



ARTICLE

EffNet-CNN: A Semantic Model for Image Mining & Content-Based Image Retrieval

Rajendran Thanikachalam¹, Anandhavalli Muniasamy², Ashwag Alasmari³ and Rajendran Thavasimuthu^{4,*}

¹Department of Computer Science and Engineering (Cyber Security), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, 600062, India

²Department of Informatics and Computer Systems, College of Computer Science, King Khalid University, Abha, 61421, Saudi Arabia

³Department of Computer Science, King Khalid University, Abha, 61421, Saudi Arabia

⁴Department of Sustainable Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, 602105, India

*Corresponding Author: Rajendran Thavasimuthu. Email: rajendran.thavasimuthusamy@ieee.org

Received: 03 January 2025; Accepted: 09 April 2025; Published: 30 May 2025

ABSTRACT: Content-Based Image Retrieval (CBIR) and image mining are becoming more important study fields in computer vision due to their wide range of applications in healthcare, security, and various domains. The image retrieval system mainly relies on the efficiency and accuracy of the classification models. This research addresses the challenge of enhancing the image retrieval system by developing a novel approach, EfficientNet-Convolutional Neural Network (EffNet-CNN). The key objective of this research is to evaluate the proposed EffNet-CNN model's performance in image classification, image mining, and CBIR. The novelty of the proposed EffNet-CNN model includes the integration of different techniques and modifications. The model includes the Mahalanobis distance metric for feature matching, which enhances the similarity measurements. The model extends EfficientNet architecture by incorporating additional convolutional layers, batch normalization, dropout, and pooling layers for improved hierarchical feature extraction. A systematic hyperparameter optimization using SGD, performance evaluation with three datasets, and data normalization for improving feature representations. The EffNet-CNN is assessed utilizing precision, accuracy, F-measure, and recall metrics across MS-COCO, CIFAR-10 and 100 datasets. The model achieved accuracy values ranging from 90.60% to 95.90% for the MS-COCO dataset, 96.8% to 98.3% for the CIFAR-10 dataset and 92.9% to 98.6% for the CIFAR-100 dataset. A validation of the EffNet-CNN model's results with other models reveals the proposed model's superior performance. The results highlight the potential of the EffNet-CNN model proposed for image classification and its usefulness in image mining and CBIR.

KEYWORDS: Image mining; CBIR; semantic features; EffNet-CNN; image retrieval

1 Introduction

In recent years, an exponential increase in the volume of visual data has been witnessed due to the distribution of digital images across various platforms. This surge in data has motivated significant interest in fields like image mining and CBIR. Image mining involves the extraction of valuable information from large collections of images. On the other hand, CBIR concentrates on image retrieval according to their visual contents, instead of textual metadata. These fields play a significant role in several applications, from healthcare to multimedia. The effectiveness of image retrieval systems is based on the robustness and



efficiency of the classification models. Conventional methods frequently struggle with the large scale and complexity of contemporary image datasets, necessitating the investigation of more advanced techniques [1]. Image mining is a subfield of data mining that extends beyond the typical field of structured data to include unstructured data types like image data. Image-mining strategies have become necessary due to the proliferation of images and image databases. Image mining is a collection of analytical approaches to investigate a substantial volume of images [2].

Fig. 1 depicts the general process image mining method. The procedure begins with selecting an image database, where the data will be saved. To increase the overall database quality, it is frequently necessary to subject inconsistent data through a preprocessing phase. Image processing techniques, which are numerical operations that affect the image's pixel values, are applied during this step. Some examples of image processing techniques include filtering, image subtraction, histogram equalization, image restorations, etc. During the feature extraction and transformation process, the images cope with a few transformations until the pertinent items in these data are discovered. After that, characteristics of these items, such as length, edge shape, and texture, are extracted from them. The mining process can then utilize data mining approaches to uncover important patterns semiautomatically or automatically if such features have been obtained first. An expert must analyze and interpret these patterns to gain knowledge that can be applied to applications and is helpful in problem understanding or decision-making [3].

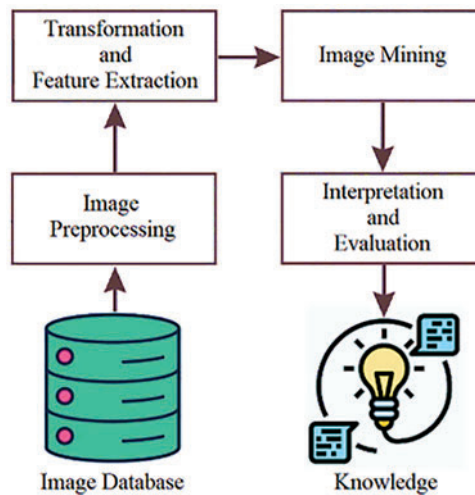


Figure 1: Pipeline of image mining process

Image mining widens the potential applications of data mining, making it a potential technology for data mining. Patterns of data that are unstructured can be distinguished from those that are structured. It cannot be easy to correctly understand and comprehend patterns extracted from image databases [4]. Image mining is therefore considered more than merely an extension of data mining. It is an attempt that incorporate knowledge from many other fields, including artificial intelligence, data mining, image processing, machine learning, computer vision, databases, and more [5]. Image processing and computer vision are not interchangeable with image mining. On the other hand, image mining aims to identify patterns stored in large image databases, whereas these other fields concentrate on extracting individual attributes from a single image. This encompasses every facet of databases, including the indexing method, storing images, and retrieving images [6]. The process of mining images consists of many components, including the following:

- Image analysis is essential to image mining and cannot be skipped. Image analysis aims to discover and extract every pertinent feature necessary to depict the images.
- Image classification: The goal is to group the various things seen in an image.
- Indexing and retrieving images require appropriate indexing to obtain timely images from databases.
- Management of data: Images contain an extremely large quantity of information. An image database allows for storing and indexing images, which determines the types of information that can be searched for and retrieved from the database [7].

1.1 Problem Statement

The challenge is that there is not currently a technique for image mining that is both efficient and accurate, particularly in the domains of image classification and CBIR. Existing methods frequently encounter difficulties extracting relevant features from huge image datasets, resulting in less-than-optimal performance and restricted applicability. As a result, there is a requirement for developing an integrated model that combines the effectiveness of the EffNet architecture with the capability of CNNs to learn new features. The MS-COCO and CIFAR datasets serve as the primary focus of the research, and the preprocessing techniques utilized include data augmentation and channel-wise normalization. For feature extraction and classification, the EffNet-CNN hybrid architecture has been implemented; assessment metrics include accuracy and effectiveness for retrieval. This research aims to propose a viable method to enhance image mining tasks.

1.2 Research Objectives

This research proposes a deep-learning image-mining model involving CNNs and the EffNet model. The following is the list of objectives for the proposed research.

- To analyze and provide a detailed review of recent current models in the field of CBIR and image mining.
- To design and develop a novel approach called “EffNet-CNN,” aimed at addressing the challenge of improving retrieval of images through CBIR and image mining. The originality of the developed model resides in its integration of EfficientNet-CNN with channel-wise normalization, random data augmentation, semantic feature extraction, and classification techniques, improving superior performance in image mining and content-based retrieval tasks.
- To apply the Mahalanobis distance metric for feature matching and highlight the significance of Mahalanobis over Manhattan distance with comparison.
- To experiment and evaluate the EffNet-CNN model on multiple evaluation metrics such as precision, accuracy, f1-score, and recall with MS-COCO, CIFAR-10 and 100 datasets.
- To validate the performance of the EffNet-CNN model by comparing it with current image mining methodologies like K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Vector Space Model (VSM), Long Short-Term Memory (LSTM), and Image Mining based Deep Belief Neural Network (IMDBN) regarding the accuracy and other metrics.

The research contribution includes EffNet-CNN, a novel image retrieval model that enhances EfficientNet with channel-wise normalization, random data augmentation, and semantic feature extraction, significantly improving classification accuracy and robustness. The proposed model outperforms existing approaches, demonstrating superior performance in image classification and CBIR. The work provides a hyperparameter tuning process, ensuring reproducibility and practical applicability in image retrieval tasks.

The remaining sections of this research are organized as follows: The works related to image mining and CBIR are discussed in the following. [Section 3](#) implements the proposed methodology, including data

augmentation, normalization, interpolation, hybrid EffNet-CNN, and CBIR. The experimental results and the interpretation of the findings are covered in [Section 4](#). [Section 5](#) presents a conclusion and a discussion of potential future research directions.

2 Related Works

2.1 ML-Based Image Mining

The challenges of semantic image retrieval and image mining play a significant part in the operation of various applications. To achieve greater precision, a system called Graph-CTree was proposed in [8], which combined a C-Tree with a neighbour graph. The classified images found on Graph-CTree were comparable using a k-Nearest Neighbour (kNN) to generate a list of visual terms. A framework for image ontology has been built using a semi-automatic process. The Simple Protocol and Resource Description Framework (RDF) Query Language (SPARQL) query was automatically produced from the visual terms and then retrieved from the ontology for the semantic image. To improve the effectiveness of this model, both the Conditional inference Tree (C-Tree) and the Graph-CTree have had their query methods and structures updated and improved. Image mining is part of the initial diagnostic process in various medical fields. The work in [9] presented the development of an algorithm that used image mining to detect micro-calcification automatically in digital mammograms. The grey-level co-occurrence matrix (GLCM) was utilized to extract pertinent features from images, whereas radial basis function-support vector machine (RBF-SVM) was applied for feature classification.

Image mining was applied in [10] to the obtained image after objects in the digital image were segmented to accomplish disease detection and severity analysis utilizing the Computer-Aided Diagnosis (CAD) system. This was done after the objects in the digital image were extracted. The cross-central filter, which combined aspects of the Gaussian and Mean filter operations, was implemented in the noise reduction process to remove unwanted noise. The segmentation process for the object segments was carried out using the Cognitive-FCM (Fuzzy C-Means) algorithm. Utilizing an SVM classifier based on Quad Programming allowed for the evolution of the classification process. The research described in [11] integrated optimization strategies to accurately forecast optimal centroid in FCM to improve the retrieving performance of the CBIR system. The FCM centroids did help to simplify things and minimized the amount of time spent computing. It was decided to use the swarm intelligence approach to tackle the problem of predicting ideal FCM centres of gravity and to get an understanding of the fundamental methodology involved in implementing particle swarm (PSO) and crow search optimization (CSO), which developed the concept of oppositional crow search optimization (OCSO).

CBIR was attained through the normal/abnormal classification in [12]. The artificial neural network approach and the GLCM feature extraction were utilized to complete the learning procedure for the CBIR successfully. The artificial bee colony algorithm was utilized to manage the normal/abnormal classification of the medical dataset. The solutions were repeated and combined with the method that was accessible to control the functionality of this method.

The work fusion of colour and shape features was used in [13] to extract descriptors from images using the colour moment and edge histogram descriptor. To obtain clusters from the dataset, the K-medoid clustering methodology was utilized. Following the application of clustering, optimum clusters were generated, in which images with comparable characteristics belonged to the same cluster or group. A fuzzy cognitive map approach based on neural networks (NN) was developed in [14]. A fuzzy cognitive map was used to extract the pertinent data features from the seed test, which then initialized the NN. The Levenberg–Marquardt approach for NN was developed to make seed purity predictions more accurate. To increase the

accuracy of the network's predictions, a deep learning technique was used to initialize the weights of the hidden layers based on a fuzzy cognitive map. Levenberg–Marquardt was the algorithm that was used to optimize the weights of the output layer. The study [15] presented a CBIR-hybrid features extrication and Hierarchical Weights-Brownian Motions Monarch Butterfly Optimizer-based features selection method for forest fire recovery. A forest fire retrieval system used the feature extraction and selection methods. The results demonstrated that the fire retrieval method was more accurate.

2.2 Image Mining-Based Methods

Using a combination of several different image attributes, the authors of [16] devised an image mining method that is accurate and efficient in image retrieval. On a sub-block, a calculation was made to determine the colour feature in the form of first, second, and third-order moments. These points in time calculated various statistical characteristics over a regional area. The center symmetric local binary pattern (CSLBP) feature was utilized during the texture extraction. An extracted texture with information about the local edges was called the Laplacian of Gaussian-CSLBP. The target test image was used to retrieve images of other images that are most comparable to it. Because of this weight component, the texture feature becomes more powerful, boosting its capacity for discrimination. The work in [17] presents a method for categorizing and recognizing flowers based on an algorithm for mining genetic association rules. This approach utilized textural characteristics, such as a grey-level cooccurrence matrix, as its foundation. Multidimensional data mining using this approach has shown to be rather successful. Clustering, one of the image mining techniques, is applied to manage the database because it contains many images spread across it.

The multimodal knowledge graphs (MKG) to clearly describe indirect multimodal knowledge linkages and integrate it into visuals-semantics embedding for images-texts retrieval was proposed using Multimodal Knowledge enhanced Visual-Semantic Embedding (MKVSE). For implicit multimodal knowledge links between text and image, MKG was created. For visual-semantic embedding, Multimodal Graph Convolution Networks (MGCN) reasoned on MKG in two stages. This strategy has been extensively tested on two benchmarks and shown to be effective [18]. A CBIR approach that extracted effective features was developed [19]. The presented feature vector combined low-and mid-level image features. Gabor wavelets transform, Auto-correlogram, and multi-level fractal dimension analysis extracted image colour, shape, and texture. Mid-level image features were retrieved using the Deep Boltzmann Machines and by comprehending minimum-level image characteristics and their correlations. In Caltech-256, Corel 5K and 10K datasets, the model fared better.

2.3 DL-Based Image Mining

Deep convolutional neural networks were used in a system to classify images proposed in [20]. CNN is reliable and could bring about satisfying outcomes. The CIFAR-10 dataset was used in groups of various patterns, colours, and shapes to explore the algorithm's efficiency in categorizing any image linked to any of those groups inside the database and to present the essential layers in the deep CNN. This was done to determine how well the system could categorize images. The CNN training technique was executed using stochastic gradient descent (SGD), and random data containing equal and small batches of each iterative learning phase were chosen randomly. The classification result was successful, and to further increase the effectiveness of the deep CNN, it is essential to estimate the number of hidden layers and the chronological sequence in which they should be applied. Effective searching and retrieval were accomplished by a CBIR system that was created in [21]. This system made use of an image mining-based deep belief neural network model. To extract feature sets that were pertinent to the problem, the colour-shape texture feature extraction technique was utilized. The retrieved characteristics were combined, then saved in the feature vector, and

then this classification step was performed on them to recover related images under a single label. This approach improved retrieval accuracy and helped the semantic gap between human visual comprehension and the representation of image features.

A deep collaboration graph hashing (DCGH) model was proposed in [22], which used multi-level semantic embeddings, intrinsic structure mining, and latent common space creation in discriminative hash code learning for large-scaled image retrieval. A dual-stream feature encoding network was utilized to simultaneously examine multi-level semantic data spanning visuals and semantics characteristics instead of a single flow visual network. A graph convolutional network preserved latent structural relationships in optimum pairwise similarity-preserved hash code. Numerous dataset studies show that this DCGH could retrieve images effectively. The study [23] presented an image retrieval system that utilized computed descriptors obtained from three distinct learning feature extraction approaches: InfoGANs, convolutional autoencoders, and vision transformers (ViT). Based on the results, the InfoGAN and ViT retrieval systems demonstrated higher efficacy in image retrieval.

A hybrid machine learning (ML) and deep learning (DL) CBIR framework combining two pre-trained DL methods, VGG16, ResNet50 and an ML method, KNN, was implemented in [24]. Transfer learning was used to extract image characteristics using these two DL models. Euclidean distance and KNN calculate image similarity. A web interface displayed comparable images, and the model with the best outcomes was measured by precision. A novel CBIR solution was provided in [25], by developing a CNN and Relevance Feedback (RF) architecture. The Generalized Discriminant Analysis (GDA) was applied to extracted features to improve image features and minimize feature vector dimensionality. The results demonstrated that this method performed well. DarkNet-53 for CBIR was enhanced by developing Shuffled-Xception-DarkNet-53 in [26]. The model introduced the Shuffled-Xception module, which employed three sets of filters in serial connection instead of a single 3×3 filter to extract finer characteristics from input images. 'Group Convolution' enhanced the Shuffled-Xception module's filter co-relations, making features more informative. To maintain information flow across groups' channels, one Channel Shuffle layer was utilized between every two serial Group Convolution layers of the same size. The findings demonstrated higher efficacy in image retrieval.

A deep open-source methodology to mine and reason the uncertainty of data sets was developed in [27] to improve performance and minimize misclassification. A label reassignment method was designed to rearrange the training labels of the network and accept defective samples with various labels. These were eliminated from their original classes and considered imprecise samples of partial elimination. Then, a new unbalanced data augmentation approach was designed for generalizing all the classes. It enabled the network to fuse imprecise and precise auxiliary data to extract more unique class features from single-labelled samples and assessed uncertainty-induced imprecision in the test set using imprecise test samples.

A CNN-based CBIR model for medical image retrieval was proposed in [28], which further utilized Modified Cosine Similarity (MCS) for feature matching. The images were pre-processed with the application of Gaussian smoothing, Contrast Limited Adaptive Histogram Equalization, and Gaussian filter. The deep features of database images were subsequently retrieved utilizing the VGG-19 and Inception-v3. The retrieved features were unified, and the optimal characteristics were chosen using the Coyote-Moth Optimization approach.

2.4 Research Gap

The analysis of related research indicates a wide variety of studies in image mining and the applications of this technique, all of which are compatible with the proposed model. These works investigate semantic-based retrieval, deep learning techniques, optimization, and feature extraction methods. They

include various topics, including medical diagnostics, horticulture, CBIR, and plant identification. This study addresses the semantic gap in CBIR systems and highlights the significance of preprocessing and feature extraction for accurate image representation. In addition, the research highlights the significance of clustering, classification, and similarity measures in image mining. Overall, the literature research gives helpful insights into existing approaches and methodologies that can be used in the proposed model for image mining utilizing EffNet-CNN. [Table 1](#) highlights the advantages and disadvantages of all the analyzed current works discussed above with the proposed research model.

Table 1: Critical analysis of related works

Ref.	Model	Application	Advantages	Disadvantages
[8]	Graph-CTree with kNN	Semantic image retrieval using ontology	Enhanced query processing, and improved retrieval efficiency.	Complexity in query processing, high computational cost.
[9]	GLCM & RBF-SVM	Micro-calcification detection in mammograms	Accurate classification, and effective feature extraction.	Limited to specific medical images, lacks scalability.
[10]	Cognitive-FCM and SVM-based Quad Programming	Disease detection and severity analysis	Improved segmentation, noise reduction, and classification.	Computationally expensive, segmentation errors.
[11]	FCM with OCSO	CBIR with optimized centroid prediction	Reduced computation time, and improved feature clustering.	Sensitive to parameter tuning, may struggle with large datasets.
[12]	GLCM, Artificial Neural Network (ANN), and Artificial Bee Colony (ABC) algorithm	Normal/abnormal classification in medical CBIR	Improved classification accuracy, and robust learning.	High computational overhead, dependency on dataset quality.
[13]	K-Medoid clustering	CBIR with optimal clustering	Efficient feature clustering, and better retrieval accuracy.	Computational complexity, sensitive to initial cluster selection.
[14]	Fuzzy Cognitive Map with NN and Levenberg–Marquardt optimization	Seed purity prediction	Accurate weight optimization, deep learning-enhanced accuracy.	High training time, prone to overfitting.
[15]	Hierarchical weights-brownian motions monarch butterfly optimizer	Forest fire image retrieval	High accuracy in fire detection, and optimized feature selection.	Algorithm complexity and requires fine-tuning.

(Continued)

Table 1 (continued)

Ref.	Model	Application	Advantages	Disadvantages
[16]	CSLBP	Texture-based image retrieval	Enhanced texture discrimination, and robust feature extraction.	Computational overhead, dependency on lighting conditions.
[17]	Genetic association rule mining with GLCM	Flower classification and recognition	Effective multidimensional data mining, and successful clustering.	Limited scalability, requires large datasets.
[18]	MGCN with MKVSE	Image-text retrieval with multimodal knowledge graphs	Effective in linking image and text information.	Requires extensive dataset preparation, and high memory usage.
[19]	Deep Boltzmann Machines	Multi-level feature extraction in CBIR	Robust image representation, and improved retrieval.	High dimensionality, complex feature fusion.
[20]	Deep CNN with stochastic gradient descent	Image classification using CIFAR-10	Reliable performance, effective classification.	Requires extensive training data, and high computational cost.
[21]	Deep belief network	Semantic gap reduction in image retrieval	Improved retrieval accuracy, and feature fusion efficiency.	High training complexity, dataset-dependent performance.
[22]	DCGH with GCN	Large-scale image retrieval	Effective hashing maintains structural relationships.	Computationally expensive, sensitive to dataset variance.
[23]	InfoGANs, Convolutional Autoencoders & ViT	Learning-based image retrieval	High retrieval performance, and efficient feature learning.	Requires large training datasets, and high model complexity.
[24]	VGG16, ResNet50, KNN	CBIR system with transfer learning	Transfer learning benefits, effective feature extraction.	High computational cost, feature fusion complexity.
[25]	CNN + Relevance Feedback with GDA	CBIR with dimensionality reduction	Improved feature representation, and enhanced retrieval accuracy.	Dependent on the training dataset, feature loss risk.
[26]	Shuffled-Xception-DarkNet-53	Deep CNN for CBIR	Improved feature extraction, and enhanced retrieval efficiency.	High computational complexity, and increased model depth.

(Continued)

Table 1 (continued)

Ref.	Model	Application	Advantages	Disadvantages
[27]	Open-source deep learning method with uncertainty mining	Uncertainty-based image classification	Improved generalization, and effective uncertainty handling.	Requires imbalanced data augmentation, and increased training time.
[28]	CNN-based CBIR with MCS and Coyote-Moth Optimization	Medical image retrieval	Enhanced feature selection, and optimized retrieval.	Model complexity, optimization sensitivity.
Proposed EffNet-CNN	EfficientNet-based CNN with Mahalanobis Distance for feature matching	CBIR, image classification, image mining	Improved similarity measurement, enhanced feature extraction, and better generalization.	Computationally intensive, parameter tuning is required.

3 Proposed Methodology

Image mining is necessary to extract useful information and knowledge from a large collection of digital images. Because of the exponential increase in data containing images, developing effective methods for evaluating and comprehending visual content has become necessary. This research presents an image-mining method that integrates the EffNet model with CNNs to achieve better results. The EffNet model is a cutting-edge deep-learning architecture that was developed to perform image categorization in an effective and precise manner. It uses an innovative compound scaling strategy that uniformly grows the network's depth, width, and resolution. This results in exceptional performance despite using fewer computational resources, which is a significant benefit.

3.1 Dataset Description

In computer vision and image classification, benchmark datasets such as CIFAR-10, CIFAR-100, and MSCOCO are utilized. They are extremely useful resources for analyzing and assessing the effectiveness of various image-mining techniques and models. The CIFAR-10 database contains 60,000 colour images organized into 10 distinct classes. The dataset comprises 50,000 to train and 10,000 to test [29]. The CIFAR-100 dataset is divided into 50,000 images to be used for training and 10,000 images to be used for testing, keeping the same ratio as CIFAR-10. The CIFAR-100 consists of 100 classes, then categorized into 20 super-classes. These super-classes are then used in the experiments to analyze the outcomes [30,31]. The MS COCO is an authoritative and important benchmark tool used in the field of object recognition and detection. It contains 117,264 training images and more than 5000 testing images with 80 classes [32]. These classes in the MS COCO dataset were grouped by 11 super-classes.

3.2 Data Preprocessing

In image-mining tasks such as image classification, object detection, and semantic segmentation, the random data augmentation technique based on contrast and brightness enhancement is extremely useful. It makes the model a more robust representation by exposing it to a broader and more varied set of augmented images. This makes it possible for the model to learn. This strategy improves the model's performance and

lowers the challenge of overfitting since it includes variations that match the image conditions seen in the real world. Randomly augmenting data entails creating random parameters within predetermined ranges and then applying those parameters to the source images [33]. Gain and bias will be denoted by g and b . These parameters control the brightness (b) and contrast (g) of the image. According to Eq. (1), the input image is denoted by $a(i)$, and the enhanced version of the image is denoted by $e(i)$.

$$e(i) = ga(i) + b \quad (1)$$

This method, in contrast to previous enhancement strategies, selects the value for g and b from the defined sets randomly. Previous strategies for enhancement rely on fixed hyperparameters and tend to overfit the dataset. The set of g and b values include positive and negative examples so that it does not overfit the data. The values are randomly selected from this set and then included in the data as an enhancement. Brightness and contrast are the only two aspects of an image that are modified by this approach to image enhancement. This improvement method applies to various datasets because g and b are data-independent.

This model employs a channel-wise normalization technique to standardize the pixel values within each image's channel individually. It helps to equalize the colour distribution and lessens the influence of differences in illumination or contrast across the various channels. The model can properly learn and extract features from the image without bias toward any channel since each channel is normalized independently before the learning process begins. This normalization technique ensures that the model may focus on the essential characteristics across all channels, improving the performance of both training and classification. In addition, channel-wise normalization contributes to improved convergence throughout the training phase, which is a significant benefit. It improves the model's overall robustness when subjected to various lighting conditions or colour fluctuations in the dataset [34]. Consider that the input is an RGB image, complete with dimensions (height (h), width (w), and channels). The red, green, and blue channels each have a value corresponding to a pixel in the image. The mean value for each channel is determined independently by taking the average values for each pixel throughout the dataset. As a result of computing the mean values for each channel (R, G, and B), there are three mean values. Calculating the mean value for each channel involves taking the average value for each pixel in the dataset that corresponds to that channel. The following equation was used for calculating the mean:

$$\mu = \frac{1}{(N \times h \times w) \times \sum (pc)} \quad (2)$$

In this case, μ indicates the mean value for a particular channel, pc specifies the pixel value of an image in that channel, and Σ indicates the summation of all the images in the collection. Calculating the standard deviation for each channel independently involves taking the square root of the variance in pixel values throughout the dataset. This results in the standard deviation for each channel. Following the computation of the standard deviation values for each channel, there are three standard deviation values. The following equation was used for calculating the standard deviation:

$$\sigma = \sqrt{\frac{1}{(N \times h \times w) \times \sum (pc - \mu)^2}} \quad (3)$$

In this context, σ denotes the standard deviation value for a particular channel, and μ stands for the mean value that was computed for that channel. After obtaining each channel's mean and standard deviation

values in the preceding steps, the pixel values were normalized within each channel using those values. The equation that describes how each channel should be normalized is as follows:

$$Z = \frac{X - \mu}{\sigma} \quad (4)$$

In this context, X refers to the value of a pixel in the image for a certain channel. The procedure for image normalization is carried out in its entirety throughout all the image's pixels, with individual focus placed on each channel. When these computations are carried out, the pixel values in each image channel effectively normalize. As a result, the pixel values are centered around zero, and their standard deviation equals one. This normalization minimizes bias and scaling difficulties that may arise owing to changes in the colour distribution or intensity across distinct channels, and it does so by eliminating the potential for those issues to occur.

Larger input image sizes, typically 224×224 or higher, are supported by EffNet-B0 because it was developed to cope with such images. Because the images in the CIFAR dataset are 32×32 pixels, they need to be enlarged to be compatible with the input size that EffNet-B0 requires. Additionally, the images in the MSCOCO dataset are 640×480 pixels, they need to be minimized. This work uses the nearest neighbour interpolation method to scale the image. The interpolation method is the most fundamental, and its processing time requirements are the lowest of the other methods. In this method, the interpolated pixel is replaced by the pixel that is physically closest to it. The linear interpolation method, known as nearest neighbour interpolation, is simple. When the image's pixels have a high resolution, it is simple to execute and produces satisfactory results [35]. When performing nearest neighbour interpolation, the interpolation kernel is expressed as follows:

$$u(x) = \begin{cases} 0 & |x| > 0.5 \\ 1 & |x| < 0.5 \end{cases} \quad (5)$$

where, x represents the distance between the point that is being interpolated and the grid point.

3.3 EffNet-CNN Model

A model scaling method called EffNet was designed specifically to speed up CNNs. It makes use of a simple compound coefficient that is extremely efficient. EffNet operates in a manner that is distinct from conventional approaches for scaling the network dimensions, including breadth, depth, and resolutions. It does this by uniformly scaling all dimensions using a predetermined scaling coefficient set [36].

In practice, scaling every dimension can increase model performances; thus, achieving a balance across all network characteristics regarding the available resources is the most effective way to improve overall performance. The effectiveness of model scaling is highly dependent on baseline networks. To accomplish this task, a newer baseline network was developed by employing the AutoML system that maximized precisions and efficacy (FLOPS), to search for a neural network. EffNet, in contrast to MnasNet and MobileNet-v2, utilized mobile inverted bottlenecks convolutional (MBConv) as the primary structural component of the network. Additionally, this network replaces the Rectifier Linear Unit (ReLU) activation function with a new function known as swish [37] rather than the ReLU activation function. EffNet-B0's architecture is depicted in Fig. 2. As seen in Table 2, the general structure of EffNet-B0 comprises one average pooling, sixteen MBConvs, two convolutions, and one classification layer. As can be observed, the most important component of an EffNet-B0 network was an MBConv unit, which is responsible for implementing the squeeze-and-excitation (SE) network attention mechanism. Adding the MBConv module

reduces the time spent on training, increases the performance, and provides the network with random depth. Simultaneously, the Squeeze-and-Excitation Network (SENet) attention scheme makes it possible for an MBConv to concentrate on the characteristics of the channel that contain the important data and suppress the attributes that are not necessary. As a result, the MBConv module enables EffNet-B0 to do feature extraction more effectively [38].



Figure 2: Architecture of EffNet-B0

Table 2: Network parameters of EffNet-B0 model

Stages	Operators	Resolutions	Layers	Channels
1	Conv 3×3	224×224	1	32
2	MBConv1, $k3 \times 3$	112×112	1	16
3	MBConv6, $k3 \times 3$	112×112	2	24
4	MBConv6, $k5 \times 5$	56×56	2	40
5	MBConv6, $k3 \times 3$	28×28	3	80
6	MBConv6, $k5 \times 5$	28×28	3	112
7	MBConv6, $k5 \times 5$	14×14	4	192
8	MBConv6, $k3 \times 3$	7×7	1	320
9	Conv1 $\times$ 1 & Pooling & FC	7×7	1	1280

This research intends to improve extracting semantic features from images by merging the EffNet model with CNNs. This research aims to exploit the efficiency and accuracy of the EffNet model. The model that has been proposed utilizes the capabilities of CNNs in addition to the benefits of the EffNet paradigm. Because CNNs are developed to learn hierarchical representations from the raw inputs automatically, they are particularly well-suited for processing visual data. This approach takes advantage of both CNNs and the EffNet model to collect and extract semantic features that represent the visual content of images [39].

The workflow of the approach that has been proposed includes several important components as depicted in Fig. 3. The images in the dataset are preprocessed first, utilizing methods such as data augmentation and channel-wise normalization to adjust the brightness and contrast of the images. By increasing the robustness and diversity of the dataset, this preprocessing step makes it possible for the model to learn more generic representations. The following step utilizes the EffNet-CNN hybrid architecture for feature extraction and classification. The EffNet model acts as the network's backbone, retrieving high-level characteristics from the images after preprocessing them. These features are then sent into a CNN, which refines the representations and learns more discriminative features that are particular to the image-mining task. After the feature extraction process, various image mining activities can be carried out, such as image classification, object detection, or CBIR. A fully connected (FC) layer and a softmax activation function were added to

the CNN architecture to facilitate multi-class classification. After training, the model can reliably predict the class labels of images.

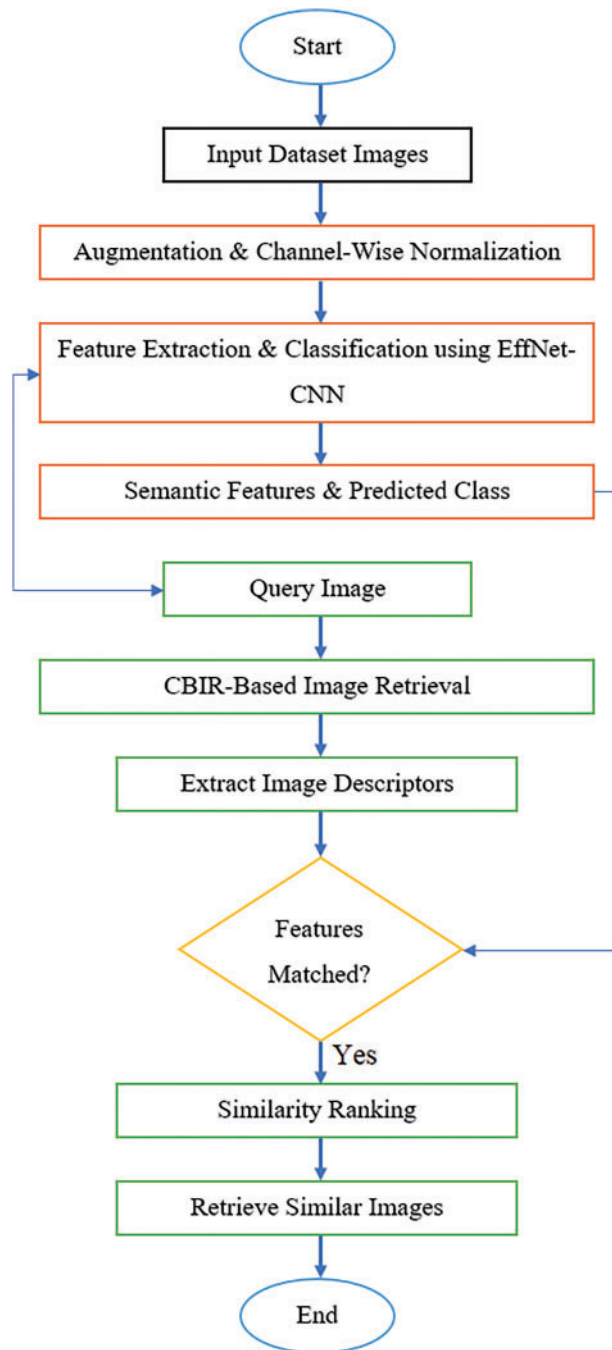


Figure 3: Flowchart of the proposed model

Additionally, the CBIR framework is incorporated into the proposed approach for conducting image mining. Users can then access images from the data set according to the visual similarities of those images to the query image they have provided. The semantic features that were retrieved from the EffNet-CNN model

are the ones that are used in the calculation of similarity estimation among the images in the dataset and query. The similarity rating produced consequently makes it easier to retrieve images with visual content that is nearly identical to the content of query images. The method mentioned above provides an effective and robust answer to problems related to image mining. The capability to effectively extract, represent, and analyze the visual information contained inside massive image collections is made possible by combining the efficiency and accuracy of EffNet with the capability of CNNs to learn new feature representations. The methodology is broken down into parts and explained in detail in the following sections.

3.3.1 Feature Extraction

EfficientNet is a model characterized by a very intricate network but exhibits computational efficiency. Fig. 4 represents the research model's architecture. As the total number of models grows, the EffNet group has expanded to include eight models, B0 and B7. While there has not been a discernible rise in the total number of calculation parameters, there has been a notable rise in the degree of accuracy. Compound scaling is the name given to this approach. According to Eq. (6), the composite coefficient is ϕ employed to provide a uniform scaling of the depth d , width w , and resolution r values. The following is the equation that can be used to estimate the coefficients of composite proportions:

$$\begin{cases} \text{depth: } d = \alpha^\phi \\ \text{width: } w = \beta^\phi \\ \text{resolution: } r = \gamma^\phi \end{cases} \quad \text{s.t., } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (6)$$

Although the coefficients of network depth, width, and resolution could be measured using the variables $\alpha \geq 1, \beta \geq 1, \gamma \geq 1, w, d, r$, respectively, the defined values could be used to calculate the sum of efficient resource extensions model, and the constants, α, β, γ . They could provide the network w, d , and three-dimensional r resources. Considering the investigation's findings, the EffNet-B0 network parameter is presented in Table 1, and the following is a list of the coefficients of optimal networks: $\gamma = 1.15, \beta = 1.1, \alpha = 1.2$. Scaling the convolution network according to the equation in (6) will approximately increase the total FLOPS the network may achieve. This is because the computational overhead of the convolutional network is unique due to the convolution operation.

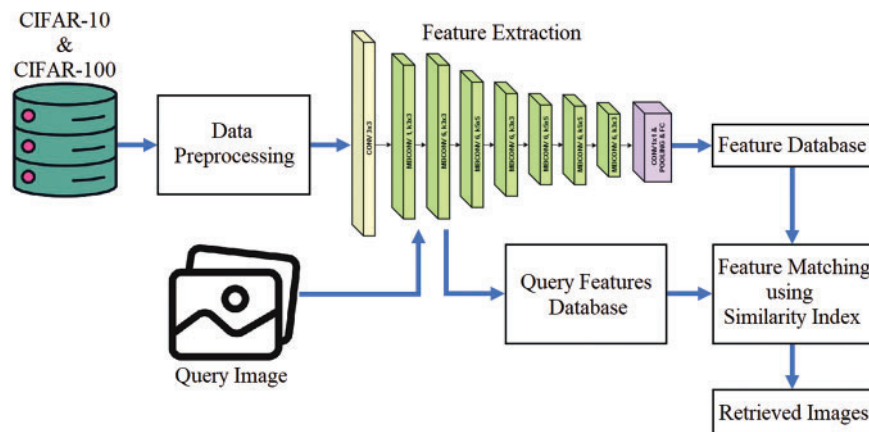


Figure 4: Architecture of the proposed model

The MBConv, an important structural component, is the primary block on which the EffNet framework was constructed. The MBConv was applied with the concepts of MobileNet models, which were then adapted to meet the requirements of MBConv. The usage of depth-wise separable convolutions, which consists of a depth-wise and a point-wise convolution layer, both of which are applied one after the other in sequential order, is one of the most significant principles. The inverted residual connections and the bottleneck linear features found in MobileNet-v2 are known as the inverted residual connection and linear bottleneck combination. Convolution was originally understood to refer to the weighted superpositions of a single function over another regarding their respective physical meanings. The function was created by an integral operator, which took the values of two other functions as input. The overlap length is integral to the values of two overlapping functions, f and g , created through shifting and flipping. It is represented by the notation $r(x)$, which denotes the integral's overlap length of the values of the overlapping functions. This notation denotes the integral's overlap length of the overlapping function.

$$r(x) = \int_{-\infty}^{\infty} f(\tau) g(x - \tau) d\tau \quad (7)$$

The system's output was the product of multiple layered inputs, one on top of the other. In image processing, the function $f(\tau)$ represented the real pixel points, while function $g(x)$ represented the action point. Both functions were given the same input value, x . Each action point was incorporated into a convolutional kernel for extra processing. After each action point on the convolution kernels has been implemented to each of the actual pixel points in the correct order, it is possible to acquire the result of the final convolution. This can only be done after every action point has been applied. The ReLU uses an element-wise operation equation to bring the positive values obtained from the convolution layers down to 0. This is accomplished by applying the equation to the nonnegative data collected from the convolutional layer. The notation 'I' was the element of the input vector.

$$\sigma(I) = \max(0, I) \quad (8)$$

Once CNN's convolution layer has been trained, the pooling layer, also called the down-sampling layer, is tasked with reducing the volume of the matrix it has created by a factor of two. This layer is also known as the down-sampling layer. Using pooling helps reduce the complexity of constructing convolutional networks by keeping the number of parameters and features to an absolute minimum. It is commonly located at the top level of the CNN since it is the final layer of the network and because it summarizes all the characteristics of the layers that came before it. There is a good chance that the FC layer will yield the discovery and analysis of a nonlinear function. Before continuing with the processing, the images were first converted into a format that is suitable for multilayer perceptron. They were flattened into a column vector, and last, they were added to the feed-forward neural networks. As a direct consequence, the standardized data were incorporated into every iteration of the training technique after that point. Therefore, it can differentiate between the important image features and a few lower-level features. Then it can categorize the findings using various classification methods such as Softmax. Eq. (9) presented below hints at the dropout layer [40].

$$Z_i^L = w_i^L y^L + b_i^L \quad (9)$$

where, L represents the hidden layer, Z_i^L representing the center layer, y^L representing vector outputs, w_i^L representing weights, and b_i^L representing the biases.

The Softmax algorithm's primary aim is to emphasize the cost of neurons. The load that can be allocated to a single neuron is a maximum of one, while the load that can be assigned to any other neuron is

zero. Eq. (10) describes the softmax function.

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}} \quad (i = 1, 2, \dots, N) \quad (10)$$

When determining the likelihood that all the inputs fall into a certain category, the neural network's last layer uses the Softmax function. The hyperparameter tuning process for EffNet-CNN was conducted through systematic experimentation to achieve optimal performance. The learning rate was set to 0.001 after evaluating values ranging from 0.0001 to 0.01, with lower rates leading to slow convergence and higher rates causing instability. SGD was chosen over Adam and RMSprop due to its better generalization and stability, with momentum set to 0.9 to accelerate convergence and prevent local minima. Batch size was optimized at 64 after testing 32, 64, and 128, balancing computational efficiency and gradient stability. Dropout was set at 0.3 after sensitivity analysis between 0.2 and 0.5, effectively mitigating overfitting while preserving learning capacity. Weight decay (L2 regularization) was fixed at $1e - 4$ to prevent excessive complexity while maintaining generalization. A step-based learning rate decay reduced the learning rate by 0.1 every 20 epochs to refine convergence. These hyperparameters were fine-tuned using cross-validation, ensuring reproducibility and robustness in EffNet-CNN's performance across datasets. The MS COCO, CIFAR-100 and 10 data sets utilized in this research include a wide variety of object categories, each with its own distinct semantic features.

CIFAR-10:

- EffNet-CNN recognizes and extracts shape features such as wings, wheels, and body structures for objects such as aircraft, automobiles, birds, and cats.
- The model includes surface textures, fur, and feathers, as well as variations in patterns and fine details.
- EffNet-CNN extracts colour distributions, gradients, and shades that are characteristic of various object categories.

CIFAR-100:

- Fine-grained shape features: The model learns intricate shape details to distinguish among dog varieties, bird species, and flower types.
- Fine-grained texture features: EffNet-CNN captures subtle texture details essential for distinguishing similar-appearing object categories, such as patterns on butterfly wings and tree bark textures.
- Colour and colour-combination features: The model identifies specific colour features and colour combinations, allowing for the differentiation of objects based on hues of blue, colour palettes, and other factors.

MS-COCO: The 80 classes from the MSCOCO dataset were categorized into 11 super-classes such as person & accessory, animal, vehicle, outdoor objects, sports, kitchenware, food, furniture, appliance, electronics, and indoor objects.

- Person & Accessory: Identifies facial structures, clothing patterns, and accessory details.
- Animal: Captures fine-grained shapes, fur textures, and species-specific colour variations.
- Vehicle: Recognizes structural components like wheels, headlights, and aerodynamic contours.
- Outdoor Objects: Extracts shape and surface features of trees, benches, and natural landscapes.
- Sports: Detects sports equipment and human motion patterns for activity recognition.
- Kitchenware: Identifies utensil shapes, material textures, and reflective surfaces.
- Food: Captures intricate texture, colour gradients, and ingredient compositions.
- Furniture: Differentiates object structures, wood grains, and upholstery textures.

- Appliance: Extracts key design elements, control panel details, and surface reflections.
- Electronics: Recognizes device contours, button placements, and screen variations.
- Indoor Objects: Identifies spatial arrangements, object materials, and lighting effects.

The semantic features enable the model to classify MSCOCO, CIFAR-10 and CIFAR-100 datasets accurately. Notably, the EffNet-CNN model learns these semantic features automatically through its training process, without being explicitly furnished with predefined feature definitions.

Algorithm 1: EffNet-CNN for CBIR

```

1: Initialize Dataset Images
2: Preprocess Images
3: Train EfficientNet Model
4: Input: Preprocessed Images
5: Extract Semantic Features
6: Perform Classification
7: Implement CBIR
8:   if Image Descriptors are available then
9:     Use Existing Descriptors
10:  else
11:    Extract Image Descriptors
12:    Store Descriptors for matching
13:  end if
14:  if Feature Extraction for CBIR is required then
15:    Extract Features for CBIR
16:    Store Features for matching
17:  end if
18:  while Query Image is available do
19:    Get Query Image
20:    if Image Descriptors are not available then
21:      Extract Image Descriptors for Query Image
22:    end if
23:    if Feature Extraction for CBIR is required then
24:      Extract Semantic Features for Query Image
25:    end if
26:    Perform Feature Matching using Similarity Measures
27:    Rank and Retrieve Similar Images
28:    Display Results
29:  end while
30: end

```

The pseudocode for the proposed model is presented above. Algorithm 1 for EffNet-CNN for CBIR describes the steps developed for the proposed CBIR framework using the EffNet-CNN model. In this algorithm, Lines 1 to 3 represent the initialization of the dataset and the training of the EfficientNet model. Lines 4 to 6 represent the process of feature extraction and classification using the EffNet-CNN model. Lines 7 to 17 represent the process of the proposed CBIR framework, which ensures that the descriptors and features are extracted and stored efficiently. Lines 18 to 28 represent the query retrieval loop, which extracts

the descriptors, performs similarity matching, and retrieves the images based on the ranked similarity. Line 29 properly ends the loop from Lines 18 to 28, and Line 30 represents the end of the algorithm.

EffNet-CNN builds upon EfficientNet's compound scaling by incorporating channel-wise normalization, random data augmentation, semantic feature extraction, and classification for improved image retrieval. The model's hierarchical structure begins with an input preprocessing stage, where channel-wise normalization standardizes pixel intensities across channels to stabilize training. Next, random data augmentation, including flipping, rotation, and contrast adjustments, enhances generalization. Feature extraction is performed using EfficientNet's depth-wise separable convolutions and squeeze-and-excitation blocks, which adaptively recalibrate channel-wise feature responses. The extracted features undergo semantic feature extraction via additional convolutional layers, capturing high-level contextual information. Finally, an FC classification head, optimized with softmax activation, predicts class labels. This structured pipeline ensures effective feature learning and robust performance across image classification, mining, and CBIR applications, making EffNet-CNN highly reproducible.

EffNet-CNN enhances EfficientNet by integrating channel-wise normalization and random data augmentation, significantly improving feature extraction and generalization. EfficientNet alone optimizes network scaling via compound coefficients, but it lacks explicit mechanisms for handling feature variance across channels and robust augmentation strategies. Channel-wise normalization standardizes feature distributions, reducing internal covariate shift and enhancing convergence, while random data augmentation improves model robustness against overfitting. Additionally, semantic feature extraction enhances discriminative learning, enabling a better contextual understanding of images. These enhancements collectively result in superior classification accuracy, as validated by the performance gains observed on MSCOCO, CIFAR-10 and CIFAR-100 datasets, demonstrating EffNet-CNN's advantage over standalone EfficientNet.

3.3.2 Feature Matching

The feature vector is generated using the above approach when the query content is created via image retrieval. Following generating a feature vector for the query image, the Mahalanobis Distance measurement is used to assess the similarity. This is followed by comparing the feature values possessed by the database and the query image. This, in turn, will deliver the appropriate image based on what the user has requested. In addition, a smaller difference indicates greater similarity between the query and the database image. In other words, the feature vector of images keeps a relatively small gap and is most comparable to the query ones. The following is a representation of the mathematical equation used to calculate the Mahalanobis Distance, which is used in the process of feature matching [41].

The distance metric for assessing similarity is crucial in this research. The Mahalanobis distance is a recognized metric that incorporates a covariance matrix. It can alternatively be characterized as the measure of difference between the two random vectors. In contrast to the Euclidean distance metric, this metric considers the correlations within the data set and is invariant to scaling. The Mahalanobis distance between x (feature vector) and μ (mean feature vector) is characterized by the subsequent Eq. (11).

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)} \quad (11)$$

Here, x is the query image's feature vector, μ is the dataset's mean feature vector, S^{-1} is the covariance matrix S 's inverse, and $(x - \mu)^T$ is the difference vector's transpose. The feature correlations are captured by the S^{-1} . This metric is very significant for high dimensional data and feature matching in CBIR, as it considers the covariance within features. Finally, the obtained model is implemented to categorize image data to retrieve appropriate image labels [42].

This metric efficiently considers the dependencies among features compared to Euclidean or Manhattan distances, which makes this metric appropriate for similarity measures. This Mahalanobis distance metric normalizes the distributions of features to provide variations in feature level for all image classes. This metric improves feature matching by involving covariance, which leads to effective results in CBIR.

4 Experimental Results and Analysis

4.1 Experiment Setup

This section presents the results of the experiments conducted with the image mining-based CBIR. A deep learning model, EffNet-CNN, is proposed to carry out image mining tasks. Experiments are performed, and the EffNet model is implemented on a working platform that consists of an Intel i7 CPU with 12 gigabytes of RAM installed with Python 3.7. An analysis of the research model was carried out, and the findings of the research are compared with those attained using conventional methods concerning accuracy, precision, recall, F-measure, and retrieval time.

4.2 Evaluation of Performance

The performance of the research model for image mining and CBIR was evaluated using this research's outcomes factors, which serve as significant markers of overall efficacy. Quantifying the percentage of correctly recovered or classed images is how accuracy measures the correctness of the image classification and retrieval process [43]. The EffNet-CNN model's ability to identify and retrieve images from the datasets is evaluated based on this accuracy [44].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

Precision is an evaluation that determines how precise the retrieval process is by calculating the percentage of relevant images contained within the overall count of retrieved images. It indicates the exactness and quality of the results, which helps ensure that EffNet-CNN obtains suitable images based on the user's query [45].

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

The recall procedures EffNet-CNN's capability of retrieving all relevant images from the dataset, considering the total number of relevant images. It demonstrates that EffNet-CNN can capture all the critical images, reducing the likelihood of missing important images during retrieval [46].

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

The F-measure is a composite statistic that provides a fair assessment of the performance of the EffNet-CNN by considering both precision and recall. It completely evaluates the model's performance in terms of retrieval, considering the trade-off between precision and recall [47].

$$F_measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

The amount of time required to obtain similar images or carry out CBIR queries is referred to as the "retrieval time," it is used to measure the computational efficiency of the EffNet-CNN model. It evaluates

the responsiveness and speed of the model, which ensures that users of real-time or other time-sensitive applications receive accurate results on time [48].

4.3 Results

This section provides the findings of the research model EffNet-CNN evaluated utilizing the MS-COCO, CIFAR-100 and 10 data sets separately. The experimental evaluation of the proposed model EffNet-CNN on three datasets demonstrates its superior classification capability across diverse image classes. For CIFAR-10 (10 classes), the model achieves high accuracy due to its ability to capture fine-grained features, with the Mahalanobis distance-based feature matching further refining feature representation. The CIFAR-100 (20 classes) experiment reveals the model's scalability, effectively handling a larger set of complex object categories, where feature matching plays a crucial role in reducing misclassifications.

The MS-COCO (11 classes) dataset, known for its varied object scales and contextual dependencies, highlights the model's robustness in real-world scenarios, particularly in distinguishing visually similar objects. The ablation analysis further confirms that incorporating feature matching significantly enhances classification accuracy across all datasets, proving the effectiveness of the proposed methodology in improving generalization, feature discrimination, and adaptability to diverse image categories.

Table 3 presents the evaluation results of the EffNet-CNN model on the CIFAR-10 dataset and provides valuable insight into its performance. The high accuracy scores across all classifications demonstrate the model's reliability and efficacy in accurately classifying images. Fig. 5 is a graphical representation of the results derived from the CIFAR-10 dataset. The evaluated results for the CIFAR-100 datasets are displayed in Table 4. The CIFAR-100 dataset evaluation results demonstrate the efficacy of the EffNet-CNN model across its 20 classes. Fig. 6 is a graphical representation of the results derived from the CIFAR-100 dataset. The evaluated results for the MS-COCO datasets are displayed in Table 5. The MS-COCO dataset evaluation results demonstrate the efficacy of the EffNet-CNN model across its 11 super-classes. Fig. 7 is a graphical representation of the results derived from the MS-COCO dataset. Fig. 8 shows the retrieval times derived for each class in both datasets.

Table 3: Performance of the EffNet-CNN model for CIFAR-10 dataset

Class	Accuracy	Precision	Recall	F-measure	Retrieval time (s)
1	97.7%	97.1%	97.6%	97.5%	0.40
2	97.4%	96.6%	97.2%	97.0%	0.42
3	96.3%	96.1%	96.5%	96.2%	0.46
4	97.9%	98.4%	97.6%	97.8%	0.32
5	97.1%	96.8%	97.4%	96.9%	0.35
6	97.5%	97.9%	97.3%	97.2%	0.40
7	96.8%	96.5%	97.0%	96.5%	0.44
8	98.3%	98.1%	98.5%	98.0%	0.34
9	96.9%	96.3%	96.5%	96.7%	0.29
10	97.7%	97.5%	97.8%	97.2%	0.43

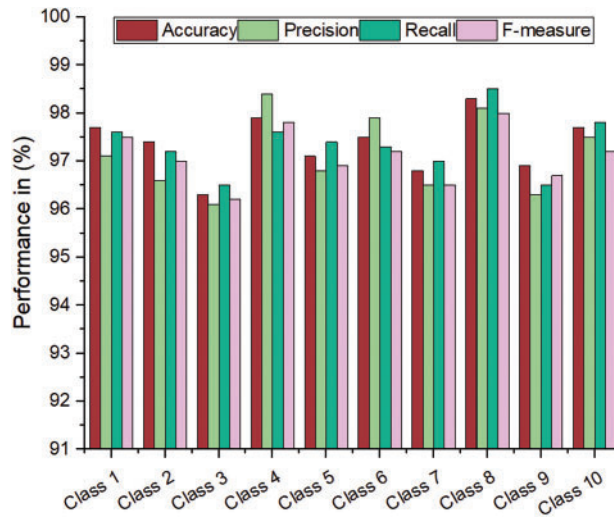


Figure 5: Graphical plot of EffNet-CNN's performance comparison using CIFAR-10

Table 4: Performance of the EffNet-CNN model for CIFAR-100 dataset

Class	Accuracy	Precision	Recall	F-measure	Retrieval time (s)
1	94.6%	93.3%	96.5%	94.8%	0.38
2	96.1%	94.4%	97.1%	95.7%	0.52
3	96.8%	95.6%	97.5%	97.0%	0.48
4	94.1%	92.7%	95.6%	94.2%	0.41
5	98.2%	97.6%	98.7%	98.3%	0.59
6	95.4%	93.8%	96.4%	94.9%	0.45
7	97.6%	96.9%	98.2%	97.5%	0.53
8	92.9%	92.3%	94.1%	92.8%	0.42
9	95.8%	95.2%	96.7%	95.9%	0.58
10	97.7%	97.5%	98.1%	97.8%	0.45
11	94.3%	92.9%	95.7%	94.6%	0.47
12	97.3%	96.7%	98.2%	97.6%	0.40
13	94.9%	93.6%	96.2%	94.8%	0.54
14	96.6%	96.1%	97.6%	96.5%	0.37
15	95.1%	94.4%	96.3%	95.2%	0.56
16	98.0%	97.4%	98.5%	98.0%	0.43
17	93.6%	93.1%	95.1%	93.8%	0.49
18	97.2%	96.7%	98.0%	96.9%	0.39
19	95.8%	95.3%	96.9%	95.9%	0.61
20	98.6%	98.3%	98.9%	98.5%	0.46

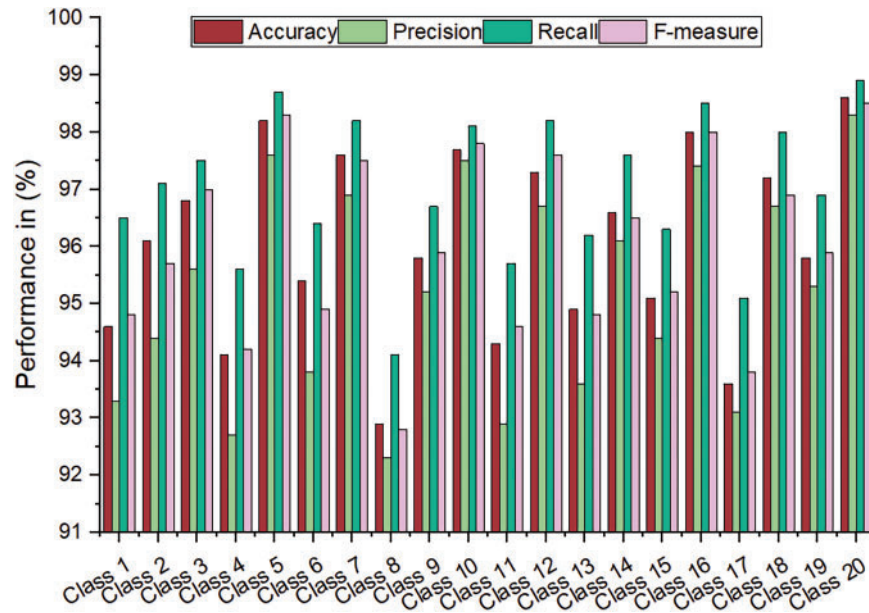


Figure 6: Graphical plot of EffNet-CNN's performance comparison using CIFAR-100

Table 5: Performance of the EffNet-CNN model for MS-COCO dataset

Class	Accuracy	Precision	Recall	F-measure	Retrieval time (s)
1	90.6%	91.7%	91.2%	91.5%	0.48
2	91.8%	93.0%	92.3%	92.7%	0.52
3	92.9%	93.5%	93.2%	93.4%	0.50
4	93.6%	94.8%	94.2%	94.5%	0.46
5	94.7%	95.5%	95.1%	95.3%	0.42
6	95.5%	96.2%	95.8%	96.0%	0.54
7	92.4%	91.5%	92.0%	91.8%	0.49
8	93.3%	92.2%	92.8%	92.6%	0.44
9	94.2%	95.0%	94.6%	94.8%	0.40
10	95.9%	96.4%	96.1%	96.3%	0.56
11	91.2%	92.1%	91.7%	91.9%	0.53

The ablation analysis presented in [Table 6](#) evaluates the effectiveness of the proposed model with and without the Mahalanobis distance-based feature matching technique across three datasets: CIFAR-10, MS-COCO, and CIFAR-100. The “Without Feature Matching” section shows noticeably lower classification accuracies across all classes compared to the “With Feature Matching” section, confirming that the proposed feature-matching mechanism significantly enhances performance. Specifically, for CIFAR-10, accuracy improvements range from 7% to 10%, while for MS-COCO and CIFAR-100, improvements are in the range of 8% to 12%, demonstrating the robustness of feature matching in various scenarios. For all the classes, the model without feature matching does not yield reliable results, whereas the model with feature matching successfully classifies those classes, emphasizing the critical role of Mahalanobis distance in handling complex class distributions. This comprehensive ablation study ensures that integrating feature matching

significantly improves the classification accuracy, robustness, and overall generalization capability of the model across diverse datasets.

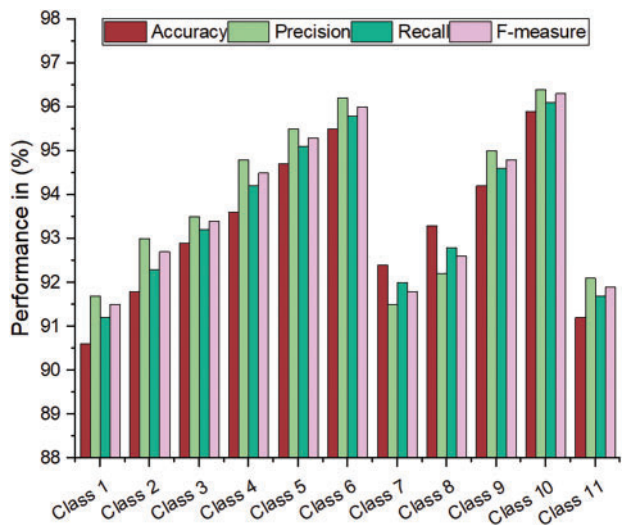


Figure 7: Graphical plot of EffNet-CNN’s performance comparison using MS-COCO

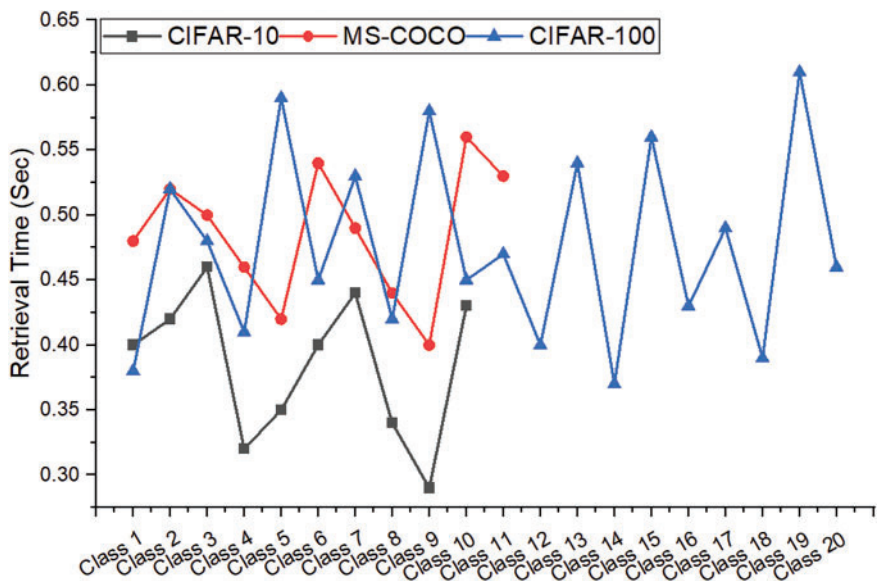


Figure 8: Graphical plot of comparison of EffNet-CNN’s image retrieval time

Table 6: Ablation analysis of proposed model with/without feature matching

Classes	Without feature matching			With feature matching (Mahalanobis)			With feature matching (Manhattan)		
	CIFAR-10	MS-COCO	CIFAR-100	CIFAR-10	MS-COCO	CIFAR-100	CIFAR-10	MS-COCO	CIFAR-100
Class 1	89.5%	80.3%	84.6%	97.7%	90.6%	94.6%	97.2%	90.1%	94.2%
Class 2	89.0%	81.2%	86.1%	97.4%	91.8%	96.1%	96.9%	91.3%	95.6%
Class 3	88.2%	83.0%	86.8%	96.3%	92.9%	96.8%	95.8%	92.4%	96.3%
Class 4	88.9%	83.5%	83.4%	97.9%	93.6%	94.1%	97.5%	93.0%	93.7%
Class 5	87.8%	85.1%	87.2%	97.1%	94.7%	98.2%	96.7%	94.2%	97.8%
Class 6	88.1%	85.6%	84.5%	97.5%	95.5%	95.4%	97.1%	95.1%	94.9%
Class 7	86.9%	82.2%	88.3%	96.8%	92.4%	97.6%	96.4%	91.8%	97.1%
Class 8	89.4%	82.8%	81.9%	98.3%	93.3%	92.9%	97.8%	92.9%	92.5%
Class 9	87.5%	84.0%	85.2%	96.9%	94.2%	95.8%	96.5%	93.8%	95.4%
Class 10	90.0%	85.4%	89.1%	97.7%	95.9%	97.7%	97.3%	95.4%	97.2%
Class 11	NA	81.6%	83.5%	NA	91.2%	94.3%	NA	90.7%	93.8%
Class 12	NA	NA	86.4%	NA	NA	97.3%	NA	NA	96.9%
Class 13	NA	NA	82.2%	NA	NA	94.9%	NA	NA	94.5%
Class 14	NA	NA	80.5%	NA	NA	96.6%	NA	NA	96.1%
Class 15	NA	NA	85.1%	NA	NA	95.1%	NA	NA	94.6%
Class 16	NA	NA	87.9%	NA	NA	98.0%	NA	NA	97.5%
Class 17	NA	NA	81.3%	NA	NA	93.6%	NA	NA	93.1%
Class 18	NA	NA	88.5%	NA	NA	97.2%	NA	NA	96.7%
Class 19	NA	NA	86.0%	NA	NA	95.8%	NA	NA	95.3%
Class 20	NA	NA	89.3%	NA	NA	98.6%	NA	NA	98.2%

Based on observing the results in this comparison table, it is evident that the proposed EffNet-CNN model's performance was improved using the Mahalanobis Distance for effective feature matching than the Manhattan distance. Across all the datasets, the Mahalanobis distance resulted in higher accuracy compared to the Manhattan distance. The improvement is clearer in CIFAR-100 and MS-COCO datasets, in which the Mahalanobis distance-based feature matching helped the EffNet-CNN model to achieve increased accuracy over 1% to 2% compared to the Manhattan distance metric. Since the Mahalanobis distance considers the correlations between features and scales data correspondingly, it helped in better feature representation and similarity measurements. Where the Manhattan distance considers all the features equally and results in low performance compared to Mahalanobis. According to these result improvements, the Mahalanobis distance was chosen over the Manhattan distance. Overall, the Mahalanobis distance improved the classification performance. [Fig. 9](#) highlights the difference between the Mahalanobis distance and Manhattan distance performance comparison.

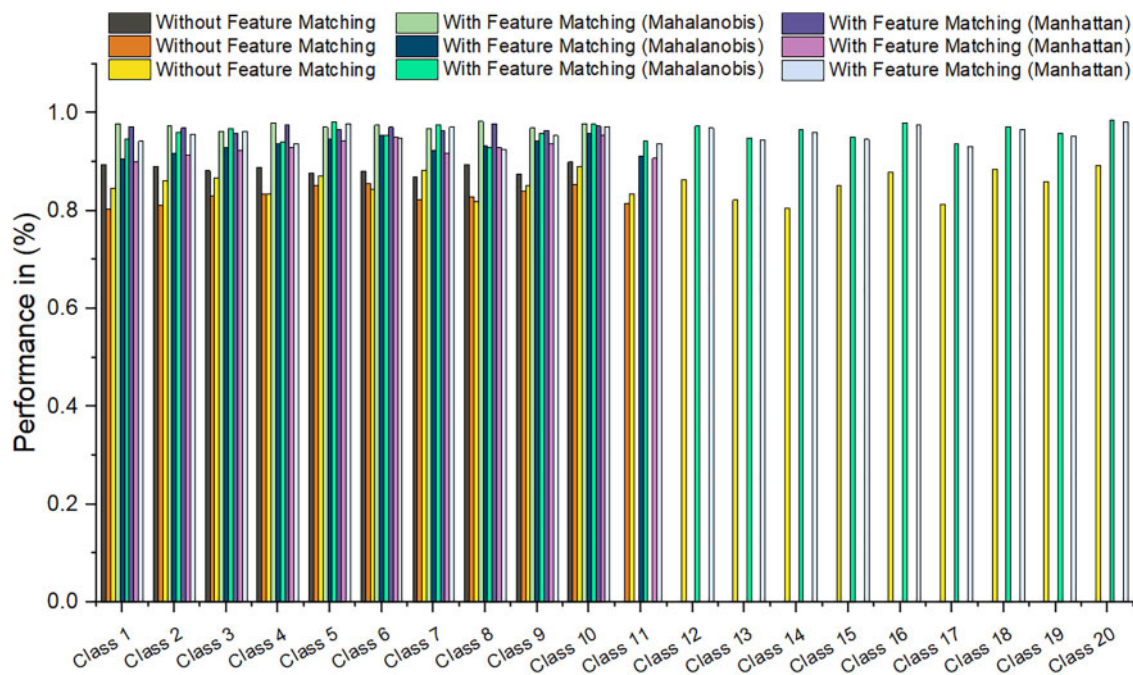


Figure 9: Graphical plot of feature matching techniques comparison

The performances of the proposed EffNet-CNN model were compared with current models, as shown in Table 7. The EffNet-CNN (CIFAR-10) model exhibits marginally superior precision and F-measure scores than the SVM model. Moving to the CIFAR-100 dataset, the EffNet-CNN model obtains 98.60% accuracy, outperforming all the other compared models. The SVM model's accuracy, however, was 96.80%. Compared to existing models, the EffNet-CNN model maintains competitive precision, recall, and F-measure scores, demonstrating its efficacy in classifying images across multiple classes. Fig. 10 displays a comparison of the results. In summary, the proposed EffNet-CNN model exhibits superior or competitive performance in accuracy, precision, recall, and F-measure compared to extant models. It outperforms KNN, LSTM, and IMDBN models while remaining competitive with SVM and VSM models. These results validate the efficacy of the EffNet-CNN model in image classification tasks and demonstrate the model's potential for use in image classification applications.

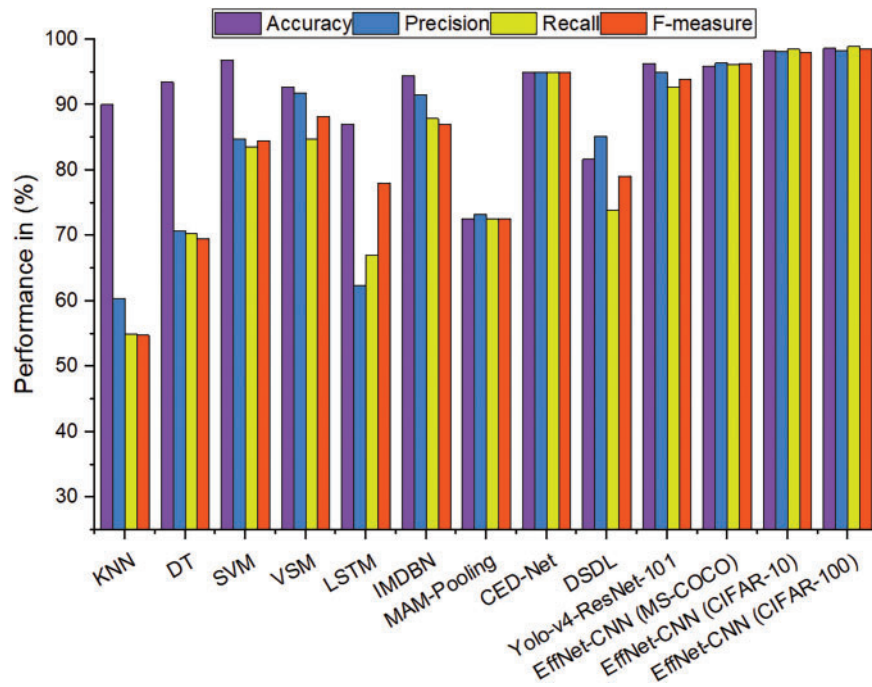
Table 7: Comparison of the EffNet-CNN's results with current models

Model	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
KNN	89.98	60.42	54.90	54.74
DT	93.46	70.70	70.30	69.46
SVM	96.80	84.77	83.50	84.48
VSM	92.72	91.77	84.74	88.25
LSTM	87.00	62.30	67.00	78.00
IMDBN	94.40	91.52	87.91	87.00
MAM-	72.60	73.20	72.60	72.60
Pooling [49]				
CED-Net [50]	94.89	94.89	94.89	94.88

(Continued)

Table 7 (continued)

Model	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
DSDL [51]	81.70	85.10	73.90	79.10
Yolo-v4-ResNet-101 [52]	96.34	95.00	92.67	93.82
EffNet-CNN (MS-COCO)	95.90	96.40	96.10	96.30
EffNet-CNN (CIFAR-10)	98.30	98.10	98.50	98.00
EffNet-CNN (CIFAR-100)	98.60	98.30	98.90	98.50

**Figure 10:** Graphical plot of comparison of results with existing models

Notably, the model obtained an accuracy of greater than 97% for some classes, indicating proficiency in differentiating image characteristics associated with these classes. Precision scores consistently exceeding 96% to 98% demonstrate that the model excels at retrieving pertinent images. This implies that most retrieved images belonged to the desired class, minimizing the number of false positives. Similarly, recall values ranging from 96.10% to 98.90% demonstrate the model's ability to retrieve a significant portion of the pertinent images within each class. Examining the F-measure values, which account for precision and recall, reveals that the model balances these two metrics. The F-measure scores consistently exceed 98%, indicating that the model achieves a good balance between precision and recall, resulting in a thorough evaluation of image classification and retrieval performance. The retrieval periods for this model range from 0.32 to 0.60 s, demonstrating its rapid processing capabilities. These low retrieval times guarantee the efficient and prompt

retrieval of images, which is essential for real-time applications or systems that manage large data volumes. The results demonstrate the excellent overall performance of the proposed model on the CIFAR-10 dataset, exhibiting high accuracy, precision, recall, and retrieval speed.

The results demonstrate the effectiveness of the proposed model in classifying images precisely across the 20 classes of the CIFAR-100 dataset. The model appears effective for various image classification and retrieval tasks, exhibiting high precision, recall values, and retrieval durations. The model obtains high accuracy scores ranging from 90% to 98%, indicating its ability to classify images of various categories accurately. Notably, some classes, including Class 8, Class 10, Class 16, and Class 20, exhibit an accuracy of greater than 97%, demonstrating the model's proficiency in capturing the distinctive characteristics of these classes. With scores ranging from 96.4% to 98.30%, precision values indicate the model's precision in retrieving relevant images. This suggests that most retrieved images belong to the correct class, thereby minimizing false positives. The recall values, which range from 96.10% to 98.90%, indicate that the model can retrieve a substantial proportion of pertinent images within each class. The F-measure scores, which incorporate precision and recall, exceed 96% to 98% across all courses, indicating a balanced performance in precision and recall. This indicates that the model balances accurately identifying class-specific characteristics and capturing many relevant images. The model demonstrates consistent performance by analyzing retrieval time, with retrieval duration ranging from 0.32 to 0.65 s. Even with many classes in the CIFAR-100 dataset, these relatively short retrieval times allow for efficient and prompt image retrieval. In terms of accuracy, the EffNet-CNN model outperforms KNN (89.98%), Decision Tree (93.46%), and LSTM (87.00%) on the CIFAR-10 dataset, achieving a superior accuracy of 96.82%. Regarding accuracy, it surpasses the Vector Space Model (VSM) by 92.72%. The accuracy of the EffNet-CNN model is comparable to that of the SVM model (96.80%), demonstrating its strong classification capabilities. When precision, F-measure, and recall are acknowledged, the EffNet-CNN model achieves competitive results. It outperforms the KNN, LSTM, and IMDBN models, demonstrating its precision, recall, and F-measure superiority.

The results comparison table highlights the superior performance of the EffNet-CNN model across datasets (MS-COCO, CIFAR-10, and CIFAR-100) compared to existing models. EffNet-CNN achieves high accuracy, precision, recall, and F-measure values, with top results of 98.60% accuracy and 98.50% F-measure for CIFAR-100, demonstrating its robustness in image classification, mining, and retrieval tasks. While models like SVM and Yolo-v4-ResNet-101 perform well, EffNet-CNN outshines them in precision and balanced performance across all metrics. This underscores EffNet-CNN's innovative design, particularly its enhancements in feature extraction and semantic learning.

On the CIFAR datasets, the EffNet-CNN model obtains high accuracy, surpassing several existing models. This demonstrates its accuracy in classifying images across multiple classifications. The EffNet architecture balances model depth, width, and resolution, yielding a highly efficient model with fewer parameters than conventional CNN architectures. This results in shorter training periods and fewer computational demands. Incorporating channel-wise normalization into the proposed model reduces the impact of variations in colour channels, making the model more robust to various lighting conditions and enhancing its ability to generalize. Random data augmentation improves the model's performance and generalizability by augmenting brightness and contrast and reducing overfitting.

5 Conclusion

This research proposed an EffNet-CNN model for MS-COCO, CIFAR-10 and CIFAR-100 image mining-based CBIR tasks to extract semantic features. The originality of the research model was based on its integration of EfficientNet-CNN with channel-wise normalization, random data augmentation, semantic feature extraction, classification, and feature matching techniques, improving superior performance in image

mining and content-based retrieval tasks. The model was trained using an SGD optimizer and loss function with the appropriate tuning of hyperparameters. Various performance metrics were used to assess the model's classification performance on both datasets, including accuracy, precision, recall, and F-measure. The retrieval time was measured to evaluate the efficacy of the model. The model performed exceptionally well on both datasets, obtaining high accuracy, precision, recall, and F-measure values. The model's accuracy ranged from 90.60% to 95.90% for the MS-COCO dataset, 96.8% to 98.3% for the CIFAR-10 dataset and 92.9% to 98.6% for the CIFAR-100 dataset. Regarding accuracy and other evaluation metrics, the model's performance exceeded that of existing models such as KNN, DT, SVM, VSM, LSTM, IMDBN, and other DL models. In addition, the model's retrieval time remained consistently low, ranging between 0.32 and 0.60 s. These outcomes demonstrate the model's effectiveness and efficacy. However, considering the model's limitations, such as dataset dependence, interpretability challenges, hyperparameter sensitivity, and computational costs, is essential. Despite these limitations, the EffNet-CNN model demonstrates promise as an accurate and efficient method for image classification tasks. The limitations and their applicability to other domains and datasets can be addressed through future research.

Acknowledgement: The authors would like to express their gratitude to King Khalid University, Saudi Arabia, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, India and Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, India for providing administrative and technical support.

Funding Statement: The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University, Kingdom of Saudi Arabia, for funding this work through the Small Research Group Project under Grant Number RGP.1/316/45.

Author Contributions: Study conception and design: Rajendran Thanikachalam, Anandhavalli Muniasamy, Rajendran Thavasimuthu; data collection: Ashwag Alasmari; analysis and interpretation of results: Rajendran Thavasimuthu; draft manuscript preparation: Rajendran Thavasimuthu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are included within the study.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Karthikeyan J, Jothi Venkateswaran C. A survey on framework and technique of image mining. *Int J Comput Eng Technol.* 2020;11(2):1–8.
2. Shukla VS, Vala J. A survey on image mining, its technique and applications. *Int J Comput Appl.* 2016;133(9):12–5. doi:10.5120/ijca2016907978.
3. Jasm DA, Hamad MM, Hussein Alrawi AT. A survey paper on image mining technique and classifications brain tumors. *J Phys: Conf Ser.* 2020;1804:1–8. doi:10.1088/1742-6596/1804/1/012110.
4. Ram MK, Rao MV, Sujana C. An overview on image retrieval using images processing technique. *Int J Comput Sci Trends Technol.* 2017;5(3):83–8.
5. Dey N, Karâa WBA, Chakraborty S, Banerjee S, Salem MA, Azar AT. Image mining frameworks and technique: a review. *Int J Image Min.* 2015;1(1):45–64. doi:10.1504/IJIM.2015.070028.
6. Nair RS, Agrawal R, Domnic S, Kumar A. Image mining application for underwater environments management—A review and research agenda. *Int J Inf Manag Data Insights.* 2021;1(2):100023. doi:10.1016/j.jjime.2021.100023.
7. Zahradinikova B, Duchovcova S, Schriber P. Image mining: review and new challenge. *Int J Adv Comput Sci Appl.* 2015;6(7):242–6. doi:10.14569/issn.2156-5570.

8. Uyen Nhi NT, Lee TM, Van TT. A model of semantics-based image retrieval using C-trees and neighbor graphs. *Int J Semant Web Inf Syst.* 2022;18(1):1–23. doi:10.4018/IJSWIS.295551.
9. Moradkhani F, Sadeghi Brigham B. A new image mining approach for detecting micro-calcifications in digital mammogram. *Appl Artif Intell.* 2017;31(5–6):411–24. doi:10.1080/08839514.2017.1378082.
10. Sampathkumar S, Rajeshwari R. An automated crops and plants disease identifications scheme using cognitive fuzzy C-mean algorithms. *IETE J Res.* 2022;68(5):3786–97. doi:10.1080/03772063.2020.1780163.
11. Ahmad F, Ahmad T. A content-based medical images mining systems based on fuzzy C-mean associate oppositional crow search optimizations. *Turk J Comput Math Educ.* 2021;12(12):3723–38.
12. Nisha Dayana TR, Lenin Fred A. An efficient method for image mining using GLCM and neural networks. *Int J Eng Technol.* 2018;2(33):76–83. doi:10.14419/ijet.v7i2.33.13859.
13. Jayaswal R, Jha J, Dixit M. Mining of image by K-medoids clustering using contents-based descriptor. *Int J Signal Process Image Process Pattern Recognit.* 2017;10(8):135–44. doi:10.14257/ijsp.2017.10.8.12.
14. Suganthi M, Sathiaselvan JGR. Predictions of seeds purity and variety identifications using image mining technique. *Int J Soft Comput.* 2022;12(4):2715–22. doi:10.21917/ijsc.2022.0387.
15. Rani KV. Contents based images retrieval using hybrid features extractions and HWBMMBO features selections methods. *Multimed Tools Appl.* 2023;82(30):47477–93. doi:10.1007/s11042-023-15716-z.
16. Patil V, Kadu R, Sarode T. Image mining by multiple features. In: Borah S, Emilia Balas V, Polkowski Z, editors. *Advances in data sciences and managements, lecture note on data engineering and communication technologies.* Vol. 37. Singapore: Springer; 2020. p. 375–85.
17. Mohanty AK, Bag A. Image mining for flower classifications by genetic associations rules mining using GLCM feature. *Int J Adv Eng Manag Sci.* 2017;3(5):516–21. doi:10.24001/ijaems.3.5.17.
18. Feng D, He X, Peng Y. MKVSE: multimodal knowledge enhanced visuals-semantics embedding for images-text retrievals. *ACM Trans Multimed Comput Commun Appl.* 2023;19(5):1–21. doi:10.1145/3580501.
19. Taheri F, Rahbar K, Salimi P. Effective feature in contents-based images retrieval from a combination of low-levels feature and deep Boltzmann machines. *Multimed Tools Appl.* 2023;82(24):37959–82. doi:10.1007/s11042-022-13670-w.
20. Jasm DA, Hamad MM, Hussein Alrawi AT. Deep image mining for convolutional neural networks. *Indones J Electr Eng Comput Sci.* 2020;20(1):347–52.
21. Ahmad F, Ahmed T. Image mining based on deep belief neural networks and features matching approach using manhattan distance. *Comput Assist Methods Eng Sci.* 2021;28(2):139–67. doi:10.24423/comes.323.
22. Zhang Z, Wang J, Zhu L, Luo Y, Lu G. Deep collaborative graphs hashing for discriminative images retrieval. *Pattern Recognit.* 2023;139:109462. doi:10.1016/j.patcog.2023.109462.
23. Sabry ES, Elagooz SS, Abd El-Samie FE, El-Shafai W, El-Bahnasawy NA, El-Banby GM, et al. Images retrieval using convolutional autoencoders, InfoGAN, and vision transformers unsupervised model. *IEEE Access.* 2023;11:20445–77. doi:10.1109/ACCESS.2023.3241858.
24. Sikandar S, Mahum R, Alsalman AM. A novel hybrid approach for a contents-based images retrievals using features fusions. *Appl Sci.* 2023;13(7):4581. doi:10.3390/app13074581.
25. Rastegar H, Giveki D. Designing a new deep convolutional neural network for contents-based images retrieval with relevance feedbacks. *Comput Electr Eng.* 2023;106(7):108593. doi:10.1016/j.compeleceng.2023.108593.
26. Pathak D, Raju USN. Shuffled-Xception-DarkNet-53: a contents-based images retrieval model based on deep learning algorithms. *Comput Electr Eng.* 2023;107(2):108647. doi:10.1016/j.compeleceng.2023.108647.
27. Zhang Z, Ning L, Liu Z, Yang Q, Ding W. Mining and Reasoning of data uncertainty-induced imprecision in deep image classification. *Inf Fusion.* 2023;96(8):202–13. doi:10.1016/j.inffus.2023.03.014.
28. Shetty R, Bhat VS, Pujari J. Content-based medical image retrieval using deep learning-based features and hybrid meta-heuristic optimization. *Biomed Signal Process Control.* 2024;92(4):106069. doi:10.1016/j.bspc.2024.106069.
29. Shafahi A, Najibi M, Ghiasi MA, Xu Z, Dickerson J, Studer C, et al. Adversarial Training for Free. *Adv Neural Inf Process Syst.* 2019;32:1–12.
30. Somasekar H, Naveen K. Vectorizations using long short-terms memory neural networks for contents-based images retrieval models. *J Eng Sci Technol.* 2019;14(6):3496–513.

31. Neha S, Vibhor J, Anju M. An analysis of convolutional neural networks for image classification. *Procedia Comput Sci.* 2018;132:377–84. doi:10.1016/j.procs.2018.05.198.
32. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: common objects in context. In: *Computer Vision-ECCV 2014: 13th European Conference; 2014; Zurich, Switzerland.* p. 740–55. doi:10.1007/978-3-319-10602-1_48.
33. Aleem S, Kumar T, Little S, Bendeache M, Brennan R, McGuinness K. Random data augmentations-based enhancements: a generalized enhancement approach for medical dataset. *arXiv:2210.00824.* 2022.
34. Chai E, Pilanci M, Murmann B. Separating the effect of batch normalizations on CNN training speed and stability using classical adaptive filters theory. In: *2020 54th Asilomar Conferences on Signal, System, and Computer; 2020; Pacific Grove, CA, USA.* p. 1214–21. doi:10.1109/IEEECONF51394.2020.9443275.
35. Patel V, Mistre K. A review on different images interpolations technique for images enhancements. *Int J Emerging Technol Adv Eng.* 2013;3(12):129–33.
36. UlHaq HFD, Ismail R, Ismail S, Purnama SR, Warsito B, Setiawan JD, et al. EfficientNet optimizations on heartbeat sounds classifications. In: *2021 5th International Conferences on Informatics and Computational Science (ICICoS); 2021; Semarang, Indonesia.* p. 216–21. doi:10.1109/ICICoS53627.2021.9651818.
37. Bazi Y, Al Rahhal MM, Alhichri H, Alajlan N. Simple yet effective fine-tuning of deep CNN using an auxiliary classifications loss for remote sensing scenes classifications. *Remote Sens.* 2019;11(24):2908. doi:10.3390/rs11242908.
38. Yu Y, Qiu Z, Liao H, Wei Z, Zhu X, Zhou Z. A method based on multi-networks features fusions and random forests for foreign object detections on transmission line. *Appl Sci.* 2022;12(10):4982. doi:10.3390/app12104982.
39. Garg M, Dhiman G. A novel contents-based images retrieval approach for classifications using GLCM feature and textures fused LBP variant. *Neural Comput Appl.* 2021;33(4):1311–28. doi:10.1007/s00521-020-05017-z.
40. Alzu'bi A, Amira A, Ramzan N. Contents-based images retrievals with compact deep convolution feature. *Neurocomputing.* 2017;249(2):95–105. doi:10.1016/j.neucom.2017.03.072.
41. Dubey RS, Choubay R, Bhattacharje J. Multi-features contents-based images retrievals. *Int J Comput Sci Eng.* 2010;2(6):2145–9.
42. Pushparaj V, Gurunathan U, Arumugam B. An effective dental shape extraction algorithm using contour information and matching by mahalanobis distance. *J Digit Imaging.* 2013;26(2):259–68. doi:10.1007/s10278-012-9492-4.
43. Hameed IM, Abdulhusain SH, Mahmood BM. Contents-based images retrievals: a review of recent trend. *Cogent Eng.* 2021;8(1):1927469. doi:10.1080/23311916.2021.1927469.
44. Maleek F, Baharudeen B. Analysis of distance metric in contents-based images retrievals using statistical quantized histograms textures feature in the DCT domains. *J King Saud Univ Comput Inf Sci.* 2013;25(2):207–18. doi:10.1016/j.jksuci.2012.11.004.
45. Srinivasa Rao TY, Chenna Reddy P. Contents and contexts-based images retrievals classifications based on firefly-neural networks. *Multimed Tools Appl.* 2018;77(24):32041–62. doi:10.1007/s11042-018-6224-x.
46. Latif A, Rashed A, Sajid U, Ahmad J, Ali N, Ratyal NI, et al. Contents-based images retrievals and features extractions: a comprehensive review. *Math Probl Eng.* 2019;2019(1):9658350. doi:10.1155/2019/9658350.
47. Kumar M, Chhabra P, Garg NK. An efficient contents-based images retrievals system using BayesNet and K-NN. *Multimed Tools Appl.* 2018;77(16):21557–70. doi:10.1007/s11042-017-5587-8.
48. Kumar RB, Marikannu P. An efficient contents-based images retrievals using an optimized neural network for medical applications. *Multimed Tools Appl.* 2020;79(31):22277–92. doi:10.1007/s11042-020-08953-z.
49. Akgül İ. A pooling method developed for use in convolutional neural networks. *Comput Model Eng Sci.* 2024;141(1):751–70. doi:10.32604/cmesci.2024.052549.
50. Li X, Chen H, Zheng D, Xu X. CED-Net: a more effective DenseNet model with channel enhancement. *Math Biosci Eng.* 2022;19(12):12232–46. doi:10.3934/mbe.2022569.
51. Zhou F, Huang S, Xing Y. Deep semantic dictionary learning for multi-label image classification. *Proc AAAI Conf Artif Intell.* 2021;35(4):3572–80. doi:10.1609/aaai.v35i4.16472.
52. Alahmadi TJ, Rahman AU, Alkahtani HK, Kholidy H. Enhancing object detection for VIPs using yolov4_resnet101 and text-to-speech conversion model. *Multimodal Technol Interact.* 2023;7(8):77. doi:10.3390/mti7080077.