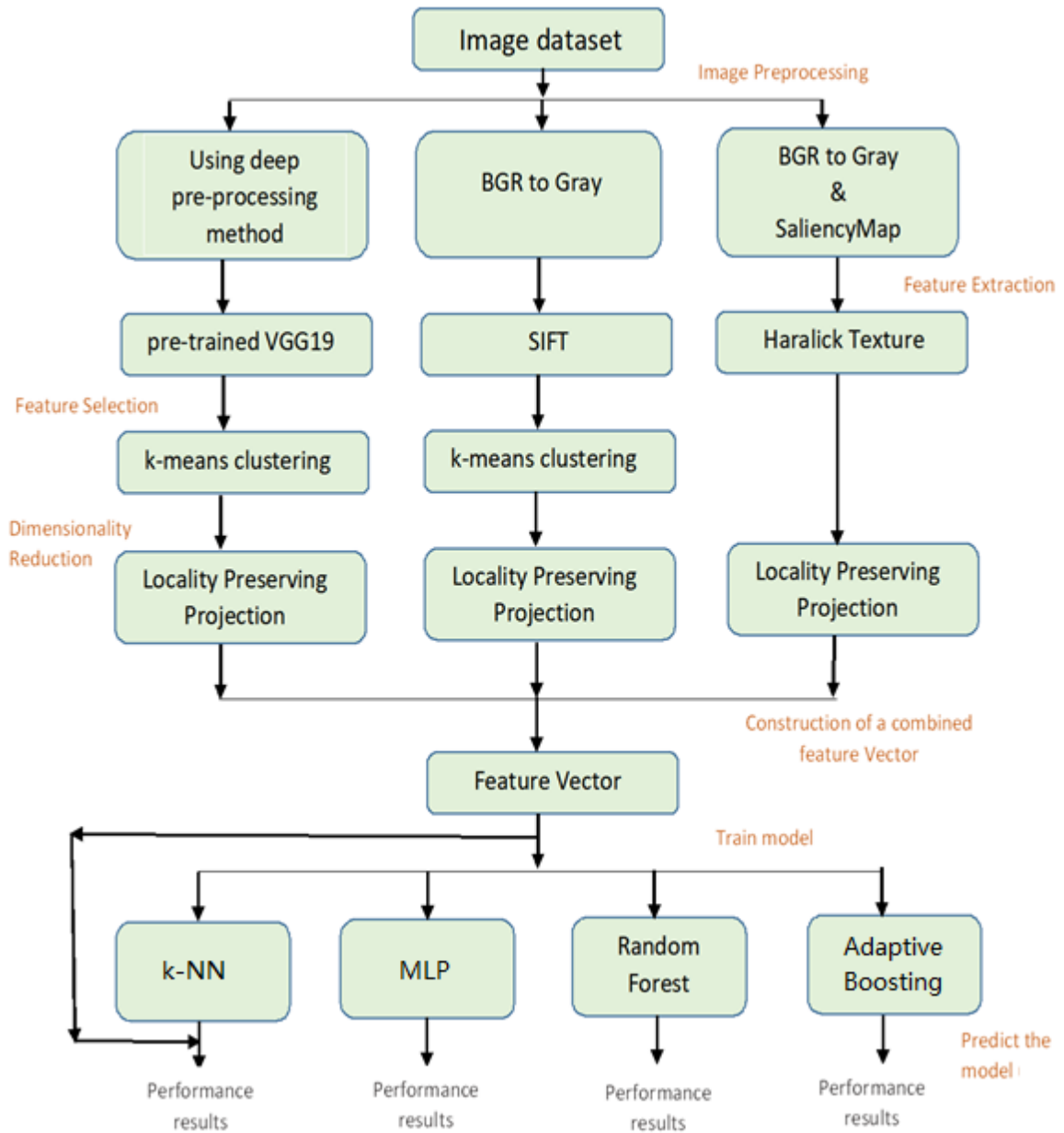
- a. The system starts with the acquisition of the image dataset. Here, a benchmark image dataset called Caltech-101 is used for the experimental work. There is a large number of images in this collection that are classified into 101 categories and one category that is dedicated to background scenes.
- a. Then features are extracted using three feature extraction methods, i.e., the pre-trained VGG19 model, SIFT, and the Haralick texture descriptor. All these methods require preprocessing before extracting the features of the images. The colour images are processed using the pre-processing method of the Keras toolkit and then input to the pre-trained VGG19 model. For the SIFT algorithm, the colour images are to be transformed into grayscale. Haralick's texture descriptor extracts the information using a gray-level co-occurrence matrix (GLCM), resulting in grayscale images. The images in the Caltech-101 dataset are noisy and have invariant illumination. So, saliency maps (proposed by Montabone and Soto, 2010) are also applied to images to differentiate the visual features of the image and make a meaningful representation of the image. Further, threshold maps are also enforced on the images to extract each salient region.
- b. After the pre-processing of the images, feature extraction algorithms are applied to the processed images. They extract a set of features from the images and store



them as a feature vector. VGG19, a pre-trained deep model, extracts 25088 features from an image. The SIFT algorithm extracts 128 features from an image, and the Haralick texture algorithm extracts 13 features from the image.

- c. Both VGG19 and SIFT obtain a large number of features, so both algorithms need a feature selection algorithm to select key features of an image. So, there is a need for a feature selection algorithm that removes the irrelevant features of the image and makes the recognition process very fast and more accurate. In the experiment, 64 key features are selected using the k-means clustering algorithm. The k-means clustering algorithm uses the Euclidian distance to process the features.
- d. Once the feature vector is selected using the above-mentioned method, the dimension of the feature vector can be significantly reduced by using a locality-preserving projection (LPP). In the experimental work, eight dimensions are computed from features of size 64 using LPP.
- e. Next, a feature vector is formed by combining the features extracted from all three feature extraction methods.
- f. Four models are now trained using four classifiers: k-NN, MLP, Random Forest, and the adaptive boosting classifier. A data-partitioning strategy is used to train the model. According to a standard 80:20 data partitioning, 80% of the data in the experiment is considered training data and 20% is considered testing data, in the experiment, according to the 80:20 partitioning.

In terms of recognition accuracy, precision, recall, and root mean square error, F1-scores, area under curve, and root mean square error indicate the success of the system.



**Fig 3.2.** Block diagram of the proposed system

### 3.6. Research Questions

The following are the ways in which this dissertation responds to each of the research questions as listed in Chapter 1 (Section 1.6):

Q1. What are the various approaches of image classification?

- In the literature review, many approaches of image classification are addressed, and a critical analysis is also provided.

Q2. Which strategy is best for image classification?

- Several studies have shown that the combination of local and deep-activated features can be used to improve the classification of images in the literature.

Q3. What are the various datasets available for image classification?

- In order to conduct experiments on image classification, there are a large number of datasets available, such as Corel-100, Caltech-101, Caltech-256, MIT-67, Flower-102, VOC-2007, VOC-2012, New-BarkTex, Outex-TC13, UIUC Sports, etc. These datasets contain varying numbers of classes and samples within each class. A few datasets for the experiment analysis are publicly accessible. We used only one dataset for our experiments in this study, Caltech-101, which is one of the publicly available datasets; we discussed this dataset thoroughly in Chapter 1 of this project work.

Q4. What are the various problems that are faced in image classification?

- A few problems with image classification include occlusion, viewport variations, affine translation, fluctuating object counts, and lighting. As a result, in this study, we investigated a hybrid learning of local features and deep-activated features to achieve acceptable image classification results.

### **3.7. Validity and Reliability**

The proposed approach given in this dissertation is assessed using the benchmark public dataset. Experimental findings based on recognition accuracy, precision, recall, the F1-score, and the RMSE were also produced to validate the proposed methodology. I have also taken Area Under Curve into consideration to confirm the classifier's efficacy for image classification.

### **3.8. Ethics and Bias**

Ethics: This material is the authors' own original work, which has not been previously published elsewhere. For the experimental study, we considered the Caltech-101 public dataset for image classification. With the help of specific feature extraction methods, the model is trained using our own created features, and WEKA machine learning software is employed for the training of the models and classification.

Bias: Image classification increases the likelihood of false arrests in computer vision and machine learning by producing a large percentage of false identifications among occlusive images. Additionally, it is far more possible that many things may be recognized in a single image, distorting the evidence and impacting investigations. Therefore, a substantial quantity of data is required to train the model using this method.

### **3.9. Limitations**

Infants are able to make causal models that predict the structure of their environment. This understanding enables them to learn from limited amounts of data and apply true generalizations. So, a large dataset is also required for the image classification and to achieve acceptable classification results. There are still significant obstacles that must be removed before we can achieve the objectives of comprehending biological vision systems and general-purpose artificial intelligence.

### **3.10. Chapter Summary**

In this chapter, I have presented a block diagram and various techniques used in this dissertation work. Various feature extraction methods covered in this chapter include the local features, namely, SIFT and Haralick descriptors, along with the deep-features, namely, VGG19, used in this proposed work. The research's images were then classified using a range of classification methods, including k-NN, MLP, Random Forest, and adaptive boosting. A few challenges with image classification are also discussed in this chapter.

# Experimental Results and Discussion

---

This chapter aims to describe the experimental dataset used for the study, our programming tools, as well as the performance evaluation parameters that were used for the performance evaluation of the study. In this chapter, the experimental results are presented using the features individually and using the different combinations of all three feature extraction techniques considered in this dissertation work. The experimental work presented in this thesis was conducted on one of the most challenging datasets, the Caltech-101 image dataset. A number of performance quality factors are used to evaluate the effectiveness of the proposed system, including recognition accuracy (ACC), precision (P), recall (R), F1-Score, area under curve (AUC), and root mean square error (RMSE).

### **4.1 Dataset**

To test the effectiveness of the proposed algorithm, the Caltech-101 image dataset was used as a challenging multi-class dataset for image classification. The images in the dataset were compiled by Fei-Fei *et al.* (2004). The dataset contains 101 object classes and one background scene with 9146 images. The proposed system is implemented for 101 object classes, which comprise a total of 8677 images. The Caltech-101 dataset is an unbalanced dataset because each class contains 40–800 images. The researchers perceived all these factors in the experiment. The Caltech-101 dataset is a collection of noisy images, images with variations in lighting, and differently geometrically shaped images.

### **4.2 Software and code implementation**

Python was used to build the proposed model since it is a simple and effective programming language that works with OpenCV to provide functionality for manipulating images. Basically, OpenCV is an open-source library that combines a number of computer vision-based algorithms and allows them to be used together in a single software package. It is an open-source BSD-licensed library that is an open-source

computer vision library that is completely open-source. The implementation of the system's main functionalities is thoroughly explained in this section. Hybrid learning of handcrafted features (SIFT, Haralick) and deep-activated features (VGG19) is considered in this work for image classification. In order to cluster the data, k-Means is used, and in order to reduce the dimensionality, LPP is used. In order to predict performance, K-NN, MLP, random forest, and adaptive boosting classifiers are used.

#### ***4.3 Performance Evaluation Parameters***

Various performance assessment metrics have been taken into account when conducting experimental work. The parameters for multi-class classification have been used to assess these measures. When using a dataset with many classes, each class must be mutually exclusive. The 101 classes in the Caltech-101 dataset each have exactly one instance or image attributed to them. For instance, a flower may only be either a lotus or a sunflower at once. The performance of multiclass classifiers can be determined by averaging the evaluation measures over a number of classes in order to determine the overall performance of the classifier. After implementing the image classification system, the confusion matrix result is achieved. Performance evaluation parameters are recognition accuracy (ACC), precision (P), recall (R), f1-score (F), area under curve (AUC), and root mean squared error (RMSE). To compute these measures, a confusion matrix is generated, as shown in Figure 4.1. As the experiment is conducted on a multi-class image dataset, all these measures are practised using multi-class classification. The mathematical definition of all these parameters over a few classes ( $n$ ) is discussed as follows:


$$\text{Accuracy (ACC)} = \frac{1}{n} \cdot \sum_{i=1}^n \frac{\text{Number of correct predictions of } i^{\text{th}} \text{ class}}{\text{Total number of actual values}}$$
$$\text{Precision (P)} = \frac{1}{n} \cdot \sum_{i=1}^n \frac{\text{Number of correct predictions of } i^{\text{th}} \text{ class}}{\text{Total number of predictions values}}$$

37

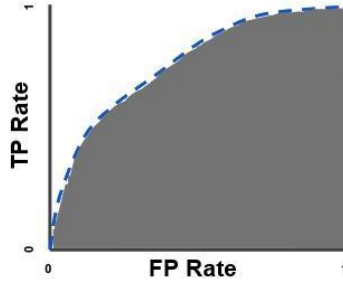
$$TPR = \frac{TP}{TP+FN}$$

$$Recall (R) = \frac{1}{n} \sum_{i=1}^n TPR_i$$

As per Godbole and Sarawagi (2004), the F1-Score is calculated by averaging the harmonic mean of the precision and recall obtained for each class.

$$F1 - Score = \frac{1}{n} \sum_{i=1}^n \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i}$$

The effectiveness of the classification model is estimated probabilistically using the area under the curve (AUC). AUC has a value between 0 and 1. The accuracy of the model's predictions increases with the AUC. TPR vs. FPR is plotted on the graph's y-axis with FPR on the x-axis to calculate AUC using the ROC curve (refer to Figure 4.2).



**Fig. 4.2.** Area Under Curve (AUC)

The standard deviation of the anticipated mistakes is known as the root mean squared error (RMSE).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (g_i - o_i)^2}{n}}$$

where  $g$  is the forecast outcomes and  $o$  is the genuine outcomes and  $n$  is the total number of features.

#### **4.4 Experimental Results**

This subsection presents the comparative analysis of four classification methods and three feature extraction algorithms, i.e., SIFT, VGG19, and Haralick texture descriptor. Caltech-101 will be used as a benchmark dataset to assess the quality of the proposed work in order to evaluate the quality of the proposed work. The experiment was carried out using four different algorithms of classification: k-NN, MLP, Random Forest,



and adaptive boosting. An 80% training set of data was used in the proposed recognition model, while 20% of the data was used in the testing set. The experimental results show that these features, i.e., SIFT, VGG19, and Haralick texture descriptors, individually do not perform well on their own for image classification. As a result, for the experimental study and to improve the results, a combination of these feature extraction algorithms is being considered. The comparison among all the methods is represented with the help of a few tables. Each table presents information about each performance measurement parameter separately. These results are also graphically depicted in Figs. 4.3–4.8. Finally, the results achieved through this experiment reveal that an image can be improved by a collection of features.

The recognition accuracy computed through the above-mentioned techniques is shown in Table 4.2. The results reveal that a combination of SIFT, VGG19, and Haralick texture descriptors has achieved an accuracy of 99.07%, which is the highest accuracy obtained using the random forest classifier. Even the MLP classification algorithm determined 98.03% accuracy, and the adaptive boosting achieved 98.67% accuracy.

The comparison of precision and F1-Score is shown in Tables 4.3 and 4.5, respectively. Here, the MLP classifier presents the highest results for precision and F1-Score, i.e., 98.13% and 98.05%, respectively. The results for recall are presented in Table 4.4, which is the maximum using random forest (99.07%). Results for the area under curve are shown in Table 4.6, where the random forest classifier achieved the highest results, i.e., 99.53%. Table 4.7 depicts the comparative analysis for the root mean squared error. Using the adaptive boosting methodology, the root mean squared error is 5.42%, which is the minimum of all other cases.

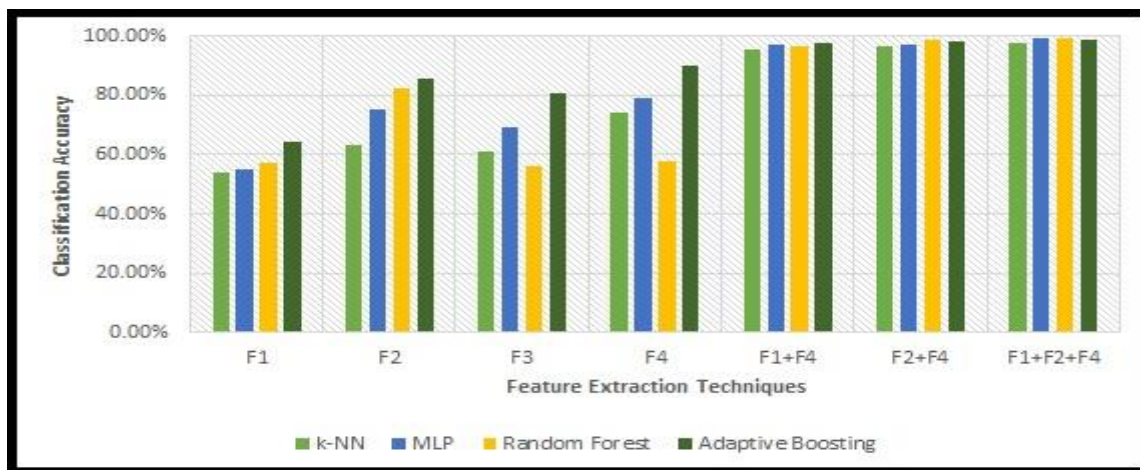
Consequently, researchers observed that random forest and MLP classifiers achieve the best results for all parameters among other methods when applied to a combination of SIFT, VGG19, and Haralick texture descriptors. The comparative analysis of the proposed feature extraction approach among these classification algorithms is depicted in Fig. 4.3 based on recognition accuracy.

**Table 4.1.** Feature Extraction Techniques

F1	SIFT
F2	VGG19
F3	Haralick (without Saliency)
F4	Haralick (with Saliency)
F1+F4	SIFT+Haralick (with Saliency)
F2+F4	VGG19+Haralick (with Saliency)
F1+F2+F4	VGG19+SIFT+Haralick (with Saliency)

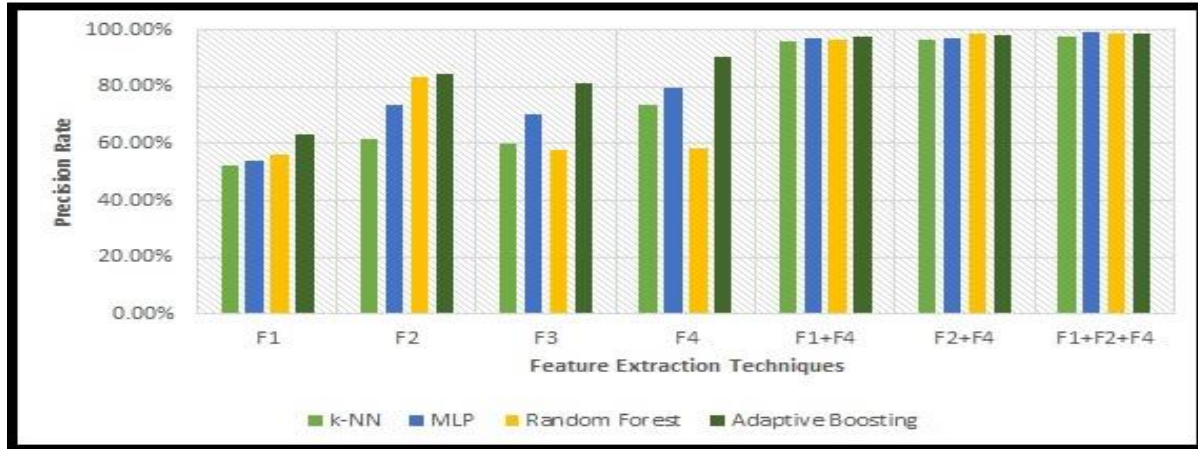
**Table 4.2.** Classifier Wise Recognition Accuracy

Features	Classification Techniques			
	k-NN	MLP	Random Forest	Adaptive Boosting
F1	54.20%	54.84%	57.13%	64.43%
F2	63.45%	74.99%	82.11%	85.64%
F3	61.00%	69.26%	55.93%	80.73%
F4	74.15%	79.32%	57.83%	90.23%
F1+F4	95.40%	97.01%	96.58%	97.86%
F2+F4	96.43%	97.25%	98.83%	98.11%
F1+F2+F4	97.65%	98.03%	99.07%	98.67%

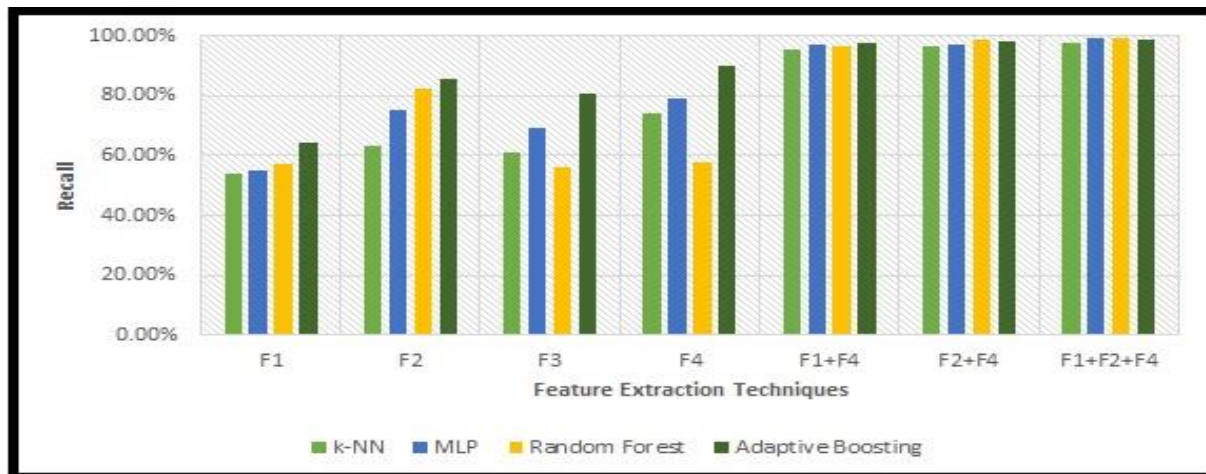
**Figure 4.3.** Classifier Wise Recognition Accuracy

**Table 4.3.** Classifier Wise Precision Rate

Features	Classification Techniques			
	k-NN	MLP	Random Forest	Adaptive Boosting
F1	52.23%	54.23%	56.04%	63.52%
F2	61.85%	73.89%	83.34%	84.76%
F3	60.20%	70.39%	57.58%	81.38%
F4	73.75%	79.71%	58.59%	90.32%
F1+F4	95.79%	97.08%	96.69%	97.88%
F2+F4	96.38%	97.21%	98.93%	98.03%
F1+F2+F4	97.70%	98.13%	98.97%	98.70%

**Figure 4.4.** Classifier Wise Precision Rate**Table 4.4.** Classifier Wise Recall Rate

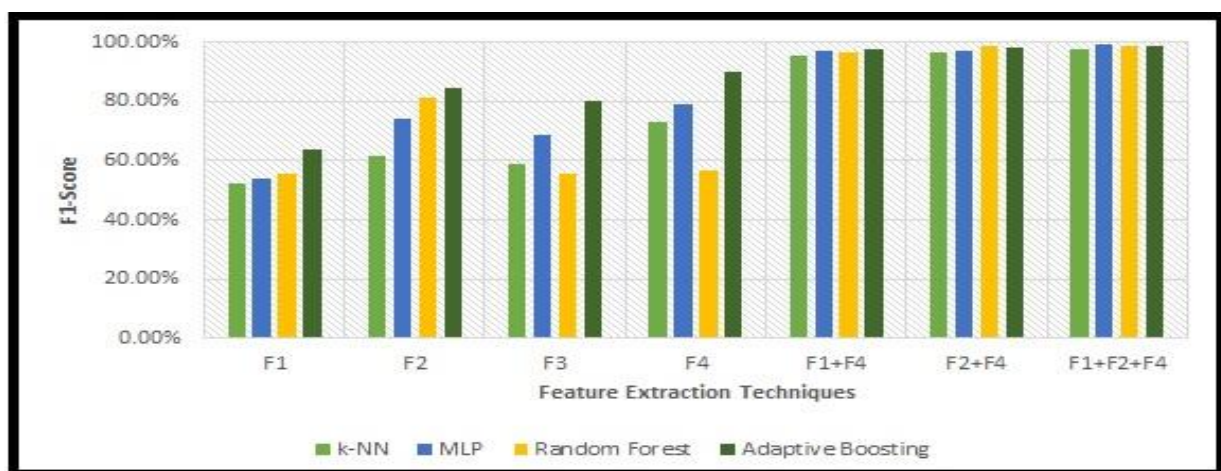
Features	Classification Techniques			
	k-NN	MLP	Random Forest	Adaptive Boosting
F1	54.20%	54.84%	57.13%	64.43%
F2	63.45%	74.99%	82.11%	85.64%
F3	61.00%	69.26%	55.93%	80.73%
F4	74.15%	79.32%	57.83%	90.23%
F1+F4	95.40%	97.01%	96.58%	97.86%
F2+F4	96.43%	97.25%	98.83%	98.11%
F1+F2+F4	97.65%	98.03%	99.07%	98.67%



**Figure 4.5.** Classifier Wise Recall Rate

**Table 4.5.** Classifier Wise F1-Score

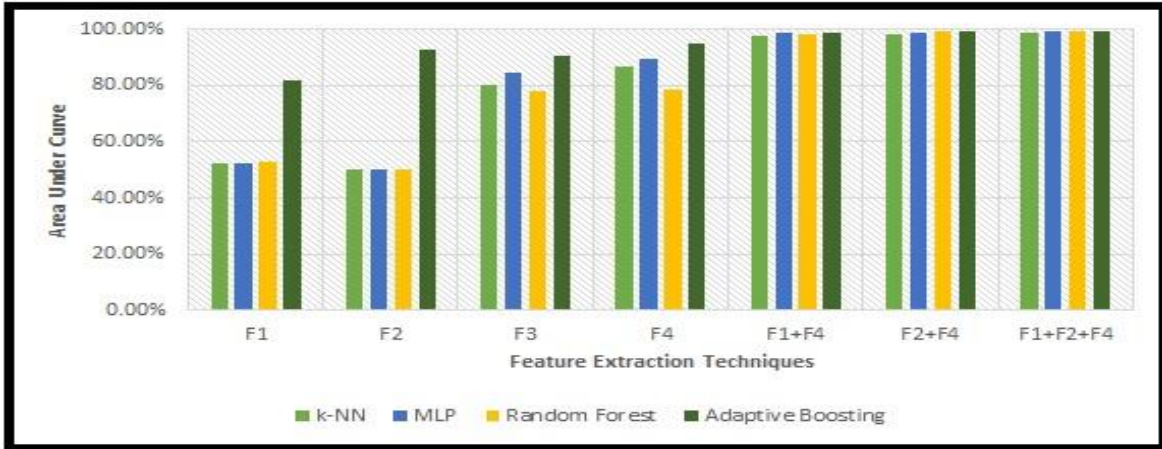
Features	Classification Techniques			
	k-NN	MLP	Random Forest	Adaptive Boosting
F1	52.32%	54.05%	55.83%	63.60%
F2	61.58%	73.92%	81.49%	84.78%
F3	59.00%	68.82%	55.61%	80.31%
F4	73.02%	79.00%	56.89%	89.81%
F1+F4	95.42%	96.99%	96.54%	97.77%
F2+F4	96.29%	97.16%	98.83%	97.99%
F1+F2+F4	97.62%	98.05%	98.98%	98.65%



**Figure 4.6.** Classifier Wise F1-Score

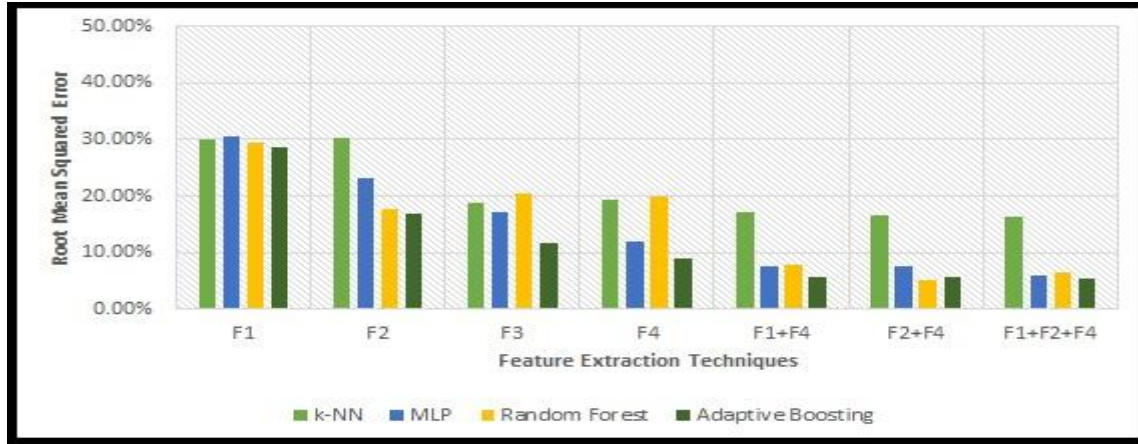
**Table 4.6.** Classifier Wise Area Under Curve

Features	Classification Techniques			
	k-NN	MLP	Random Forest	Adaptive Boosting
F1	52.48%	52.56%	52.70%	81.96%
F2	50.22%	50.31%	50.43%	92.73%
F3	80.36%	84.52%	77.83%	90.30%
F4	86.96%	89.59%	78.79%	95.08%
F1+F4	97.65%	98.49%	98.27%	98.92%
F2+F4	98.17%	98.61%	99.41%	99.04%
F1+F2+F4	98.78%	98.50%	99.53%	99.33%

**Figure 4.7.** Classifier Wise Area Under Curve**Table 4.7.** Classifier Wise Root Mean Squared Error

Features	Classification Techniques			
	k-NN	MLP	Random Forest	Adaptive Boosting
F1	30.02%	30.67%	29.35%	28.60%
F2	30.37%	23.08%	17.66%	16.82%
F3	18.67%	17.09%	20.36%	11.76%
F4	19.44%	12.06%	20.00%	9.01%
F1+F4	17.10%	7.59%	7.93%	5.63%
F2+F4	16.56%	7.52%	5.77%	5.65%
F1+F2+F4	16.40%	6.09%	6.51%	5.42%





**Figure 4.9.** Classifier Wise Root Mean Squared Error

#### 4.5 Chapter Summary

In this chapter, I have considered the hybrid learning of handcrafted local features and deep-activated features for image classification. Handcrafted local features are extracted using SIFT and Haralick descriptors, and deep-activated features are extracted using the VGG19 model. This chapter reveals that the fusion of these methods achieves efficient results as compared to individual techniques. Four multi-class classifiers, i.e., k-NN, MLP, Random Forest, and adaptive boosting, are experimented with for each feature extraction method and with the combination of all three features considered in this work. The proposed system states that an image can be precisely described by the combination of features, and the random forest classifier is the best classifier among all. Experiments are conducted with a dataset that comprises 101 categories with a total of 8677 images. The results of the experiments are presented in the chapter as accuracy, precision, recall, F1-Scores, area under curves, and root mean squared errors as measures of recognition accuracy, precision, and recall. The accuracy of the fusion of features obtained using random forest is 99.07%. Other computations for random forest areas are precision (98.97%), recall (99.07%), F1-Score (98.98%), and the area under the curve (99.53%), which is the best among other classifiers. The adaptive boosting classifier shows the minimum root mean squared error (5.42%). Consequently, the proposed system using combined features and a random forest classifier is potentially useful for image classification.

# Conclusion and Future Directions

---

### 5.1. Introduction

An image classification system's aim is to recognize and classify image, what are the various approaches and objects that remain in these images. Many researchers have failed to achieve satisfactory results despite numerous attempts due to issues such as occlusion, lighting effects, geometrical variance, cloudy images, background clutter, processing complexity, etc. This dissertation is a research project for 2D image classification that provides an efficient solution for 2D image classification. A technique for classifying images using local and deep activation features has been proposed in this work, which combines local and deep activation features. This dissertation shows how various techniques are used in the various stages of an image classification system.

In this chapter, we presented concluding notes based on the literature survey conducted during this dissertation work. A large number of datasets are available for image classification experiments, and these datasets can be used to increase the accuracy of the results, such as Corel-100, Caltech-101, Caltech-256, MIT-67, Flower-102, Cub-200, and so on. These datasets contain varying numbers of classes and samples within each class. And these samples include a wide range of local and global image classification features.

This dissertation is divided into five chapters. The first chapter describes the introduction of the dissertation work, the framework, and the dataset used in this dissertation work. The Caltech-101 dataset is one of the most challenging datasets that I have used in my dissertation. This work's literature review is mentioned in Chapter 2. This chapter includes information about the approaches used in various phases of the system. In Chapter 3, a proposed methodology based on hybrid learning of SIFT, the Haralick texture descriptor, and the VGG19 deep learning model is described. Chapter 4 depicts experimental results

obtained using the methodology proposed in this work. Chapter 5 of the project provides concluding notes and future directions for this work.

## **5.2. General Conclusion**

In this dissertation, I have considered the hybrid learning of deep and local features to improve recognition accuracy for image classification. Deep learning and local features are extracted in the project work using the pre-trained deep learning models VGG19, SIFT, and Haralick texture descriptors. The dissertation work reveals that the fusion of these methods achieves efficient results as compared to individual techniques. Four multi-class classifiers, i.e., k-NN, MLP, random forest, and adaptive boosting, are experimented with for each feature extraction technique and with the combination of all features under consideration. The dataset contains 101 object classes (ant, chair, cup, faces, helicopter, etc.) and one background scene. The Caltech-101 dataset is one of the most difficult to analyse because it contains a variety of noisy, varying lighting, and irregularly shaped geometric images. Each image is about 300 by 200 pixels, and each class has 40–800 images. The proposed system states that an image can be precisely described by the combination of features, and the random forest classifier is the best classifier among all. The dataset comprises a total of 8677 images. During the dissertation research, experimental results are presented based on the results of the experiments in terms of recognition accuracy, precision, recall, F1-Scores, the area under the curve, and the root mean squared error, as well as the results of the experiments. The accuracy of the proposed methodology (hybrid learning of local features, namely, SIFT, Haralick, and deep-activated features VGG19) and random forest is 99.07%. Other computations for random forest are like precision (98.97%), recall (99.07%), F1-Score (98.98%), and the area under the curve (99.53%). Adaptive boosting shows the minimum root mean squared error (5.42%). There are five chapters in this dissertation, and a brief summary of each chapter with concluding notes is as follows:

In Chapter 1, I have discussed various phases of the 2D-image classification system and challenges related to the image classification system. This chapter also discusses the system's various applications. In Chapter 2, a review of the literature on the various stages of image classification systems has been presented. This review of



the literature provides essential background knowledge for image classification. In addition, I have also presented a detailed analysis of the results achieved by other researchers. In Chapter 3, a proposed methodology using hybrid learning of SIFT, the Haralick descriptor, and the VGG-19 deep learning model has been presented. Performance comparisons are conducted on an individual basis as well as combined basis between the three features considered in this project work. The k-means clustering algorithm was used for feature selection, and locality-preserving projection algorithm was used for dimensionality reduction. The classifiers used in this work are k-NN, MLP, Random Forest, and Adaptive Boosting.

The detailed experimental results based on the proposed methodology are presented in Chapter 4 using a combination of local features, namely SIFT and haralick texture descriptors, as well as the deep-activated feature VGG19. The classifiers employed were k-NN, MLP, and Random Forest. The work has achieved classification accuracy of 97.65%, 98.03%, and 99.07% using k-NN, MLP, and Random Forest, respectively, on the combination of SIFT, Haralick texture, and VGG19 features. The adaptive boosting technique is also explored to improve the classification results. The experiment has shown 98.67% classification accuracy using adaptive boosting. Various other performance parameters such as true positive rate, precision, recall, RMSE, and area under the curve have also been evaluated for k-NN, MLP, Random Forest, and adaptive boosting.

### **5.3. Research question conclusions**

These are the findings of each research question considered in this dissertation.

Q1. What are the various approaches of image classification?

Ans: There are various phases for image classification, like image acquisition, preprocessing, feature extraction, and classification. Several methods can be used in order to collect data when it comes to the process of acquiring an image, such as using a camera, a sensor, and so on. We can also use a public dataset for the experimental study. During the preprocessing phase, we can explore different techniques, like converting the RGB to grayscale format. In the feature extraction process, there are a

number of features like regional and texture features, LBP features, SIFT, SURF, ORB, histograms of oriented gradients, shape based features, texture based features, content-based features, color-based features, etc. For classification, there are a number of techniques like Gaussian Naïve Bayes, k-NN, MLP, ANN, eXtreme Gradient Boosting, Decision Tree, Random Forest, etc.

Various methods of image classification are discussed in the literature review, and a critical review is also given in Chapter 2. The various techniques used for image classification on different datasets are thoroughly examined in this chapter. Feature extraction and classification work has been done for various image datasets that have been examined. This chapter provides a discussion of several performance metrics used on a multi-class image dataset. The various feature extraction methods used to extract features from an image, as well as the classification algorithms used to classify the images, are also thoroughly addressed. The accuracy gained on multiple datasets, including a separate analysis of accuracy on the Caltech-101 image dataset, and has been elaborated in an analysis of various image classification approaches provided by various researchers. This chapter provides a discussion of several performance metrics used on a multi-class image dataset. A study on multiple methods for the fusion of distinct feature extraction methodologies has been described.

Q2. Which one strategy is the best for image classification?

Ans: One of the most difficult tasks is finding images via searching, indexing, and browsing the vast image databases. In addition, manually managing and annotating each image in huge datasets is laborious and subject to error. In light of this, there is a need for an efficient image classification system, and during the past few decades, this area of study has become quite active. Image classification algorithms are used in order to obtain the images from a large database, depending on the contents needed. Recently, research groups have given more attention to image classification algorithms, which recover an image based on its unique features. A classification approach to image retrieval involves extracting and comparing the properties of a query image with a large dataset of features that are stored for other images in the dataset. A comparison has been made between the two feature vectors to determine the degree of similarity, and the results have been

correlated to provide a yield to the client. Image classification algorithms are used in various applications, including the military, architectural designs, face detection systems, and many others. The transformed domain and the spatial domain are the two domains from which image features are extracted. In the transformed domain, features are extracted using transformed data, whereas in the spatial domain, image features are extracted using pixel values. The domains of space and transformation both extract either global or local characteristics. Various image classification strategies have been put out in the last ten years, all of which are based on the global and local properties of images. The image descriptor can be considered one of the most important factors that affect the speed and accuracy of the image classification system. In order to extract features from an image, it is necessary to map the pixels into a feature space that contains the features of the image. The image is translated into digital form via image processing, and following the conversion, operations are carried out on the image to extract information from it. The fundamental goal of feature extraction is to represent the objects in an image such that their key characteristics and traits may be more clearly seen. The following are a few processes in image processing and matching:

- Local feature detection in images saved in the database
- To create a big dictionary for matching the test images and use the descriptor extractor to describe the characteristics.
- Passing a feature vector to the machine will train it and help it learn about the stored images.
- Identifying and defining the test image characteristics, then comparing them to the feature vector.
- Returning the output
- Features are extracted into two domains for image classification is:
  - Transformed domain: - It uses transformed data to extract features.
  - Spatial domain: - It extracted image features using pixel values in the spatial domain.

In this study, local features, namely, SIFT and Haralick descriptors, and deep features, namely, VGG19 feature extraction methods for extracting image features, were taken into

consideration. These methods are performing well for image classification. They are briefly discussed in Chapter 3.

Q3. What are various datasets available for image classification?

Conclusion: There are a large number of datasets available for image classification experiments, such as Corel-100, Caltech-101, Caltech-256, MIT-67, Flower-102, VOC-2007, VOC-2012, New-BarkTex, Outex-TC13, UIUC Sports, etc. These datasets contain varying numbers of classes and samples within each class. A few datasets for the experiment analysis are publicly accessible. But, in this study, we only used one publicly available dataset, Caltech-101, for experimentation; a thorough description of this dataset is covered in Chapter 1.

Q4. What are the various problems that are faced in image classification?

Conclusion: When attempting to identify items in an image, there are several problems. The following issues came up during image classification:

- **Feature extraction:** It is still a challenge to use a more efficient feature extraction approach to increase retrieval accuracy. There is no set of ideal measurements that would lead to the perfection of image retrieval, despite the fact that several basic image feature measures such as edges, RGB, lines, grey scale, spatial, etc., were employed.
- **Image classification mapping:** Similar images taken from various angles are often regarded as separate.
- **Response time:** For large databases and a large number of users, the image classification system's real-time query response time as well as feature processing time will be lengthy.
- **Performance:** To achieve better performance, we need to optimize the space complexity and feature-dimensionality of the algorithm.
- **Procedure:** Assessment of the practicality of image classification techniques in handling real-time queries in large and diverse image collections is still divided in the absence of clear evidence.

- **Retrieving:** The problem involves uploading an image as a query into software meant to separate visual characteristics using image classification techniques. This method is used to find images in the image database that superficially resemble the query image.
- **Index structure:** A good index structure is necessary for retrieval outcomes to be competent and versatile. The index structure needs to take into account how important it is for image characteristics to change continuously in response to precise queries and user input.

### 5.3. Recommendations

In this research, we explored our model on a 2D image dataset. This study could be expanded to include a 3D image classification system. An object in the 3D image classification system may contain multiple images from various perspectives. This approach can be used for a variety of other image datasets. In the real world, an image cannot be simply identified with a single feature; hence, we proposed a novel methodology using hybrid learning of local and deep features. This proposed technique can be used in a variety of other applications, such as medical imaging, security surveillance, 3D image classification systems, and so on.

### 5.4. Errors and limitations

These limitations reduce the image classification rate. The following are some of the difficulties with image classification:

- **Occlusion** – When an object that needs to be identified is obscured by another object.
- **Variable number of objects** – A image might have several items in it. The identification of every object is then required. Additionally, the system should offer the identities of all instances of an item that has to be identified in an image if there are several of them.
- **Illumination** – Some lighting on an object may have an impact on it. In this scenario, some areas of the item can have more lighting than other areas. After

recognizing the lighting effects, the image classification produces the correct results.

- **Viewpoint variation** – Illumination of an object may have an effect on it. In this case, different parts of the object can have different illumination levels. The image classification system gives accurate results after recognizing the lighting effects.
- **Intra-class variation** – An image may seem differently, for instance, a chair may have a variety of designs. All these implications should be addressed by the image classification system.
- **Inter-class similarity** – An image that we are classifying could be recognized as belonging to a different category, such as a cat that could be mistaken for a tiger.

## REFERENCES

---

- Abdellatef E, Omran EM, Soliman RF *et al.* (2020) Fusion of deep-learned and hand-crafted features for cancelable recognition systems. *Soft Computing*. DOI: 10.1007/s00500-020-04856-1.
- Affonso C, Rossi ALD, Vieira FHA and de Leon Ferreira ACP (2017) Deep learning for biological image classification. *Expert Systems with Applications*, 85:114-122.
- Agarwal A, Samaiya D and Gupta KK (2017) A Comparative Study of SIFT and SURF Algorithms under Different Object and Background Conditions. *Proceedings of the International Conference on Information Technology (ICIT)*, Bhubaneswar, 42-45. doi: 10.1109/ICIT.2017.48.
- Ahmed K T and Iqbal M A (2018) Region and texture based effective image extraction. *Cluster Computing*, 21(1):493-502
- Arnou T and Bovik AC (2008) Foveated Object Recognition Using Corners. *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation*, 53-56.
- Bansal M, Kumar M, and Kumar M (2021) An efficient technique for object recognition using Shi-Tomasi corner detection algorithm. *Soft Computing*, 25:4423–4432.
- Bansal, M., Kumar, M., Sachdeva, M. et al. Transfer learning for image classification using VGG19: Caltech-101 image data set. *J Ambient Intell Human Comput* (2021). <https://doi.org/10.1007/s12652-021-03488-z>
- Bay H, Ess A, Tuytelaars T and Van Gool L (2008) Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346-359.
- Boyras P and Bayraktar E (2017) Analysis of feature detector and descriptor combinations with a localization experiment for various performance metrics. *Turkish Journal of Electrical Engineering and Computer Sciences*, 25(3):2444-2454.
- Bosch A, Zisserman A and Munoz X (2007) Image Classification using Random Forests and Ferns. *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV)*, 1-8.
- Breiman L (2001) Random Forests, *Machine Learning*, 45(1):5-32.

- Brodersen KH, Ong CS, Stephan KE and Buhmann JM (2010) The Balanced Accuracy and its Posterior Distribution. *Proceedings of the 20th International Conference on Pattern Recognition*, 3121-3124.
- Chao W L, Changpinyo S, Gong B and Sha F (2016) An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. *Proceedings of the European Conference on Computer Vision*, 52-68.
- Chao Y, Zhennan W, Xuepu JIANG, Hongyan YU, Sai LIU and Liangzhu D (2018) Fast Object Classification Method Based on Saliency Detection. *Proceedings of the 11th International Symposium on Computational Intelligence and Design (ISCID)*, 1:374-377.
- Chen T, Yin X, Yang J, Cong G, and Li G (2021) Modeling Multi-Dimensional Public Opinion Process Based on Complex Network Dynamics Model in the Context of Derived Topics. *Axioms* 10 (4), 270. doi:10.3390/axioms10040270
- Chien H J, Chuang CC, Chen CY and Klette R (2016) When to use what feature? SIFT, SURF, ORB, or A-KAZE features for monocular visual odometry. *Proceedings of the International Conference on Image and Vision Computing*, New Zealand (IVCNZ), 1-6.
- Chum O and Zisserman A (2007) An Exemplar Model for Learning Object Classes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.
- Diplaros A, Gevers T and Patras I (2016) Combining color and shape information for illumination-viewpoint invariant object recognition. *IEEE Transactions on Image Processing*, 15(1):1-11.
- Fei-Fei L, Fergus R and Perona P (2004). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop*, 178-178
- Fergus R, Perona P and Zisserman A (2003) Object class recognition by unsupervised scale-invariant learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.



- Freund Y, Schapire RE (1999) A Short Introduction to Boosting, *Journal of Japanese Society of Artificial Intelligence*, 14(5):771-780
- Garcia-Gasulla D, Parés F, Vilalta A, Moreno J, Ayguadé E, Labarta J, Cortés U and Suzumura T (2017) On the behavior of convolutional nets for feature extraction. *Journal of Artificial Intelligence Research*, 61:563-592
- Godbole S and Sarawagi S (2004) Discriminative methods for multi-labeled classification. *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 22-30.
- Guo GD and Zhang HJ (2001) Boosting for fast face recognition. *Proceedings of the IEEE ICCV Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, 96-100.
- Gupta S, Kumar M and Garg A (2019) Improved object recognition results using SIFT and ORB feature detector. *Multimedia Tools and Applications*, 78(23):34157-34171.
- Haralick R M, Shanmugam K and Dinstein I H (1973) Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 6:610-621.
- He X and Niyogi P (2004) Locality Preserving Projections. *In Advances in Neural Information Processing Systems*, 153-160.
- Helmer S and Lowe DG (2004) Object Class Recognition with Many Local Features. *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 187-187.
- Huang X, Xu Y and Yang L (2017) Local visual similarity descriptor for describing local region. *Proceedings of the Ninth International Conference on Machine Vision (ICMV 2016)*, 10341: 103410S
- Ijjina EP and Mohan CK (2014) View and illumination invariant object classification based on 3D Color Histogram using Convolutional Neural Networks. *Proceedings of the Asian Conference on Computer Vision*, 316-327.
- Jurie F and Triggs B (2005) Creating efficient codebooks for visual recognition. *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV)*, 1: 604-610.

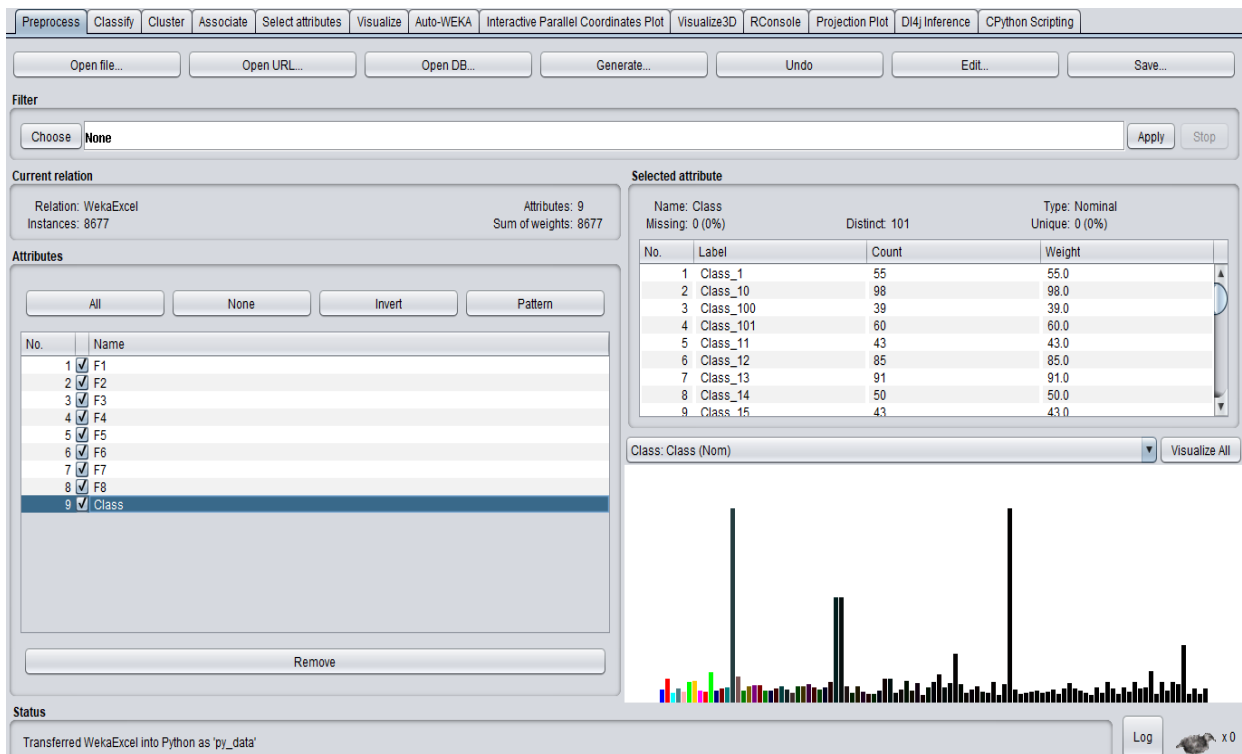
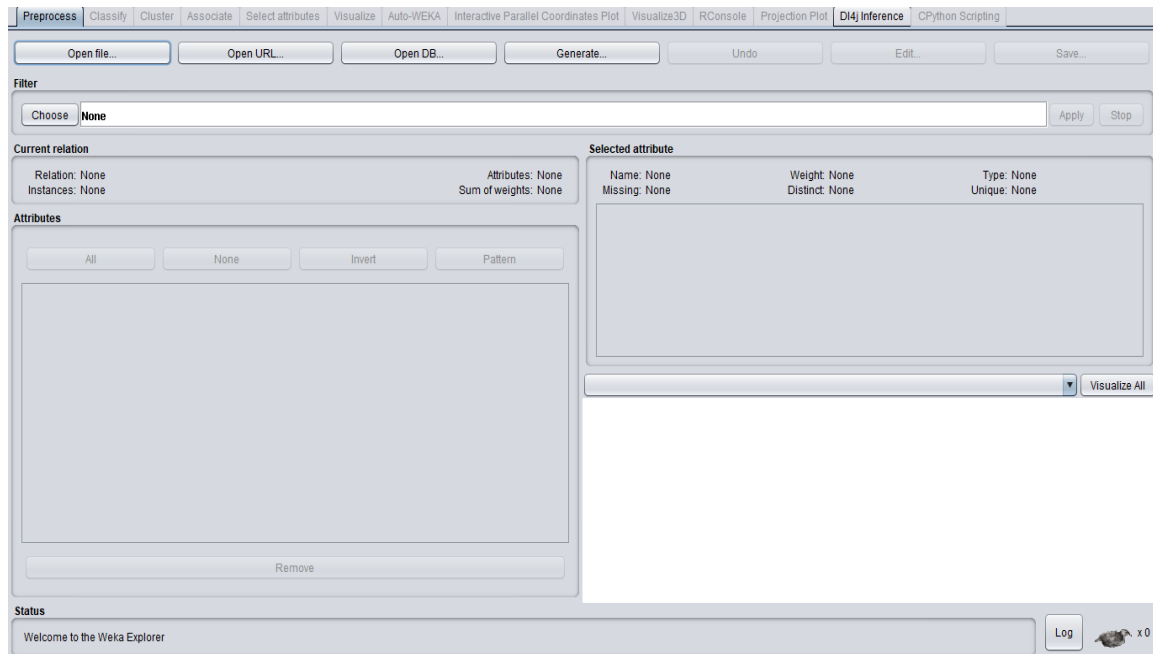
- Kabbai L, Abdellaoui M and Douik A (2019) *Image classification by combining local and global features. The Visual Computer*, 35:679-693.
- Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R and Wu AY (2002) An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 24(7):881–892.
- Khan MA, Sharif M, Akram T, Raza M, Saba T and Rehman A (2020) Hand-crafted and deep convolutional neural network features fusion and selection strategy: an application to intelligent human action recognition. *Applied Soft Computing*, 87:105986
- Kim J, Kim BS and Savarese S (2012) Comparing image classification methods: K-nearest-neighbor and support-vector-machines. *Proceedings of the Applied Mathematics in Electrical and Computer Engineering*, 133-138.
- Li W, Dong P, Xiao B and Zhou L (2016) Object recognition based on the region of interest and optimal bag of words model. *Neurocomputing*, 172:271-280.
- Li F and Xiong Y (2018) Automatic identification of butterfly species based on HoMSC and GLCMolB. *The Visual Computer*. 34:1525-1533.
- Liu Y, Yu M, Xue C and Yang Y (2018) A Novel Image Classification Method Based on Bag-of-Words Framework. *Proceedings of the IEEE 8th Annual International Conference on CYBER Technology in Automation, Control and Intelligent Systems (CYBER)*, 534-539.
- Luo L (2021) Research on Image Classification Algorithm Based on Convolutional Neural Network, *Journal of Physics: Conference Series*, Vol. 2083, 2.
- Mahantesh K, Aradhya VNM and Niranjana SK (2015) Coslets: A Novel Approach to Explore Object Taxonomy in Compressed DCT Domain for Large Image Datasets. *Advances in Intelligent Systems and Computing*, 320.
- Mahmood A, Bennamoun M, An S and Sohel F (2017) Resfeats: Residual network based features for image classification. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 1-9. doi:10.1109/icip.2017.8296551
- Montabone S and Soto A (2010) Human detection using a mobile platform and novel features derived from a visual saliency mechanism. *Image and Vision Computing*, 28(3):391–402.

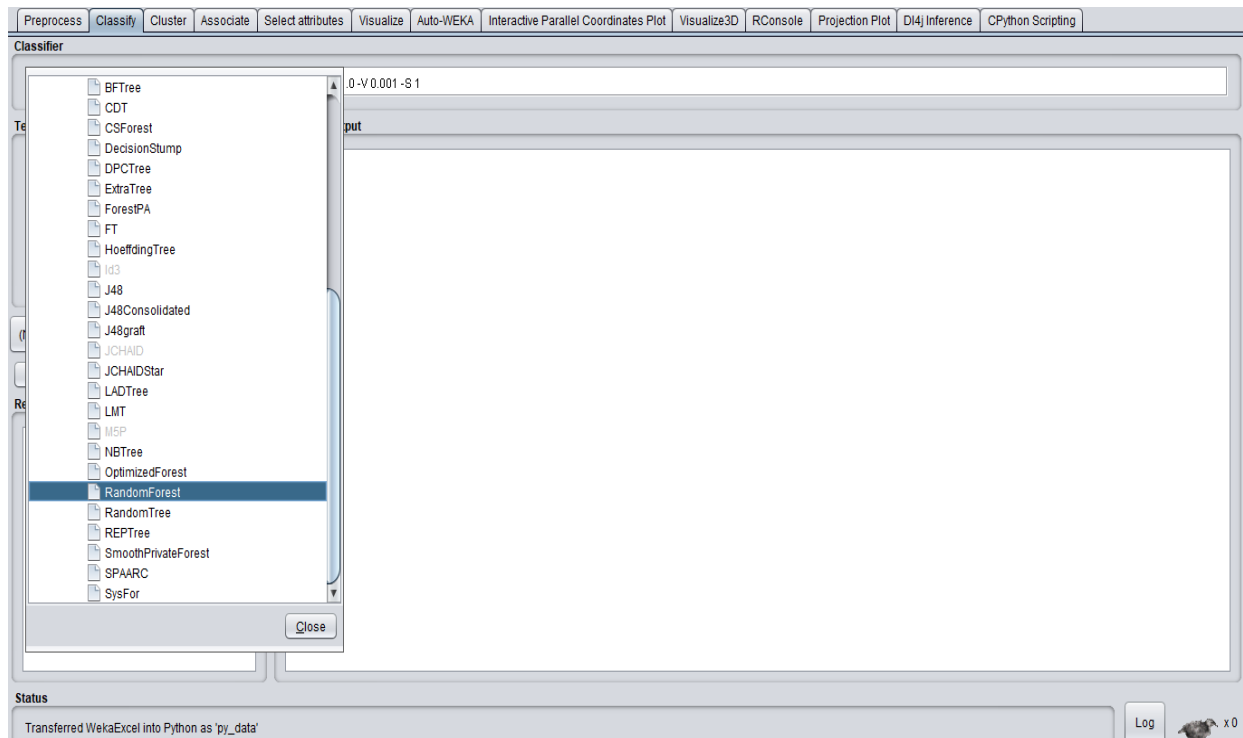
- Muralidharan R (2014) Object recognition using k-Nearest Neighbor supported by Eigen value generated from the features of an image. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(8):5521-5528.
- Murphy-Chutorian E and Triesch J (2005). Shared features for scalable appearance-based object recognition. *Proceedings of the 7th IEEE Workshops on Applications of Computer Vision*, 1(1):16-21.
- Prajapati N, Nandanwar AK and Prajapati GS (2016) Edge Histogram Descriptor, Geometric Moment and Sobel Edge Detector Combined Features Based Object Recognition and Retrieval System. *International Journal of Computer Science and Information Technologies (IJCSIT)*, 7(1):407-412.
- Quinlan JR (1986) Induction of Decision Trees. *Machine learning*, 1(1):81-106.
- Rashid M, Khan MA, Sharif M, Raza M, Sarfraz MM and Afza F (2018) Object detection and classification: a joint selection and fusion strategy of deep convolutional neural network and SIFT point features. *Multimedia Tools and Applications*, 78:15751–15777.
- Rastegari M, Ordonez V, Redmon J and Farhadi A (2016) Xnor-net: Imagenet classification using binary convolutional neural networks. *Proceedings of the European Conference on Computer Vision*, 525-542.
- Roy K and Mukherjee J (2013) Image Similarity Measure using Color Histogram, Color Coherence Vector and Sobel Method. *International Journal of Science and Research (IJSR)*, 2(1):538-543.
- Shang J, Chen C, Pei X, Liang H, Tang H and Sarem M (2017) A novel local derivative quantized binary pattern for object recognition. *The Visual Computer*, 33(2):221–233.
- Shermina J (2010) Application of locality preserving projections in face recognition. *International Journal of Advanced Computer Science and Applications*, 1(3):82-85.
- Shi J and Tomasi C (1994) Good Features to Track. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 593–600.

- Shivakanth AM (2014) Object Recognition using SIFT. *International Journal of Innovative Science, Engineering & Technology (IJSET)*, 1(4):378-381.
- Shu X, Tang J, Qi G-J, Li Z, Jiang Y-G and Yan S (2018) Image Classification with Tailored Fine-Grained Dictionaries. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(2):454–467.
- Simonyan K and Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
- Sivic J, Russell BC, Efros AA, Zisserman A and Freeman WT (2005) Discovering objects and their location in images. *Proceedings of the Tenth IEEE International Conference on Computer Vision*, 1:370-377.
- Srivastava D, Bakthula R and Agarwal S (2019) Image classification using SURF and bag of LBP features constructed by clustering with fixed centers. *Multimedia Tools and Applications*, 78(11):14129-14153.
- Tareen SAK and Saleem Z (2018) A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB and BRISK. *Proceedings of the International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 1-10.
- Tesfaye AL and Pelillo M (2020) Multi-feature fusion for image retrieval using constrained dominant sets. *Image and Vision Computing*, 94: 103862
- Uijlings JRR, Van de Sande KEA and Gevers T (2013) Selective Search for Object Recognition. *International Journal of Computing Vision (IJCA)*, 1-14.
- Vo T, Nguyen T and Le CT (2019) A hybrid framework for smile detection in class imbalance scenarios. *Neural Computing and Applications*, 1-10.
- Wei H, Chengzhuan Y and Yu Q (2017) Contour Segment Grouping for Object Detection. *Journal of Visual Communication and Image Representation*, 48:292-309.
- Wu H and Zhou Z (2021) Using Convolution Neural Network for Defective Image Classification of Industrial Components, *Mobile Information Systems*, Vol. 2021, 9092589, DOI:10.1155/2021/9092589
- Wu M, Ramakrishnan N, Lam SK and Srikanthan T (2012) Low-complexity pruning for accelerating corner detection. *Proceedings of the IEEE International Symposium on Circuits and Systems*, 1684-1687.

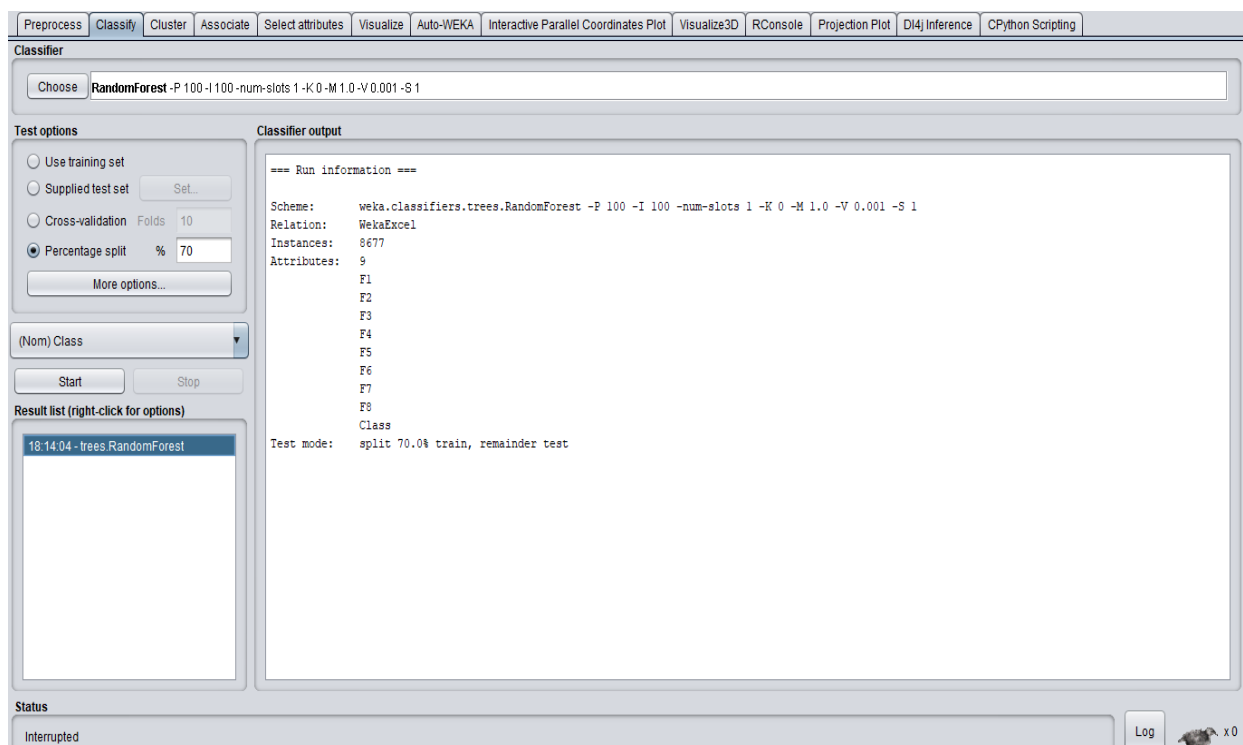
- Xiao T, Liu L, Li K, Qin W, Yu S and Li Z (2018) Comparison of Transferred Deep Neural Networks in Ultrasonic Breast Masses Discrimination. *Biomed Research International*. <https://doi.org/10.1155/2018/4605191>.
- Xie GS, Zhang XY, Shu X, Yan S and Liu CL (2015) Task-driven feature pooling for image classification. *Proceedings of the IEEE International Conference on Computer Vision*, 1179-1187.
- Xie GS, Zhang XY, Yan S and Liu CL (2017) SDE: A novel selective, discriminative and equalizing feature representation for visual recognition. *International Journal of Computer Vision*, 124(2):145-168.
- Zhu Q, Zhong Y and Liu Y, Zhang L and Li D (2018) A Deep-Local-Global Feature Fusion Framework for High Spatial Resolution Imagery Scene Classification. *Remote Sensing*. 10(4):1-22. DOI:10.3390/rs10040568.

# Appendix





**Classification model selection (For example, here we selected Random Forest)**



**Training of the proposed model with Random Forest Classifier**

Preprocess Classify Cluster Associate Select attributes Visualize Auto-WEKA Interactive Parallel Coordinates Plot Visualize3D RConsole Projection Plot D[4] Inference CPython Scripting

**Classifier**

Choose RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

**Test options**

☐ Use training set

☐ Supplied test set

☐ Cross-validation Folds 10

☒ Percentage split % 70

(Nom) Class

**Result list (right-click for options)**

18:14:04 - trees RandomForest

18:14:57 - trees RandomForest

**Classifier output**

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_1
0.192	0.010	0.156	0.192	0.172	0.164	0.957	0.127	Class_10
1.000	0.000	0.900	1.000	0.947	0.949	1.000	1.000	Class_100
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_101
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_11
0.375	0.011	0.243	0.375	0.295	0.294	0.950	0.217	Class_12
0.107	0.007	0.150	0.107	0.125	0.119	0.926	0.132	Class_13
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_14
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_15
0.209	0.014	0.205	0.209	0.207	0.193	0.966	0.226	Class_16
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_17
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_18
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_19
0.534	0.070	0.444	0.534	0.485	0.428	0.933	0.479	Class_2
0.150	0.006	0.286	0.150	0.197	0.198	0.939	0.180	Class_20
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_21
0.217	0.006	0.238	0.217	0.227	0.221	0.971	0.203	Class_22
0.304	0.012	0.189	0.304	0.233	0.231	0.958	0.320	Class_23
0.500	0.013	0.209	0.500	0.295	0.317	0.984	0.309	Class_24
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_25
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_26
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_27
0.167	0.010	0.100	0.167	0.125	0.121	0.967	0.144	Class_28
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_29
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Class_3
0.450	0.004	0.450	0.450	0.450	0.446	0.990	0.333	Class_30

Status

OK  x 0

## Experimental Results with Random Forest Classifier