Image Mining Based on Deep Belief Neural Network and Feature Matching Approach Using Manhattan Distance

Faiyaz AHMAD*, Tanvir AHMAD

Department of Computer Engineering
Jamia Millia Islamia Central University
New Delhi, India; e-mail: tahmad2@jmi.ac.in
*Corresponding Author e-mail: ahmad.faiyaz@gmail.com

Over the past few decades multimedia content, particularly digital images, has increased at a rapid pace, with several complex images being uploaded to various social websites such as Instagram, Facebook and Twitter. Therefore, it is difficult to search and retrieve the relevant image in seconds. Search engines retrieve images based on traditional text-based methods that depend on metadata and captions. In the last few years, a wide range of research has focused on content-based image retrieval (CBIR) based on image mining approaches. This is a challenging research area due to the ever-increasing multimedia database and other image libraries. In order to offer an effective search and retrieval, a novel CBIR system is proposed using the image mining-based deep belief neural network (IMDBN) technique. The proposed method is designed to enhance retrieval accuracy while diminishing the semantic gap between human visual understanding and image feature representation. To achieve this objective, the proposed system carries several steps like preprocessing, feature extraction, classification, and feature matching. Initially, the input database images are fed into the proposed image mining-based CBIR system, whereas colour-shape-texture (CST) feature extraction technique is applied to extract relevant feature set. The extracted features are fused and stored in the feature vector and are subjected to the proposed IMDBN classification step to retrieve similar images in one label. Whenever a new query content is created, the most relevant images are retrieved. This, in turn, achieves 94% accuracy, which is higher than in existing approaches.

Keywords: mining, deep belief neural network, Manhattan distance, median filter, energy, correlation, contrast, dissimilarity, homogeneity.

1. Introduction

Information is collected from various sources such as newspapers and books, and it is digitized nowadays. Hence it is available in the form of digital images. For accessing these types of larger databases, there is a need to develop

an efficient method for analysing the indexed form of an image database. In the digitized world, large amount of digital images must be handled by social networks, smart devices, and other data data resources as a result of rapid increase in the number of such images being uploaded daily. Due to this massive growth, mining a particular image from a vast collection is challenging. Nowadays, deep learning has shown great advantages in many fields [1, 2]. Additionally, image retrieval has been an active research field. The term image retrieval is defined as a process of browsing, searching, and retrieving images from a large collection of database images. The stored and shared multimedia contents are rising daily, so searching and retrieving relevant images from social media and other platforms play a major research challenge [3]. However, it is problematic to discover relevant information from a huge database. This, in turn, makes it essential for the image retrieval system to search and arrange suitable images, which tends towards a visual semantic connection along with reasoning from the user side. Conventionally many search engines retrieve images based on textual information that takes caption as input [4]. This is done by entering a query to the search engine in terms of keywords or text, and the outcome is obtained by similarity matching score. In this manner, there is a high chance of retrieving irrelevant images. The irrelevant outcome is due to the considerable variation in manual annotation/labeling and human visual perception [5]. Also, it is not always possible to label manually the existing large number of the image contained in archives [6].

With the drawbacks mentioned above, an automatic image annotation model was developed in later years for image retrieval analysis; it seems the images are labeled here on the basis of image contents [7]. This model works under the principle of how accurately the system is modeled to detect texture, colour, spatial information, shape, and edge layout materials [8, 9]. Still, various researches are conducted to perform automatic image annotation approach; however, the retrieval process can lead to mismatch outcomes because of possessing variation in visual perception. To overcome the existing shortcomings, researchers proposed a new scheme called CBIR [10]; this is a model built on visual analysis of contents with respect to images as derived from a part of the query image. The user query image is an incoming source in the CBIR approach. It matches visual contents of query image with an image available in multimedia [11]. Among the visual content of the query image, it finds a map with a multimedia image. It provides visual closeness depending on the feature of the image. Feature vector also shows an organized group of low-level features such as shape, spatial information, texture, colour and shape extracted with respect to the feature extraction method [12]. Query image and its similarity matching value are gathered to design relevant image as output. Early research [30] reported a modest approach termed as query-by-image content (QBIC) for an image retrieval me-

chanism that works by extracting low-level visual semantics. However, this leads to a mismatched outcome most of the time; hence, CBIR plays a major role in several applications such as military surveillance, medical image analysis, textile industry, crime detection, remote sensing, and video analysis [13]. Likewise, in applications such as image classification via network pruning and drop path pruning using deep convolutional neural networks (CNNs) [14], there are also some drawbacks such as the insufficient amount of training data and reduced efficiency. Methods such as the full stage of the data augmentation framework [15] introduced a problem of a complex network that suffers from minimizing costs. Similar techniques such as the spectrum interference-based two-level data augmentation method [16] in deep learning ought to expand signals and present variations that retain respective features rather than the failed ones to enhance the performance.

The main merits of using CBIR over the textual based image retrieval (TBIR) are listed as follows: (a) low-level features such as colour, shape, and texture information are extracted automatically, (b) there is no need to appoint experts to label caption, tags, keywords for an image, (c) offers more accurate outcome images, (d) manual error gets greatly reduced due to avoiding human intervention because of its automatic performance, and (e) there is no need for describing the image in textual format, hence it will not cause language barrier or ambiguity. With these benefits, finding suitable feature weight factors with regard to each label augments the retrieval system performance. Several existing types of research have been developed to enhance the retrieval accuracy of the system using low-level visual features. The extraction of features such as colour, shape, and texture from images was utilized to a great extent for classifying patterns while diminishing the computational burden of the system. However, establishing the most appropriate features for the retrieval system cause a recurring challenge that the computer vision community has yet to fully meet. The feature extraction approach can be solely applied in various applications such as detection of saliency, landscapes, biometrics, wood, biometrics, rock minerals, background subtraction, and cloth classification. In general, CBIR techniques, selection and extraction of features enable data representation from an image. This allows further classification and matching analysis at a reasonable time [17]. Even though data augmentation in different techniques helps to improve the ability of deep learning model with respect to classification, the performance of neural networks without deep learning is affected by various factors such as complex structure and weight initialization. The effectiveness of data augmentation has to be measured correspondingly with respect to effective neural networks.

This paper is organized as follows: Sec. 2 provides the detailed background and related works so far proposed. The subsequent section explains the problem statement identified and also the motivation of the research article. Section 4

presents the design of the proposed methodology in brief notes. Section 5 carries out the experimental result and discussions. The final section concludes the entire research work.

2. Background and related work

The CBIR model was initiated and developed by Kato in 1992 [24] and detailed in research article "Database architecture for content-based image retrieval". The author invented and discussed an automatic image retrieval system by extracting low-level features such as colour and shape. The basic requirement of the image retrieval system was to identify and arrange similar images from the social web with very low human interaction to the system. Various CBIR methods have been developed previously, and their overview and limitations are presented below.

Zhao *et al.* [18] have presented an adversarial patch that created an adversarial network; this modified the entire image and caused deep neural networks (DNNs) based on the image retrieval model to revive non-positive outcomes. An adversarial patch generative adversarial network (AP-GAN) was trained in an unsupervised method that requires only a low quantity of unlabelled data to train. After training, query-specific perturbations were generated for the query image to facilitate the query image. However, long experiments resulted that AP-GAN suited various application cases based on deep features within-person ID, vehicle ID and image retrieval. Singh and Batrav [19] have designed a bi-layer image retrieval content-based framework that consisted of two modules. First module eliminated highlights of dataset images with respect to shape, shading and surface. Similarly, in second module, there were two layers of comprisal: at first, all images were contrasted, and question images for shape and surface feature space and indexes of M most comparable pictures to the query image were retrieved. Next, M pictures retrieved from previous layer were matched with query pictures for shape and shading highlight space, and F pictures such as the query image were returned as an output Xie *et al.* [20] have introduced a technique that initially applied the content on the layout to recognize and remove the reliable zone of a picture and calculate the prevailing colour descriptor feature on the pixels in this consistent zone. Also, the translation and rotation invariance of the Hu moments was applied to remove the shape data in a similar predictable zone of the picture. At last, the mix of the predominant colour descriptor and the Hu minutes was used for content-based picture retrieval. The calculation proposed for the introduced CBIR was tested on three data sets: Corel-1k, Corel-5k, and Corel-10k. Kavitha and Saraswathi [21] have presented an image segmentation method using a fuzzy C-means clustering algorithm on satellite imagery. This image retrieval method split detection into three stages;

the initial stage improved the image using Gaussian filter then segmented the image using fuzzy C-means clustering algorithm. This technique was utilized to diminish the computational cost.

Likewise, Rudrappa and Vijapur [22] have built up a framework that utilizes ground-based pictures of clouds to characterize the cloud as significant-level, middle-level, and low-level. The authors used k-means clustering and CBIR strategies for cloud classification. The developed framework classified the clouds as low level, middle level, and high level clouds. Cloud class assumes a fundamental part of choosing the rainfall precipitation. The result of this cloud characterization can be used to contribute to a framework that progressively chooses rainfall precipitation. For this reason, the identification of the low-level clouds is of high significance. Wang *et al.* [23] have presented a technique dependent on shading for recovering pictures based on picture content, which is set up from the combination of shading and surface highlights. This gives a viable and adaptable assessment of how early humans dealt with visual substance. The combination of shading and surface highlights offers an incredible list of capabilities for shading picture recovery draws near. Results obtained from the trials showed that the proposed strategy recovered pictures more precisely than the other conventional techniques. Nonetheless, the element measurements were not more than various methods and needed a high computational cost. However, pairwise relation for both low-level highlights was used to calculate the closeness measure, which could be an issue. Ashraf *et al.* [24] introduced another CBIR procedure that utilizes the mix of colour and texture features to extract local vector, which is used as a feature vector. Colour moments are used for colour features, and to extract the texture features a discrete wavelet and Gabor wavelet methods are employed. To enhance feature vector representation, colour and edge descriptor is also included in the feature vector. At that point, this technique is compared with other remaining CBIR strategies, and the proposed approach is efficient in terms of feature extraction, and the efficiency and effectiveness of the proposed research outperform the existing research in term of average precision and recall values. Bu *et al.* [25] have introduced a CBIR technique using texture and colour features extracted from multi-resolution and multi-directional filter gain to easily create multi-resolution analysis results. It acquired hue, saturation, and value (HSV) colour pattern as varying characteristics to verify the human visual system and extract global and local features from the area of low to high frequency for all colour space. The experiments showed that precision with respect to recall presents a low dimension vector.

Thiagarajan *et al.* [26] presented a study over supervised local sparse coding of sub-image feature vectors to retrieve an image. The sparse representation was widely used in the image modeling domain and in computer vision application. The authors performed supervised local sparse coding with large overlapping

regions [27–29] proposed multiple local and global features and designed a dictionary for supervised local sparse coding of sub-image heterogeneous features. However, this method failed to reach the desired image retrieval limit.

Medina *et al.* [30] have introduced an approach for medical image retrieval with the help of a fuzzy object-relational database management system (FORD-BMS). The system is used to store medical images and details for image content such as absence or presence of desired pathology indicators. This method allows to retrieve medical images based on this indicators making it possible to extract images from patients with similar diagnosis. To illustrate the capabilities of the FORDBMS, their paper focussed primarily on X-ray images of patients suffering from scoliosis. The major drawback of their study was scoliosis description. A set of images was retrieved by analysing the query image. Then the retrieved images were manually verified by the medical experts.

Yildizer *et al.* [31] have presented an approach to facilitate retrieving relevant images, which used the well-known clustering algorithm k-means. Their study employed cluster validity indexes combined with the majority voting to verify the appropriate number of clusters and predicted similar images; also, it considered images from nearby clusters. It used parameters named cS and cG to predict the distant range to be predicted for each cluster [32]. However, results indicated that the help of data mining techniques might decrease the efficiency and accuracy of the CBIR task. Based on the literature, visual feature for a different system based on the requirements of the user was selected. In addition, feature representation is one more necessity for any retrieval image system. To make a uniquely robust and feature based on representation, expensive computation and low-level feature fusion are necessary for the most valuable outcome. Thus, the undesired selection of features lowered the performance of the image retrieval system [33]. To enhance CBIR performance, an image feature vector might be used as input for deep learning and machine learning methods. In addition, it could be implemented associated with testing and training framework for unsupervised and supervised scenarios [34]. In recent years, studies have been focused on deep belief neural network (DBNN) for image retrieval to produce better output at a high computational cost.

3. Problem statement and motivation of the research

In the digital era, multimedia information plays a vital role in applications used in entertainment, education, e-commerce, medicine, and aerospace. With the rapid development of the internet, vast multimedia content is accessible to users. Hence, many digital images are created, and if it is examined periodically, there are tons of useful information offered to the users. With the proliferation of multimedia databases, the practicality of such large information is based on

how well it is browsed, retrieved and how useful is the obtained knowledge while minimizing the searching time [35]. Based on this purpose, CBIR has been developed for a long time and it is considered to fulfil these aims. The main aim of the CBIR approach is to retrieve images that obtain a maximum similarity score by comparing both the database and the query image. The modest method followed by the CBIR technique in extracting local and global feature sets is to ease image representation. The retrieval scheme followed here is entirely based on the distance calculation of feature vectors between query and database images. This, in turn, returns the most relevant image that is closest to the query image and provides a pair of images. Even though the problematic design appears to be modest to resolve, it poses various issues, such as images carrying different illumination, resolution, viewpoint criteria, etc. In addition to this, the occurrence of distractors or background features such as trees, cars, and people may produce challenges for the algorithm when retrieving the appropriate images because the computer-based automated design does not depend on the experience and knowledge humans possess. This makes the retrieval process to fail in obtaining the relevant image [36]. These issues are very complicated in distributed applications, as many cameras can produce a different image of the same scene in terms of translation, viewpoint, scaling and illumination. Nevertheless, advancements in the CBIR system can overcome these challenges and offer widespread applications in electronic guides for tourists, geo-localization at fine-grained level and in augmenting reality. Thus, it is essential to design a retrieval system with better accuracy to speed up retrieval of images and to respond within a promising timing period.

However, it is not always helpful to acquire both performances for various causes with respect to the specific dataset challenge. Hence, the only permitted final target is to identify a good trade-off between the two requirements. Also, due to a lack of proper extraction techniques, the users are unable to extract the relevant information from the available databases. The above drawbacks have motivated the authors to conduct the research work in this area.

4. Design of the proposed methodology

All over the world, the current trend of research area focuses on image mining approaches. Hence, mining image data is one of the indispensable approaches in the present scenario because of the analysis of image data in many fields, for example, construction engineering, business marketing, web publication, hospital surgery, etc. The main research area in the image mining technique is CBIR, which performs the retrieval under the similarity term accompanied by the extracted features. The complexity of the retrieval algorithm depends on the retrieval time. Yet, most of the scholars have not given much attention to study

retrieval time. In addition, dealing with colour images and extracting features have some drawbacks such as large dimensions, time-consuming computations, and sensitivity to noise interference. In order to overcome these issues, in this paper, a novel methodology is proposed by incorporating several steps like pre-processing, feature extraction, feature fusion, classification of classes, and feature matching retrieval system. The schematic diagram of the proposed retrieval system is shown in Fig. 1.
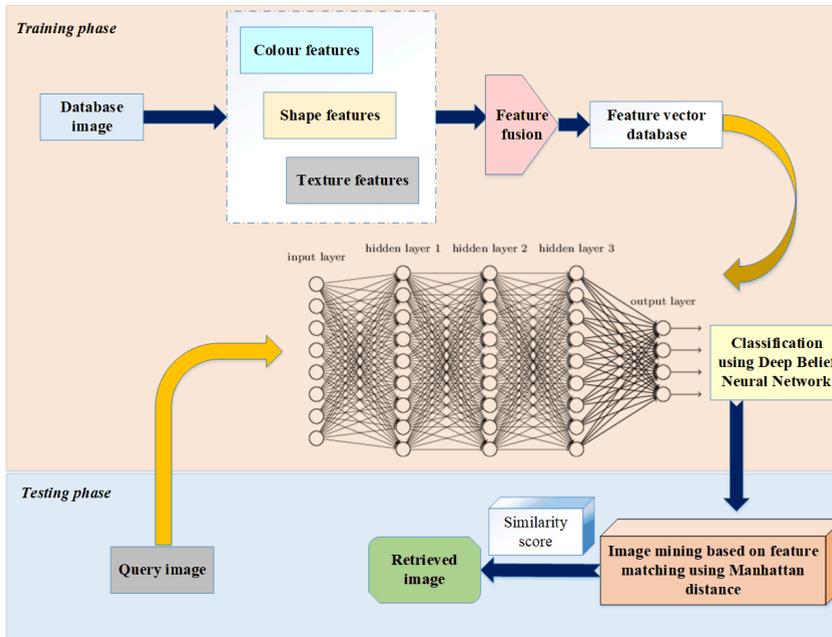

Fig. 1. Overall architecture of the proposed methodology.

The main contributions of this approach area high-performance CBIR model using IMDBN and feature matching process using Manhattan distance. The proposed technique helps to improve user interaction with image retrieval systems by fully exploiting the similarity information.

4.1. Noise removal using a median filter

The initial step of the proposed methodology is to eradicate noisy contents contained in an original image before subjecting it to the proposed image retrieval system. For removing the noisy contents, the primary task to be performed in the proposed research is resizing. These two are tasks involved in a typical pre-processing procedure. Image resizing aims to carry out all the dataset images of similar size to decrease computational burden and retrieval time. The performance metric retrieval time is calculated based on the time taken to retrieve

a set of relevant database images for the query input given. After performing resizing, the image size taken for analysis is 256×256. Then for removing noises, the resized image is applied with the median filter by substituting the median filter value of a pixel by a median of all pixels in adjacent values. The following equation mathematically formulates it

$$J(y, z) = \text{median} \left\{ J\left(y, z\right) \right\}, \tag{1}$$

where $(y, z) \in (1, 2, ..., \text{height}) \times (1, 2, ..., \text{width})$ and $J(y, z)$ determines the outcome of the filtered image. This improves the quality of an image by preserving edges, which improves so system performance improves.

4.2. CST-based image feature extraction technique

One of the significant processes in an image retrieval system is the feature extraction approach. In terms of extracting colour features, primarily colour space must be established. The term colour space denotes the representation of a set of colours and can be generated from RGB colour information. In addition to RGB colour space, an HSV transformation is performed to improve the image retrieval rate. Hue influences the chrominance of an image, saturation defines the predominance of a specific hue in colour space, and value represents the intensity of an image. These combinations are formed by means of primary colours incorporated. With respect to this, the HSV transformation is mathematically expressed by the following equations:

$$\text{Hue} = \cos^{-1} \frac{1/2 \left[(\text{Red} - \text{Green}) + (\text{Red} - \text{Blue}) \right]}{\sqrt{(\text{Red} - \text{Green})^2 + (\text{Red} - \text{Blue})(\text{Green} - \text{Blue})}}, \tag{2}$$

$$\text{Sauration} = 1 - \frac{3 \left[\min(\text{Red, Green, Blue}) \right]}{\text{Red} + \text{Green} + \text{Blue}}, \tag{3}$$

$$\text{Value} = \left[\frac{\text{Red} + \text{Green} + \text{Blue}}{3} \right]. \tag{4}$$

Typically, HSV is in the form of cylindrical geometry, whereas the angular dimension of hue starts at 0 deg, then the angular dimension passes 120 deg to the green region and then to blue with 240 deg. Then it again reaches the red region at 360 deg.

Step 1: After the transformation gets over, colour moments are evaluated for each colour space image to extract colour features. The first moment (mean) describes the image brightness and is calculated from the average intensity of all histogram images taken. Similarly, the second moment (standard deviation)

of an image evaluates the intensity distribution value about the mean in all the histogram images taken. The third moment (skewness) indicates the calculation of intensity measure of all histogram images taken with the unequal distribution. The mathematical representation of those colour moments is described as follows:

$$\mu_j = \frac{1}{O} \sum_{k=1}^{O} g_{jk}, \tag{5}$$

$$\sigma_j = \left(\frac{1}{O} \sum_{k=1}^{O} (g_{jk} - \mu_j)^2 \right)^{1/2}, \tag{6}$$

$$t_j = \left(\frac{1}{O} \sum_{k=1}^{O} (g_{jk} - \mu_j)^3 \right)^{1/3}, \tag{7}$$

where the term g_{jk} represents the measure of j-th colour component of an image k and the term O denotes the count of pixels carried by the image, μ_j indicates mean, σ_j symbolizes standard deviation and t_j represents the term skewness.

Step 2: In this second stage, shape features of the proposed CST feature extraction method are defined, which helps to highlights the object present in the given image. Here, Hu moments are extracted for shape feature description and are obtained from the geometric characteristic of a given image. The geometric characteristic is represented by the weighted average pixel intensity of images. The intensity of an image pixel with orientation (y, z) is indicated by $J_{(y,z)}$. In general, the moment of an image is evaluated by establishing the total amount of intensity values associated with the image pixel and is mathematically expressed as follows:

$$Mi = \sum_{y} \sum_{z} J_{(y,z)}, \tag{8}$$

where the term Mi denotes the moment of an image. The above equation defines the weighted value of the pixel that is generated based on intensity measure and irrespective of imaging location. From this, the above equation is remodeled by considering the intensity value of the pixel and the location of the image. It is mathematically given as follows:

$$Mi_{j,k} = \sum_{j} \sum_{k} y^j z^k J_{(y,z)}, \tag{9}$$

where j, k term ranges from 0 to infinity. In Hu moment shape descriptor, initially, the central moment of an image is calculated and is represented by μ_{jk}.

The central moment of an image is attained by deducting the centroid value $(\overline{y}, \overline{z})$ from (y, z) in moment calculation. It is mathematically given by the following equation:

$$\mu_{jk} = \sum_j \sum_k (y - \overline{y})^j (z - \overline{z})^k J_{(y,z)}, \tag{10}$$

where

$$\overline{y} = \frac{Mi_{10} \text{ (Summation of } y \text{ coordinate of image pixel)}}{Mi_{00} \text{ (Image Area)}}, \tag{11}$$

$$\overline{z} = \frac{Mi_{01} \text{ (Summation of } z \text{ coordinate of image pixel)}}{Mi_{00} \text{ (Image Area)}}. \tag{12}$$

The above-calculated feature extraction-based moments are invariants to translation, scale, and rotation. Likewise, central moments are invariant to the position. To make the moment that is invariant to the scaling factor, central moments are normalized by the following equation:

$$\eta_{jk} = \frac{\mu_{jk}}{\mu_{00}^{\left(1 + \frac{j+k}{2}\right)}}. \tag{13}$$

The proposed shape feature-based Hu moments combine seven sets of moments that are a non-linear arrangement of normalized central moments (η_{jk}). The seven Hu moments are mathematically solved by the subsequent equations:

$$Hu_1 = \eta_{20} + \eta_{02}, \tag{14}$$

$$Hu_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \tag{15}$$

$$Hu_3 = (\eta_{30} - \eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2, \tag{16}$$

$$Hu_4 = (\eta_{30} - \eta_{12})^2 + (\eta_{03} + 3\eta_{21})^2, \tag{17}$$

$$Hu_5 = (3\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right]$$
$$+ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right], \tag{18}$$

$$Hu_6 = (\eta_{20} - \eta_{02})^2 \left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$
$$+ 4\eta_{11}(\eta_{30} + \eta_{12}) + (\eta_{21} + \eta_{03}), \tag{19}$$

$$Hu_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) \left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$
$$+ (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03}) \left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]. \tag{20}$$

The expressed seven moments are invariants to translation, scale, reflection, and rotation.

Step 3: In the third step of the CST feature extraction approach, some of the significant texture grey level co-occurrence matrix (GLCM) features are extracted, namely contrast, inverse difference moments, dissimilarity, energy, correlation, and angular second moment. These texture features help to evaluate the spatial relationship among pixels in an image by extracting texture information. Here the total number of columns and rows in an image is similar to the total number of grey levels in an image. In four different angular orientations, such as 0 deg, 45 deg, 90 deg, and 135 deg, the co-occurrence matrix is constructed. This texture feature extraction technique is in the form of a square matrix possessing the dimension of M_h, where M_h relates the number of grey levels in an image. The matrix elements are generated by adding the occurrence of a pixel that attained a value j adjacent to a pixel measure with k, then distributing the entire matrix by the total amount of comparisons created. Thus, every tuple is considered a probability with a pixel of value j adjacent to a pixel of value k.

The term contrast determines the variation in luminance over a whole image among a pixel and neighboring pixel. The contrast attains zero for the constant image with no variation. The mathematical derivation is offered as follows

$$G_{\text{con}} = \sum_{j=1}^{M_h} \sum_{k=1}^{M_h} q_{j,k}(j-k)^2, \tag{21}$$

where $q(j,k)$ represents the joint probability distribution of pixel pair over an image with grey levels (j,k). The next texture feature to be extracted is the inverse difference moment. It defines the closeness of the element distributed in GLCM to its diagonal. The value obtained by this texture feature will be higher at the time the local greyscale is uniform. Due to its weight value, this feature is very much influenced by image local homogeneity. The inverse difference moment texture feature is otherwise known as homogeneity

$$G_{\text{idm}} = \sum_{j=1}^{M_h} \sum_{k=1}^{M_h} \frac{1}{1+(j-k)^2} q(j,k). \tag{22}$$

The texture feature dissimilarity defines the difference between grey-level measures of pixels over the whole image. The dissimilarity texture function of GLCM is mathematically expressed as follows

$$G_{\text{dis}} = \sum_{j=1}^{M_h} \sum_{k=1}^{M_h} q_{j,k}(j-k). \tag{23}$$

The sum of the squared term of elements in GLCM is obtained by texture feature energy. It is otherwise known to be angular second moment or uniformity. It attains a higher measure when all the pixels possess a similar value or has a uniform texture. It is mathematically represented by the following equation:

$$G_{\text{ene}} = \sum_{j=1}^{M_h} \sum_{k=1}^{M_h} q(j,k)^2. \tag{24}$$

The texture feature correlation determines the linear dependency measure of the grey level of neighboring pixels. It occurs as either -1 or 1 for a perfectly negative or positive image that is correlated and is infinite for the constant image

$$G_{\text{cor}} = \frac{\sum_{j=1}^{M_h} \sum_{k=1}^{M_h} (j,k)q(j,k) - \mu_j \mu_k}{\sigma_j \sigma_k}. \tag{25}$$

After computing the CST feature set, the values are fused together to form a feature vector F_v and are denoted in the form of $F_v = \{\mu_j, \sigma_j, t_j, Hu_1, ..., Hu_7, G_{\text{con}}, G_{\text{idm}}, G_{\text{dis}}, G_{\text{ene}}, G_{\text{cor}}\}$. The constructed feature vector F_v of all datasets and images are stored in a feature database F_{db}. This set of features is unique for different images; hence, it helps to recognize the relevant image in the classification task.

4.3. Proposed image mining technique based on deep belief network

For performing the classification task in the proposed CBIR system, the deep belief network (DBN) model is used to classify different image classes. The proposed classification approach is an energy-based probability generation model and is stacked by several restricted Boltzmann machines (RBM). It carries two phases, namely training and testing, where it is executed in a layer-by-layer manner. Here the layers incorporated in DBN are trained over the group of pre-trained RBM layers. It contains two layers: a visible layer (input) and a hidden layer carrying neurons. Both of them are connected by edges, whereas neurons in the same layer are restricted. The visible layer enclosed here is represented as

$$vis = \{vis_1, vis_2, ..., vis_j, ..., vis_m\}; \quad (vis_j \in \{0, 1\}). \tag{26}$$

Likewise, any one of the hidden layers employed in DBN is expressed as

$$hid = \{hid_1, hid_2, ..., hid_j, ..., hid_m\}; \quad (hid_j \in \{0, 1\}), \tag{27}$$

where the visible layer is indicated as vis, and the hidden layer is represented as hid. The count of the visible layer is defined based on the input feature vector

$F_v = \{\mu_j, \sigma_j, t_j, Hu_1, ..., Hu_7, G_{con}, G_{idm}, G_{dis}, G_{ene}, G_{cor}\}$ that is extracted from the proposed CST based feature extraction technique. These CST features are subjected to a visible layer and can be further transmitted to the hidden layer. Both the layers are connected to each other via weight connection, and neurons of each layer are not connected to each other. In this architecture, three parameters are defined initially and are denoted as $\theta = \{W, b, c\}$, where weight matrix is denoted by W, the bias of the hidden layer and the visible layer is indicated by B and C, respectively.

The proposed RBM framework has n number of hidden neurons and also possesses m number of input neurons, vis_j denotes a visible unit of j-th layer, and hid_j symbolizes a hidden unit of j-th layer. The structure of parameters is mathematically given in the following equation:

$$W = \{w_{j,k} \in S^{m \times n}\}, \tag{28}$$

where $w_{j,k}$ denotes the weight value of visible and hidden neurons in j-th layer and k-th layer. The bias function of the hidden layer B is mathematically expressed as follows:

$$B = \{b_j \in S^n\}. \tag{29}$$

In the above equation, b_j the term denotes threshold measure of bias function of the hidden layer with j-th hidden neuron. Similarly, for the visible layer, the bias function is determined as follows:

$$C = \{c_k \in S^m\}. \tag{30}$$

In the above equation, the term c_k denotes threshold measure of bias function of the visible layer with k-th visible neuron. With the help of the energy function in RBM, the probability function is learned among hidden and visible layers. For achieving the probability distribution, the energy function is mathematically derived by the following equation:

$$E\left(vis, hid|\theta\right) = -\sum_{j=1}^{m} b_j vis_j - \sum_{k=1}^{n} c_k hid_k - \sum_{j=1}^{m} \sum_{k=1}^{n} vis_j W_{jk} hid_k, \tag{31}$$

where $\theta = \{W_{jk}, b_j, c_k\}$ are the parameters of the RBM model, and E term defines the energy function among every hidden and visible layer node. Likewise, m and n indicate the total number of visible and hidden neurons, correspondingly.

With the exponential and regularisation of the energy function, the probability function tends to be defined in further processes. The joint probability

distribution of the RBM model among visible and hidden layers can be mathematically expressed as follows:

$$P\left(vis, hid|\theta\right) = \frac{1}{Z\left(\theta\right)}e^{-E\left(vis, hid|\theta\right)}. \tag{32}$$

The above equation is based on the Gibbs distribution function of the RBM model. The partition function is derived based on the following equation:

$$Z(\theta) = \sum_{vis,hid} e^{-E\left(vis,hid|\theta\right)}, \tag{33}$$

where the $Z(\theta)$ term stands for a distributed or normalized function that symbolizes the state of energy condition of every visible and hidden layer. It is generated by summing the energy evaluation attained by both visible and hidden layers. For acquiring the parameter values, the probability function is determined. For this purpose, the joint distribution of both visible and hidden layers is formulated primarily by the notation $P\left(vis, hid|\theta\right)$. Then the marginal distribution function of the visible layer is calculated using $P\left(vis|\theta\right)$ and can be mathematically described as follows:

$$P\left(vis|\theta\right) = \frac{1}{Z\left(\theta\right)} \sum_{hid} e^{-E\left(vis,hid|\theta\right)}. \tag{34}$$

The above marginal distribution function of the visible layer is evaluated by performing a summation operation of the overall network criteria of the hidden layer. Similarly, the marginal distribution function of the hidden layer is calculated by the following equation:

$$P\left(hid|\theta\right) = \frac{1}{Z\left(\theta\right)} \sum_{vis} e^{-E\left(vis,hid|\theta\right)}. \tag{35}$$

As per RBM architecture, it possesses both inter-layer and layer-by-layer connections because of its exponential formation. Also, both the visible and hidden layers of RBM are independent in nature; hence its conditional probability values are determined by the following equations:

$$P\left(vis|hid\right) = \prod_{j} P\left(vis_j|hid\right), \tag{36}$$

$$P\left(hid|vis\right) = \prod_{k} P\left(hid_k|vis\right). \tag{37}$$

The proposed RBM structure is comprised of a binary state of elements; hence, its activation function to be utilized is of the sigmoid term, which is given as follows:

$$\text{sigmoid}(y) = \frac{1}{(1 + e^{-y})}. \tag{38}$$

Based on the above activation function, the probability of both visible and hidden layer activation function in RBM structure is formulated and is given as follows:

$$\text{sigmoid}(y) = \frac{1}{(1 + e^{-y})}, \tag{39}$$

$$P\left(vis_j = 1 | hid\right) = \frac{1}{1 + \exp\left(-b_j - \sum\limits_{k=1}^{n} w_{jk} hid_k\right)}, \tag{40}$$

$$P\left(hid_k = 1 | vis\right) = \frac{1}{1 + \exp\left(-c_k - \sum\limits_{j=1}^{m} w_{jk} vis_j\right)}. \tag{41}$$

The further process to be carried out in the proposed classification framework is to update the rules for the involved parameters $\theta = \{W, b, c\}$. In some situations, the Gibbs distribution function of the RBM model will not be implemented in minimum time. Therefore, to minimize time consumption, a fast-learning algorithm named contrast divergence (CD) is proposed in such circumstances. The training process is conducted stage by stage for RBN networks. Initially, inputs from visible layers are trained based on the hidden layer h1. This data is biased with biasing values and weights in the first stage. The second stage of RBM is performed based on the previous stage outcome. Similarly, a sequential training process will make the high-level classification. Based on the above learning method, the parameter values are updated and are mathematically offered in the equations below

$$W^{\text{time}+1} = W^{\text{time}} + \varepsilon\left(P\left(hid | vis^{(0)}\right)\left[vis^{(0)}\right]^{\text{time}}\right.$$
$$\left. - P\left(hid | vis^{(1)}\right)\left[vis^{(1)}\right]^{\text{time}}\right), \tag{42}$$

$$b^{\text{time}+1} = b^{\text{time}} + \varepsilon\left(vis^{(0)} - vis^{(1)}\right), \tag{43}$$

$$c^{\text{time}+1} = c^{\text{time}} + \varepsilon\left(P\left(hid | vis^{(0)}\right) - P\left(vis^{(1)}\right)\right), \tag{44}$$

where "time" indicates step time with iteration, and ε denotes the learning rate. The learning rate of the network is set to 0.001 rate with 100 epochs. The above steps are repeated until the parameter values are updated to form a more

abstract representation that is more abstract and has more represent ability than the lower layer. With the help of the above-stated equations, the training algorithm RBM achieved the feature extraction purpose. In the proposed DBN framework, the training method RBM can avoid complex functionalities carried out by the overall training process of DBN. DBNN is a structural model with many in-built layers called RBMs. A training method, namely the CT method, is used to train this proposed DBNN model. This method consists of three steps. In the initial step, the first RBM is pre-trained by the use of attributes. In the next stage, when the parameters from the first RBM were obtained, the output layer of the first RBM is considered as the input layer for the second one. This process gets iterated until training is finished. Hence, a higher level of features is obtained by the whole process, which is helpful in improving the performance of the retrieval system. By this classification, the task is accomplished for the training phase. After training, the classified class labels can be attained in the testing phase with the help of the testing dataset.

4.4. Feature matching using Manhattan distance

In the proposed image retrieval system, when the query content is created, the same procedure mentioned above is followed to generate the feature vector. After producing a feature vector for the query image, the similarity score value is evaluated using Manhattan distance measurement and compared feature values of both database images and query images. This, in turn, returns the relevant image as per the user query. Moreover, the smaller the difference, the more similar are both the query and database images. In other words, the feature vector of images retains little distance that is most similar to query ones. The mathematical representation of Manhattan distance for the feature matching process is represented as follows:

$$d_M = \sum_{j=1}^{m} |Q_j - T_j|, \tag{45}$$

where d_M denotes the similarity calculation of Manhattan distance and m denotes the length of the feature vector. This is computed by taking the total amount of absolute variation among feature vectors of the query image Q_j and database training image T_j. The reason behind choosing the Manhattan distance metric is that it displays robustness to outliers. Based on Eq. (45), database images possessing minimum distance values are retrieved and visualized as similar images. This is how the matching function is evaluated to obtain a similarity score among training and testing images. Finally, the model obtained is applied to classify image data for retrieving relevant image labels. The algorithmic flow of the proposed methodology is illustrated in Algorithm 1.

Algorithm 1. Pseudo-code of the proposed methodology.

Input: Input image $\boldsymbol{T_j}$ and Query image $\boldsymbol{Q_j}$;
Output (C_m): Image result based on proposed IMDBN;
Get the image
//**Noise removal**//
   ❖ *Resize the input image to $\boldsymbol{256 \times 256}$;*
   ❖ *Noise removal using median filter using Eq. (1);*
//**CST feature extraction**//
   ❖ *Perform from Eq. (2) to (25);*
//**Create IMDBN structure**//
Input set of feature vectors obtained;
Fix maximum number of iteration;
Set weights and bias functions;
Train IMDBN with its corresponding layers;
   //**Create DBN based RBM training model**//
   do
   {
   Procedure
   for all hidden units $\boldsymbol{k}$
   calculate $\boldsymbol{P(hid_k = 1|vis)}$ using Eq. (41);
   sample $\boldsymbol{hid_k \in 0,1}$ from $\boldsymbol{P(hid_k = 1|vis)}$;
   end for
   for all visible units $\boldsymbol{j}$
   calculate $\boldsymbol{P(vis_j = 1|hid)}$ using Eq. (40);
   sample $\boldsymbol{vis_j \in 0,1}$ from $\boldsymbol{P(vis_j = 1|hid)}$;
   end for
   Procedure classify LABEL (Image);
   end Procedure
   }

5. Experimental results and discussions

This section presents the experimental outcome of the proposed image mining approach-based CBIR system. For performing image mining task, a deep learning technique named the image mining technique based on deep belief network (IMDBN) is proposed. To show the superiority of the proposed methodology, the experimentation is carried out, and its implementation is done in the working platform of Python 3.7 processor – Intel i5 with 16GB RAM and 4GB graphics processor. For GPU analysis, the input images are tested on an i3 processor with 4GB RAM and NVIDIA GeForce GTX 1650. Here 80% of the dataset is used

for the training phase, and the remaining 20% of images are used for the testing phase. The proposed method is validated, and tested outcomes are compared with the existing techniques in terms of accuracy, precision, recall, F-measure, and retrieval time. IMDBN model is then compared with previous methods such as long short-term memory (LSTM) networks, multilayer perceptron (MLP), radial basis network (RBN), DBN and neural network (NN) to show that the proposed model is superior to these methods.

5.1. Dataset description

To show the validation of the proposed method, extensive experimentation is done on a dataset [31] that contains 1355 photographic images with 18 classes. The images contained in the dataset are of different pixel size and are in JPEG format. This larger data source is currently available on the public website. Some of the sample images are displayed in Fig. 2.
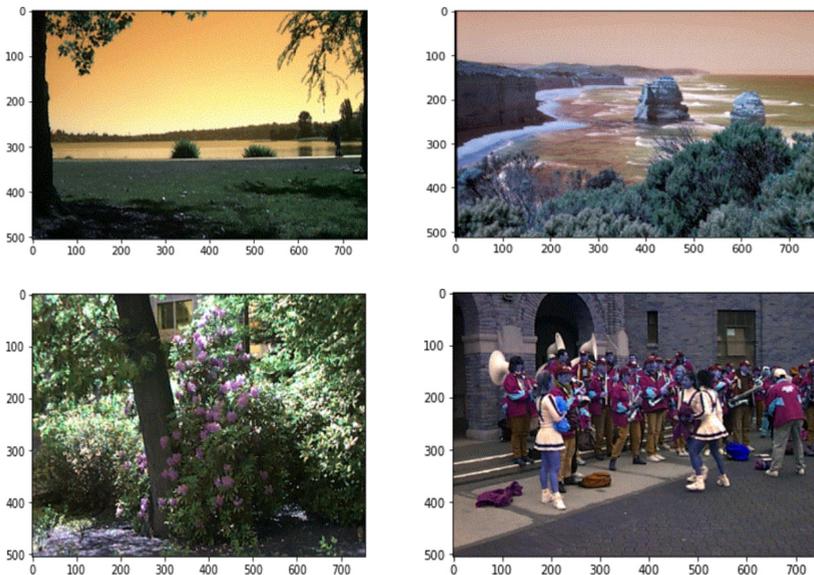
FIG. 2. Sample input images.

The above figure shows the sample input images taken to analyze the proposed CBIR system. For achieving the proposed retrieval system, several steps are included to obtain a better retrieval model. The steps involve: resizing, filtering, feature extraction, classification, and matching. All the functions are done with an effective approach to perform the proposed system better. The outcome of images obtained by two stages, namely resizing and filtering, is displayed in Table 1.
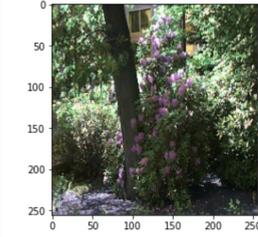
TABLE 1. Outcome of different images.
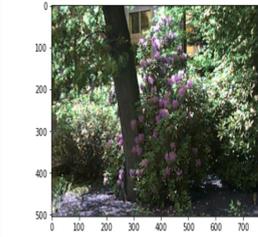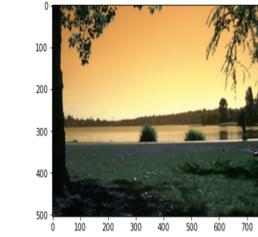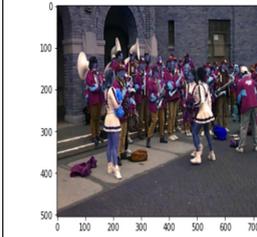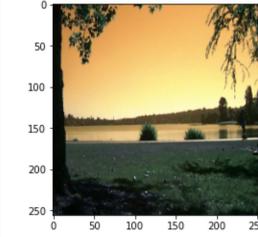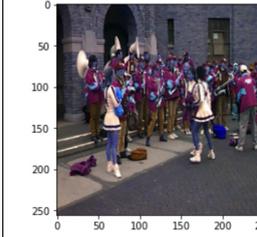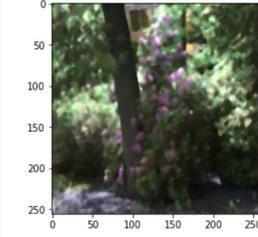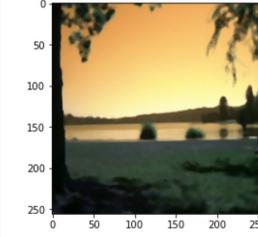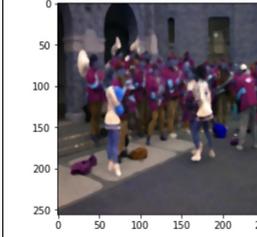
Table 1 shows the outcome of resized and filtered image labels such as Australia, cherries, football, green lake, spring flowers. The further section describes the evaluation metrics taken for analysis and offers the outcome obtained by the proposed and existing technique.

5.2. Evaluation metrics

The efficiency of the proposed image retrieval system is evaluated based on performance attained by feature extraction, classification rate, and similarity measurement. In this sub-section, some of the major evaluation metrics such as accuracy, precision, recall, and F-measure are adopted not only to validate the effectiveness of the proposed methodology but also to show the stability of the results. To demonstrate the effectiveness of the proposed system clearly,

different test metrics: accuracy, precision, F1-score, and recall are evaluated. The mathematical expressions are illustrated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}, \tag{46}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{47}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{48}$$

$$F1 - \text{Score} = \frac{2 \times (\text{precision} \times \text{recall})}{\text{precision} + \text{recall}}. \tag{49}$$

Based on the above mathematical notation, the performance of the proposed and existing techniques are evaluated. Therefore, its performance analysis is offered in the subsequent sections.

5.3. Performance analysis

After employing the proposed IMDBN method in the retrieval system, its obtained accuracy and other measure values are graphically represented in the following figures to show the clarity of the proposed methodology. Comparisons of these previous methods with respect to performance evaluation parameters are listed in the table below.

Table 2 shows the measurement performed for IMDBN and other previous methods with respect to sensitivity, specificity, error and kappa. It has been countably efficient for the IMDBN model, other NN such as DBN, MLP, RBN and NN performed slower than this model; the proposed one outperforms them due to the heavy performance at the output side.

TABLE 2. Comparison results.

Method	Type of network	Sensitivity	Specificity	Error	Kappa
Proposed method	*IMDBN*	*71.06*	*93.82*	*0.2*	*7.5*
Liang *et al.* [38]	LSTM	53.88	85.65	0.10	3.2
Kaur and Singh [40]	DBN	67.00	91.00	0.6	2
Antani *et al.* [39]	MLP	62.59	86.12	4.1	4.3
Loboda *et al.* [37]	RBN	47.06	76.88	2.8	4.8
Hussain *et al.* [41]	NN	69.65	91.12	2.1	5

5.3.1. Accuracy. Accuracy term is defined as the number of images retrieved correctly. When the accuracy metric attained is higher, i.e., nearer to 100, then

the retrieval system achieves the best performance and shows the system is an appropriate method for retrieving a relevant image set. The acquired accuracy value of both proposed and existing techniques is graphically shown in Fig. 3.
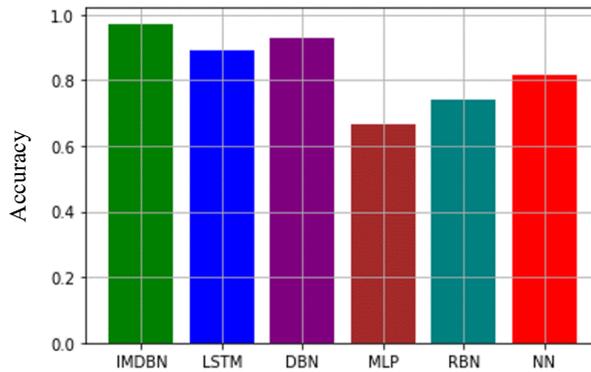

F IG. 3. Obtained accuracy for proposed and existing algorithms.

In Fig. 3, it is clearly visible that the accuracy obtained for the proposed model is 94.4%. By contrast, the existing techniques such as LSTM, DBN, MLP, RBN, and NN attained lesser accuracy measures, which is not up to the desired level for an appropriate retrieval. In this comparison result, it is evident that IMDBN attained high accuracy due to its data augmentation technique and was trained using the CD technique. LSTM exhibited 87% as it was less efficient when it comes to image retrieval. DBN, without appropriate preprocessing and training, was unable to attain the required accuracy. MLP and RBN respectively attained 65% and 70% accuracy levels, which was insufficient for retrieval applications. Hence the proposed framework achieves higher classification measures than existing approaches. Likewise, other measures like F-measure, precision, and recall are evaluated for the CBIR system, and this is graphically shown in the graphs below.

5.3.2. Precision. Measurement of retrieved relevant images in the query of the total retrieved image is termed precision. The graphical outcome of precision measure is depicted in Fig. 4.

Figure 4 examines the outcome measure of the proposed IMDBN and existing LSTM, DBN, MLP, RBN, and NN techniques. 0.9152 is obtained by the proposed methodology, whereas the precision measure for LSTM is 0.623, traditional DBN is 0.761, MLP is 0.701, RBN is 0.482, and NN is 0.76. Due to high true positive values obtained at the result, it was possible to obtain a high precision value using the IMDBN model, as it gives 98 images positively correct at the output, thus improving the precision rate. By contrast, other methods suffer to maintain precision values due to their smaller accuracy range than the proposed model.
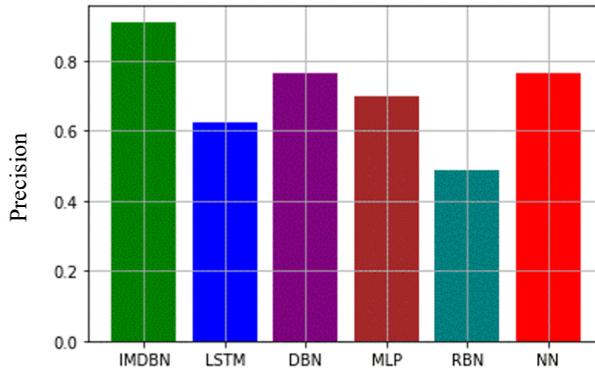
Fɪɢ. 4. Precision comparison in graphical representation.

From this, it is clear that the proposed methodology attains a maximum outcome compared to the existing techniques. This shows that the proposed retrieval system performs better in terms of measuring precision value.

5.3.3. *Recall.* The recall in image retrieval can be defined as the measurement of retrieved relevant images to the total database images. The graphical representation of the recall measure obtained is plotted in Fig. 5.
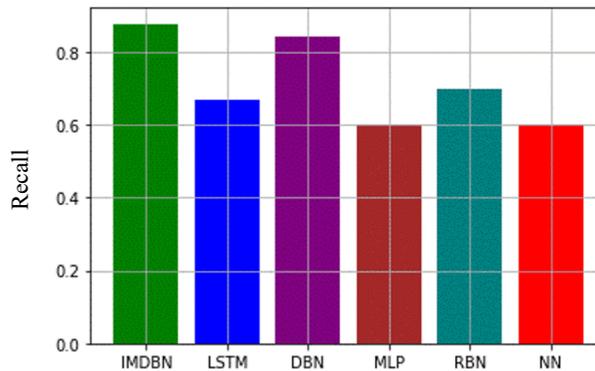

Fɪɢ. 5. Overall recall acquired by proposed and existing approaches.

Figure 5 deliberates the outcome of the recall measure obtained by both the proposed and existing approaches. For instance, recall obtained for the proposed IMDBN methodology is 0.8791, which is comparatively higher than the existing LSTM, DBN, MLP, RBN, and NN classifiers attaining 0.67, 0.84, 0.76, 0.76, and 0.60, respectively. Due to fewer false negative values obtained at the result side, a model efficiently improves its recall values. Recall values are improved only when there is a proper retrieval process taking place. The IMDBN model using deep learning approach have appropriately improved recall rate. LSTM and DBN

models still require improvement for recall rate and MLP suffers due to lack of training data and overfitting. This shows that the proposed methodology attains the maximum outcome and shows outstanding retrieval performance.

5.3.4. F-measure. Combined measurement of precision and recall helps to show the efficiency of the system via F-measure. F-measure is measured for proposed and previous models and compared in the form of the graph shown in Fig. 6.
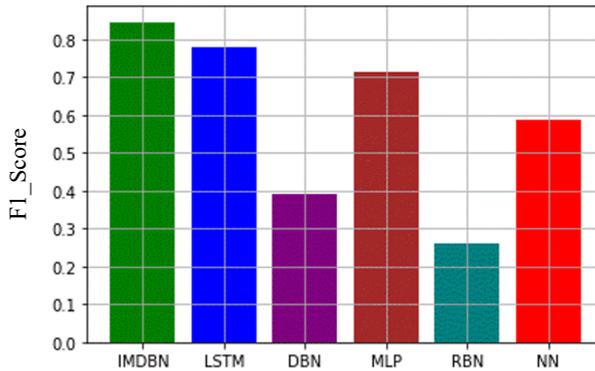

FIG. 6. F1-measure analysis for proposed and existing techniques.

In Fig. 6, it is visible that F-measure is as high as expected for the IMDBN model. Furthermore, it is clearly seen that the previous methods such as LSTM gives 0.78 value, there is 0.4 for DBN, 0.7 for MLP network, 0.25 for RBN and 0.6 for NN model. The obtained F-measure function attained is higher for the proposed approach, it is nearly 0.87; hence, it shows a better outcome for the proposed retrieval system. Likewise, the calculated loss function is provided in the subsequent section.

5.3.5. Loss function. The error or loss function is calculated by taking the number of irrelevant images obtained for the query input accessed from the database images.

Due to the minimum error value attained by the proposed IMDBN, the proposed retrieval system has the capacity to generate the most appropriate images related to query and database images. It shows that the occurrence of the loss function is 0.056 for the proposed method; it is higher for existing models. Therefore, the proposed method leads to the lower number of less irrelevant images compared to the existing approaches.
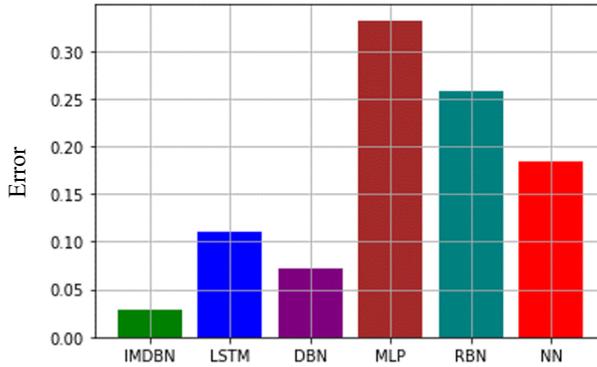
FIG. 7. Comparison plot for the loss function.

5.3.6. Retrieval time. The retrieval time is calculated based on the time taken for retrieving relevant images related to query and database information. It is measured in seconds, and the obtained values for retrieval time are graphically plotted in the graph in Fig. 8.
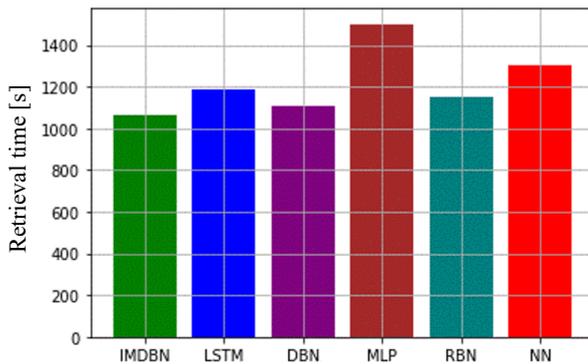

FIG. 8. Retrieval time acquired for proposed and existing techniques.

It is a well-known fact that when the retrieval time is smaller then the system performance automatically improves. In Fig. 8, analysis results showed that there was an improved performance for the IMDBN model with 1000 sec. LSTM shows 1200 sec, which was slightly higher than the IMDBN model and DBN exhibits 1100 sec due to its complex training stage and its testing stage taking a longer time than required. MLP and RBN techniques show 1500 sec and 1190 sec, respectively. NN is a previous technique that gives retrieval result at 1250 sec. By satisfying the required time for retrieval at 1010 sec, the proposed system outperforms well by attaining a minimum retrieval value. Thus, from the analysis, it is concluded that overall performance is improved enough for the proposed retrieval system.

For the dataset ICPR-2004 1355 photographic images with 18 classes, the accuracy rate was 94.4% without losing the time complexity and overfitting issues. Feature extraction seems to be efficient compared to previous algorithms as IMDBN starts to preprocess, extract features and classifies within minimum retrieval time compared to normal NNs. IMDBN provides a good classification as it could retrieve 98 images correctly, which was relatively higher than the previous ones discussed here. The proposed framework has been implemented and evaluated extensively using different parameters; thus, it will be more helpful when used for image retrieval applications.

6. CONCLUSION

The proposed technique introduces a novel approach named IMDBN for the classification of input images. The primary objective of the proposed method is to categorize relevant images in a retrieval system in an efficient manner. For achieving this objective, the proposed methodology involves phases like preprocessing, feature extraction, and classification. The preprocessing step carries several steps to eliminate noisy contents contained in the input image. The next step is to perform feature extraction, which is done to extract the most appropriate features. This helps to make the network model learn faster. Finally, classification is performed through the proposed IMDBN model; this, in turn, improves classification accuracy by achieving a wider range of features. Thus, the performance of the proposed method improved by reducing the computational complexity. The proposed methodology achieved an efficient outcome compared with the existing approaches in terms of accuracy, F-measure, precision, and recall.

REFERENCES

1. Q. Zheng, M. Yang., J. Yang, Q. Zhang, X. Zhang, Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process, *IEEE Access*, **6**: 15844–15869, 2018, doi: 10.1109/ACCESS.2018.2810849.

2. Q. Zheng, X. Tian, N. Jiang, M Yang, Layer-wise learning based stochastic gradient descent method for the optimization of deep convolutional neural network, *Journal of Intelligent & Fuzzy Systems*, **37**(4): 5641–5654, 2019, doi: 10.3233/JIFS-190861.

3. T.A. Ahmed, S. Ummesafi, A. Iqbal, Content based image retrieval using image features information fusion, *Information Fusion*, **51**: 76–99, doi: 10.1016/j.inffus.2018.11.004.

4. M.S. Haji, M.H. Alkawaz, A. Rehman, T. Saba, Content-based image retrieval: A deep look at features prospectus, *International Journal of Computational Vision and Robotics*, **9**(1): 14–38, 2019, doi: 10.1504/IJCVR.2019.098004.

5. Q. Zheng, X. Tian, M. Yang, H. Wang,Differential learning: a powerful tool for interactive content-based image retrieval, *Engineering Letters*, **27**(1): 202–215, 2019.

6. S.P. Rana, M. Dey, P. Siarry, Boosting content based image retrieval performance through integration of parametric & nonparametric approaches, *Journal of Visual Communication and Image Representation*, **58**: 205–219, 2019, doi: 10.1016/j.jvcir.2018.11.015.

7. U. Chaudhuri, B. Banerjee, A. Bhattacharya, Siamese graph convolutional network for content based remote sensing image retrieval, *Computer Vision and Image Understanding*, **184**: 22–30, 2019, doi: 10.1016/j.cviu.2019.04.004.

8. R.R. Saritha, V. Paul, P.G. Kumar, Content based image retrieval using deep learning process, *Cluster Computing*, **22**(2): 4187–4200, 2019, doi: 10.1007/s10586-018-1731-0.

9. U. Sharif, Z. Mehmood, T. Mahmood, M.A. Javid, A. Rehman, T. Saba, Scene analysis and search using local features and support vector machine for effective content-based image retrieval, *Artificial Intelligence Review*, **52**(2): 901–925, 2019, doi: 10.1007/s10462-018-9636-0.

10. A. Rashno, E. Rashno, Content-based image retrieval system with most relevant features among wavelet and color features, *arXiv preprint*, arXiv:1902.02059, **[eess.IV]**, 2019.

11. P. Tschandl, G. Argenziano, M. Razmara, J. Yap, Diagnostic accuracy of content-based dermatoscopic image retrieval with deep classification features, *British Journal of Dermatology*, **181**(1): 155–165, 2019.

12. A. Sarwar, Z. Mehmood, T. Saba, K.A. Qazi, A. Adnan, H. Jamal, A novel method for content-based image retrieval to improve the effectiveness of the bag-of-words model using a support vector machine, *Journal of Information Science*, **45**(1): 117–135, 2019, doi: 10.1177/0165551518782825.

13. Y.D. Mistry, Textural and color descriptor fusion for efficient content-based image retrieval algorithm, *Iran Journal of Computer Science*, **3**(3): 169–183, 2020, doi: 10.1007/s42044-020-00056-0.

14. R. Ashraf, M. Ahmed, U. Ahmad, M.A. Habib, S. Jabbar, K. Naseer, MDCBIR-MF: multimedia data for content-based image retrieval by using multiple features, *Multimedia Tools and Applications*, **79**(13): 8553–8579, 2020, doi: 10.1007/s11042-018-5961-1.

15. N.F. Haq, M. Moradi, Z.J. Wang, A deep community based approach for large scale content based X-ray image retrieval, *Medical Image Analysis*, **68**: 101847, doi: 10.1016/j.media.2020.101847.

16. A.R. Zubair, O.A. Alo, Content-based image retrieval system using second-order statistics, *International Journal of Computer Applications*, **176**(36), 9 pages, 2020, doi: 10.5120/ijca2020920475.

17. H.-H. Bu, N.-C. Kim, B.-J. Yun, S.-H. Kim, Content-based image retrieval using multi-resolution multi-direction filtering-based CLBP texture features and color autocorrelogram features, *Journal of Information Processing Systems*, **16**(4), 991–1000, 2020, doi: 10.3745/JIPS.02.0138.

18. G. Zhao, M. Zhang, J. Liu, Y. Li, J.-R. Wen, AP-GAN: Adversarial patch attack on content-based image retrieval systems, *Geoinformatica*, 1–31, 2020, doi: 10.1007/s10707-020-00418-7.

19. S. Singh, S. Batra, An efficient bi-layer content based image retrieval system, *Multimedia Tools & Applications*, **79**(25/26): 17731–17759, 2020.

20. G. Xie, B. Guo, Z. Huang, Y. Zheng, Y. Yan, Combination of dominant color descriptor and Hu moments in consistent zone for content based image retrieval, *IEEE Access*, **8**: 146284–146299, 2020, doi: 10.1109/ACCESS.2020.3015285.

21. P. Kavitha, P.V. Saraswathi, Segmentation for content based satellite image retrieval using fuzzy clustering, *International Journal of Advanced Science and Technology*, **29**(9s): 3042–3049, 2020.

22. G. Rudrappa, N. Vijapur, Cloud classification using K-means clustering and content based image retrieval technique, [in]: *2020 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0700–0704, 2020, doi: 10.1109/ICCSP48568.2020.9182211.

23. X.-Y. Wang, B.-B. Zhang, H.-Y. Yang, Content-based image retrieval by integrating color and texture features, *Multimedia Tools and Applications*, **68**(3): 545–569, 2014.

24. T. Kato, Database architecture for content-based image retrieval, [in:] *Image storage and retrieval systems*, International Society for Optics and Photonics, Vol. 1662, pp. 112–123, 1992.

25. H.-H. Bu, N.-C. Kim, C.-J. Moon, J.-H. Kim, Content-based image retrieval using combined color and texture features extracted by multi-resolution multi-direction filtering, *Journal of Information Processing Systems*, **13**(3): 464–475, 2017, doi: 10.3745/JIPS.02.0060.

26. J.J. Thiagarajan, K.N. Ramamurthy, P. Sattigeri, A. Spanias, Supervised local sparse coding of sub-image features for image retrieval, [in:] *2012 19th IEEE International Conference on Image Processing*, pp. 3117–3120, 2012, doi: 10.1109/ICIP.2012.6467560.

27. A. Latif, A. Rasheed, U. Sajid, J. Ahmed, N. Ali, N.I. Ratyal, Z. Bushra, S.H. Dar, M. Sajid, T. Khalil, Content-based image retrieval and feature extraction: a comprehensive review, *Mathematical Problems in Engineering*, **2019**, Article ID 9658350, 21 pages, 2019, doi: 10.1155/2019/9658350.

28. Y. Hou, Q. Wang, Research and improvement of content-based image retrieval framework, *International Journal of Pattern Recognition and Artificial Intelligence*, **32**(12), 1850043, 2018.

29. S. Jabeen, Z. Mehmood, T. Mahmood, T. Saba, A. Rehman, M.T. Mahmood, An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model, *PloS ONE*, **13**(4): e0194526, 2018, doi: 10.1371/journal.pone.0194526.

30. J.M. Medina, S. Jaime-Castillo, C.D. Barranco, J.R. Campana, On the use of a fuzzy object-relational database for flexible retrieval of medical images, *IEEE Transactions on Fuzzy Systems*, **20**(4): 786–803. 2012, doi: 10.1109/TFUZZ.2012.2201726.

31. E. Yildizer, A.M. Balci, T.N. Jarada, R. Alhajj, Integrating wavelets with clustering and indexing for effective content-based image retrieval, *Knowledge-Based Systems*, **31**: 55–66, 2012, doi: 10.1016/j.knosys.2012.01.013.

32. N. Jyothi, D. Madhavi, M.R. Patnaik, Optimization of log Gabor filters using genetic algorithm for query by image content systems, [in:] S. Choudhury, R. Mishra, R. Mishra, A. Kumar [Eds], *Intelligent Communication, Control and Devices*, pp. 799–806, Springer, Singapore, 2020.

33. M. Kokare, P.K. Biswas, B.N. Chatterji, Rotated complex wavelet based texture features for content based image retrieval, 17th International Conference on Pattern Recog-

nition, ICPR 2004, Cambridge, UK, August 23–26, pp. 652–655, 2004, doi: 10.1109/ICPR.2004.1334250.

34. Q. Zheng, X. Tian, M. Yang, Y. Wu, H. Su, PAC-Bayesian framework based drop-path method for 2D discriminative convolutional network pruning, *Multidimensional Systems and Signal Processing*, **31**: 793–827, 2020, doi: 10.1007/s11045-019-00686-z.

35. Q. Zheng, M. Yang, X. Tian, N. Jiang, D. Wang, A full stage data augmentation method in deep convolutional neural network for natural image classification, *Discrete Dynamics in Nature and Society*, **2020**: Article ID 4706576, 11 pages, 2020, doi: 10.1155/2020/4706576.

36. Q. Zheng, P. Zhao, Y. Li, H. Wang, Y. Yang, Spectrum interference-based two-level data augmentation method in deep learning for automatic modulation classification, *Neural Computing and Applications*, **33**: 7723–7745, 2021, doi: 10.1007/s00521-020-05514-1.

37. I. Loboda, Y. Feldshteyn, V. Ponomaryov, Neural networks for gas turbine fault identification: Multilayer perceptron or radial basis network?, *International Journal of Turbo and Jet Engines*, **29**(1): 37–48, 2012, doi: 10.1515/tjj-2012-0005.

38. Q. Liang, Y. Wang, W. Nie, Q. Li, MVCLN: multi-view convolutional LSTM network for cross-media 3D shape recognition, *IEEE Access*, **8**: 139792–139802, 2020, doi: 10.1109/ACCESS.2020.3012692.

39. S. Antani, L.R. Long, G.R. Thoma, R.J. Stanley, Vertebra shape classification using MLP for content-based image retrieval, [in:] *Proceedings of the International Joint Conference on Neural Networks*, Vol. 1, pp. 160–165, 2003, doi: 10.1109/IJCNN.2003.1223324.

40. M. Kaur, D. Singh, Fusion of medical images using deep belief networks, *Cluster Computing*, **23**(2): 1439–1453, 2020, doi: 10.1007/s10586-019-02999-x.

41. S. Hussain, M. Hashmani, M. Moinuddin, M. Yoshida, H. Kanjo, Image retrieval based on color and texture feature using artificial neural network, [in:] B.S. Chowdhry, F.K. Shaikh, D.M.A. Hussain, M.A. Uqaili [Eds], *Emerging Trends and Applications in Information Communication Technologies, IMTIC 2012, Communications in Computer and Information Science*, vol. 281, pp. 501–511, Springer, Berlin, Heidelberg, 2012, doi: 10.1007/978-3-642-28962-0_47.