IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A SURVEY ON IMAGE MINING METHODS AND CLASSIFICATION BREAST MALIGNANCY

Mr. K. VigneshKumar[1], Dr. N. Sumathi[2], Dr. S. Ragul[3]
[1]Research Scholar (PhD) / Department of Computer Science,
Sri Ramakrishna College of Arts & Science, Coimbatore,
[2]Head / Department of Information Technology,
Sri Ramakrishna College of Arts & Science, Coimbatore,
[3]MBBS, DNB-Ortho, Narayana Hrudhalaya, Bangalore

Abstract-Image mining finds useful information in large image datasets. Data mining, digital image processing, machine learning, image retrieval, and AI underpin image mining. Image mining extracts hidden data and patterns from images. Two key issues might arise while mining gathered images: first, the image must be acceptable for mining, and second, the image's selected items and characteristics must be processed to retrieve the most effective way to save time and effort. This research describes the image mining technique and classifies breast cancers using image mining. In addition, a generic scheme is provided to complete the procedures and analyse the current malignancy classification methodologies compared to training groups and accuracy. The report also compares current relevant publications. Deep convolutional neural network accuracy was reported at 98%. (DCNN).

Keywords. Data Mining, Image Mining, Breast Malignancy, Image Classification

I.INTRODUCTION

Data mining is an approach that identifies potentially intriguing structures hidden inside large amounts of data. Data may be represented in a variety of ways, including structurally marked patterns, the link between the many pieces of gathered data, and mathematical or probabilistic models of a data. In addition to this, it may be put to use in the analysis of a huge quantity of data with the purpose of uncovering new and potentially useful information [1]. The advantages of utilising data mining include speeding up the way the data is processed or the pace of retrieving the data from the database, discovering or identifying the patterns that are hidden or nonexplorer, simplifying the data, and saving time [2]. In order to retrieve the newly revealed information, many algorithms and a variety of approaches have been used, depending on the steps that were required. These aims, which have been addressed so far in this article, have been broken down into the parts that follow. In the first section, we spoke about how the digital photographs were processed. The second half is a study of the most significant methods of photographic research, and the third segment investigated the uses of digital photos. The last section of this study is devoted to a discussion of the research that relates to image retrieval that is considered to be the most significant. Due to the exponential growth of the worldwide social network [3,] there is now an abundance of data; as a result, data mining is playing an increasingly important role in a wide variety of disciplines, including the commercial, satellite, and medical industries. The extraction of patterns from massive collections of photos is the focus of image mining, in contrast to the poor computer visions and image processing methods that are also available. This distinction is what sets image mining apart from these other fields. Picture mining is a technology that treats photos of unstated data or information and shows the connection of many patterns that are indirectly examined from the image. Image mining may also be thought of as visual pattern recognition. The classification of the image mining has been accomplished using a variety of methods. One categorization method involves taking information from either a database or a collection of photos and using that information to make a determination. Mining a combination of the gathered photos and the related alphanumeric data is another method that may be used for categorization [2]. Image mining in its most common

form may be broken down into five fundamental stages (see Figure.1) [4].

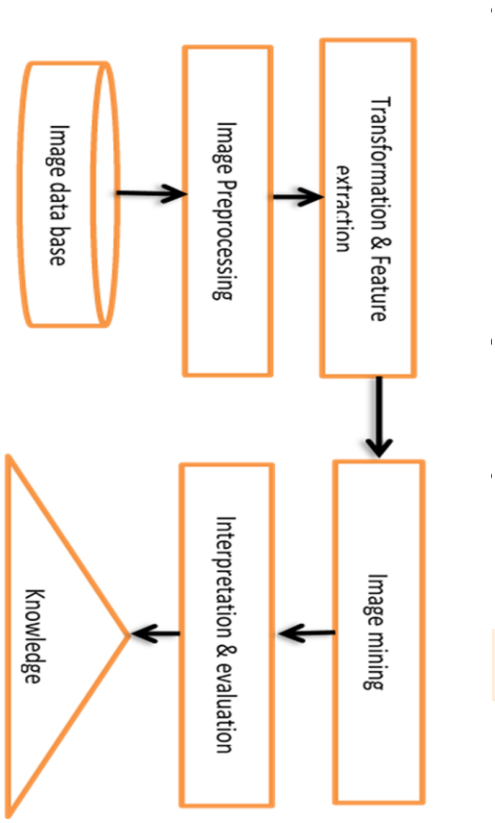

Figure 1. image mining process

The process of image mining includes a number of processes, including refining the picture database, pre-processing, converting and retrieving mining and data interpretation, and evaluation as the last step.

The following sequence of stages, which occur in order, may be used to describe the process's steps [5].

- Retrieving the picture from its storage location in the image database.
- image preparation is essential for the consumers as well as the application that is necessary. Improving the image's overall quality is one of the tasks involved in pre-processing, which comes before the feature extraction procedure. It is essential to do pre-processing on a picture in order to remove noise from the image, since this might lead to undesirable outcomes [3].
- Image Modification and Feature Extraction Image feature retrieval and extracting must be based on colour, texture, and shape, evaluated, and comparing the pictures. Image transformation is the process of changing an image so that its features are more easily retrieved.
- Image mining: The primary goal of image mining is to extract relevant data from an enormous picture database using a variety of image analysis techniques. The mining process is carried out by using the appropriate data mining tools in order to uncover the essential trends.
- Interpretation and analysis: after acquiring the resultant pattern, the pattern is eventually reviewed and interpreted to yield the final information, which may either be implemented individually or structured for a future application [4].

II. Related Work

[6] S. Chauhan et al. examined the identification and categorization of breast cancers. The suggested approach employs median filtering in pre-processing for MRI brain imaging. The coloured based segmentation and edge detection were used to separate the lesion from the picture. Moreover, the pictures were shown using different feature extraction approaches, including histogram of directed gradients and grey level co-occurrence matrix. To store all of the extracted characteristics, a transactional database was employed. The IBkLG classifier (Instance-based K-Nearest utilising Log and Gaussian weight Kernels) was used in WEKA 3.9 to classify the tumour as benign or malignant. The categorization was classified with an accuracy of 86.6%.

S. Kumar et al. (2017), [7] used a hybrid technique to classify breast cancers. The method comprises employing the discrete wavelet transform (DWT) to extract the features, the evolutionary algorithm to reduce the number of features, and the vector machine to support it (SVM). When compared to state-of-the-art procedures for the same range of context, the yield of using the strategy announced an increase in accuracy and a reduction in error for the RMS. Linear accuracy ranges from 80% to 90%.

Z. Ullah et al. (2018) [8] proposed an intelligent system for classifying magnetic resonance imaging of the breast cancers. Pre-processing, feature extraction, feature reduction, and classification are the four steps of classification. At the first step, noise is removed from DM images by using a median filter and then transforming the picture to RGB. In the second step, Discrete Wavelet Transform (DWT) is used to extract picture characteristics, which are then reduced using colour moments. The last stage of decreased features is categorised into normal and DM image using the k-Nearest Neighbours (k-NN). The findings demonstrate that the system outperforms the other strategies in terms of lowering the number of characteristics. This approach has an overall accuracy of 94.97%.

According to N. A. Mazin et al. (2018), [9], employing ANN from DM images may enhance the automated breast cancers. segmentation and identification

technique without the need for human intervention. K-means clustering is used as part of the primary organisation in the process to enhance the DM image's detection of grey scale in the district areas. Next, in the training phase, ANN is used to choose the proper item. The textural characteristic of the breast cancers. region may then be extracted into the division stage. The ANN method has a 94.07% accuracy.

M. Kaur and B. Prajapat (2018) presented an automated approach for malignant breast cancers. categorization [10]. The hypothesis is based on the Haarlet transform and a probabilistic neural network. For data processing, threshold-based segmentation and binarization, as well as the Haarlet transform, were utilised. The feature extraction process produces twelve features for each training picture, which are then utilised to train a probabilistic neural network. The suggested system's accuracy was 96.3%.

N. Abiwinanda et al. (2019), [11] sought to train a Convolutional Neural Network (CNN) to distinguish the most prevalent three forms of breast cancers. (Glioma, Meningioma, and Pituitary).

In order to find the best CNN architecture, the simplest feasible CNN architecture was developed (architecture 2). Architecture 2 is made up of two convolution, activation (ReLu), and maxpool layers, followed by a hidden layer of 64 neurons. Among all designs, architecture 2 shows a steady-state declining trend in validation loss as the number of epochs increases. The training accuracy attained was 98.51%, with the greatest validation accuracy being 84.19%.

F. zyurt et al. (2019), [12] proposed a hybrid technique based on Neutrosophy and Convolutional Neural Network (NS-CNN). The neutrosophic set-expert maximum fuzzy-sure entropy (NS-EMFSE) technique is used to segment the MRI images in the first step. At the classification step, the CNN is used to retrieve the characteristics of the segmented brain pictures, which are then classified using SVM and KNN classifiers. The outcome of the investigation indicated that the CNN features exhibited a good classification performance with various classifiers. The experimental findings were consistent with the SVM simulation output data of 95.62% average success, and the CNN features performed better in classification.

H. H. Sultan et al. (2019), [13] said that they used a DL model based on a deep neural network to categorise various kinds of brain tumours. Two publicly accessible datasets were employed in the suggested method. The first database categorises cancers as (meningioma, glioma, and pituitary tumour), whereas the second distinguishes between the three glioma grades (Grade II, Grade III, and Grade IV). For the first and second datasets, total pictures were 3064 and 516 on T1-weighted contrast-enhanced images from 233 and 73 patients, respectively. This suggested network structure obtained an overall accuracy of 96.13%.

K. Vigneshkumar et al. (2021), [14] suggested employing various stages of cancer (DM). neural networks (PNNs) and a genetic algorithm are used in the suggested technique (GA). CANCER types were correctly identified 96% of the time. The efficiency of the suggested strategy has been proved to categorise the breast cancer using DM Images. Because of the method's adaptability, it may be used to aid in the early detection of breast cancers.

K. Vigneshkumar & Dr.N. Sumathi et al. (2021) [15] explored the Deep convolutional neural network (DCNN)(VGG-19). The suggested approach reports a mean classification accuracy of 98%.

Table 1. Methods Used in the Previous Related Works with the Proposed Techniques

S.NO	AUTHORS	YEARS	METHOD CLASSIFICATION	TRAINING	ACCURACY
1	S. Chauhan et al.	2017	k-Nearest Neighbors (k-NN)	60%	86.6%
2	N. A. Mazin et al.	2018	Artificial Neural Network (ANN)	70%	94.07%.
3	S. Kumar et al.	2017	hybrid approach:(DWT) & (SVM)	80%	85%
4	Z. Ullah et al.	2018	K-Nearest Neighbors (k-NN)	70%	94.974%.
5	M.Kaur &B. Prajapat	2018	probabilistic neural network (PNN)	68%	96.3%
6	N. Abiwinanda et al.	2019	Convolutional neural network (CNN)	80%	84.19%
7	H. H. Sultan et al.	2019	Deep neural network (DNN)	80%	95.13%
8	F. Özyurt et al.	2019	Neutrosophy and Convolutional Neural Network (NS-CNN)	80%	95.62%
9	K. Vigneshkumar & Dr.N. Sumathi	2021	Deep convolutional neural network (DCNN)(VGG-19)	80%	98%
10	K. Vigneshkumar et al.	2021	hybrid approach: (PNN) & (GA)	80%	96%

III. Image Mining Technique

In order to mine the information contained inside a picture, a number of different methods, including object identification, retrieval, images indexing, image classification, image clustering, mining association rules, and neural networks, are used. The image mining processes are shown in Figure.2 and Figure.4.

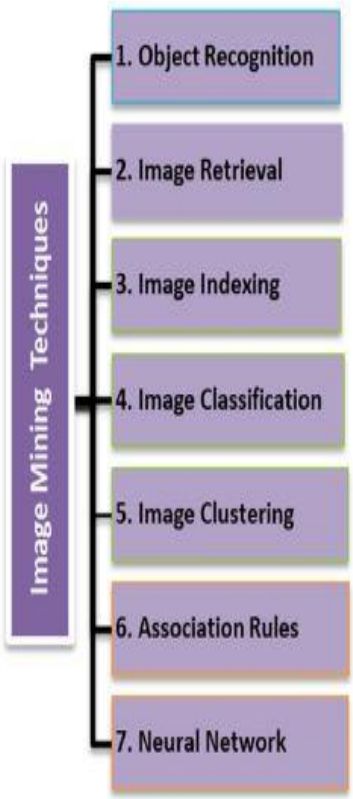

Figure 2. Image Mining Techniques

3.1 Object Recognition

This method is now being investigated in the realm of image processing research as an active method. In order to extract the object from the photos included inside the database, object models and recognition techniques are employed. After the machine has identified and recognised a small number of items, useful information may be extracted easily, and the system can begin to automatically acquire new skills. Object recognition may be thought of as a kind of supervised labelling that is based on object models of the things that are being looked at [4, 5].

3.2 *Image Retrieval*

The process of image retrieval can be broken down into three distinct levels, each of which corresponds to a higher level of difficulty [16]. This is due to the fact that more and more people are accessing the information that is accessible via the web and on the interactive media of as the world concentrate on intricate and accelerating retrieval.

Level 1: consideration involves low-level features such as colour, texture, and form when retrieving pictures.

Level 2: Taking into consideration aspects of a higher level of logic, such as people or things of a certain sort.

Level 3: This stage involves thinking about high-level aspects, such as those concerning the meaning or the situations that are presented.

3.3. Image Indexing

The purpose of picture indexing is to get from the ImageNet dataset any images that are a match for the query image that was given. The distinguishing characteristic of each individual picture [4]. The picture may be measured in a similar fashion based on the characteristics such as the colour, the intensity, the texture, the location, and the form. Image index can be accomplished by comparing the features.

3.4. Image Classification

The classification facilitates the decision-making process. The processing of information that takes place during the categorization phase is what makes it possible to divide the pixels that make up a digital picture into their respective categories [3], It is a kind of learning called supervised learning, and it is used to categorise pictures based on certain outcomes that have already been categorised.

3.5. Image Clustering

It is a procedure that collects data objects that are comparable to one another. A second cluster is made up of dissimilar objects. It is the process of looking at the comparisons between the data and pointing such similarities out based on the characteristics of the data. Picture clustering, also known as unsupervised clustering, is a technique that classifies photographs into categories based on the information contained within them [3]. This approach does not need any previous information.

3.6. Association Rules Mining

In both the process of discovering information and mining data, it is an effective tool for recognising patterns [4]. There are two primary approaches that are represented as the basis of the work that is done by the standard association rule algorithm. The first step is to identify all of the important item sets that are compatible with the minimal support requirement. The second part of this process is the transfer of the rules that were developed for each item; this helps to complete the minimal confidence constraint [5].

3.7. Neural Network

In the process of retrieving and mining pictures, one method that is used is called a neural network. In its most basic form, it is employed inside a computational system

that is constructed from elementary processing units called neurons. It is built up in layers, each of which is linked either partly or completely. Neurons are primarily responsible for three functions: receiving input values from their neighbour neurons (which are considered their output), assessing those output values based on the weight of their input, and lastly, transmitting those output values to their neighbours [5].

IV. Applications

Because of an increase in the amount of data that is available, there has been a rise in the demand for image mining applications in a variety of fields. Some of these fields include the study of space - based cloud images, the study of neural networks, the classification of medical images, and others, some of which will be discussed further down.

Satellites: A vast quantity of information is gathered from spacecraft via remote sensing instruments. As a result, several mining strategies are used in the many applications that make use of satellite imagery [17].

Medicare: Image classification as well as rule mining methods are the most sought after in medical treatment, according to Medicare. methods of image mining for the diagnosis of illnesses

Bank: Image mining methods are used in the banking industry for the purposes of safeguarding the data, retrieving the data, and scanning customers' eyes for fraudulent activity prevention.

Television: These systems capture and analyse, on the fly and in an anonymous manner, information gleaned from the channel view, broadcasts, and programming.

Traffic Image: The role of the image processing and analysis subsystem is to differentiate the movements from the picture sequences and to provide the time and space properties such as object type, location, orientation, and speed, among other things [17].

Biological Data: Tissue images extraction is a creative and powerful technology when the pictures of the tissue are indexed, saved, and mined for information [18]. Tissue image mining is an example of this.

V. Using Image Mining for Breast Malignancy

Adults currently die most from breast cancer. The illness and growing instances continue to impact public health. Mammography pictures help compare screenings and early detection clinically. Mammography reduced morals by 40% to 60% after four to eight years. Mammographic photographs, X-ray film, and supervised film processing provide better mammograms. High-resolution mammograms improve contrast.

Artificial Neural Networks, Deep Learning, Probability Reasoning, and Data Processing lead AI ethical research. AI, ML, and DL are related. ML & DL help AL solve data-driven problems through algorithms and neural networks. Humans learn through experience, and robots obey instructions. It emphasises system design, learning, and prediction based on past experience. The health-care business will benefit most from AI's $1 trillion impact on the world economy by 2030. Nuance AI technologies help physicians save paperwork time and increase report quality. Nuance prediction services leverage AI to anticipate customer needs and respond accordingly.

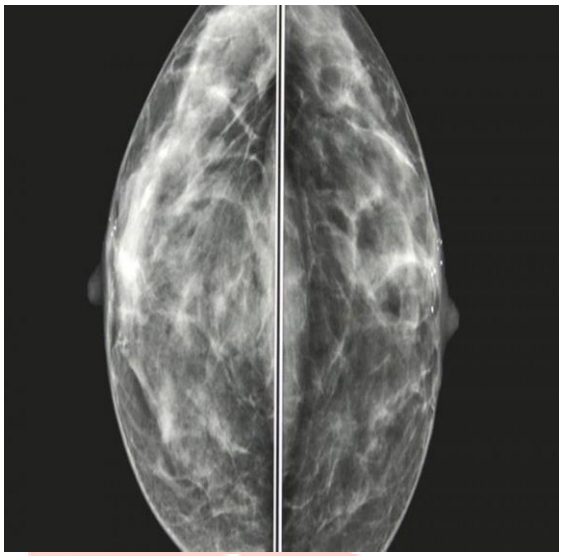

Figure 3. Normal and Cancer Mammogram Images of Breast Cancer

Tumors form when normal cells have faults (mutations) in their DNA. These mutations speed up cell growth and division while also allowing them to survive when healthy cells perish. As a result, a clump of aberrant cells grows, becoming a tumour. Pre-processing helps to improve the image's quality so that it may be used in subsequent treatments. The RGB input picture is transformed to grayscale in this step. Several phases are also linked with the process, such as picture enhancement, contrast improvement, and image sharpening. The procedure that follows pre-processing is segmentation. Instead of being pixel based, segmentation algorithms are area orientated. Segmentation depends on separating a picture into relevant parts, i.e., breaking the image into linked portions [22]. The primary goal of image segmentation is to separate a digital picture into several parts. Feature extraction is the technique of extracting the cluster from the segmentation output in order to diagnosis the tumour. Lastly, image mining is utilised to identify and categorise tumours, determining whether the mass is malignant or benign. Figure 4 depicts a block schematic of the tumour categorization processes.

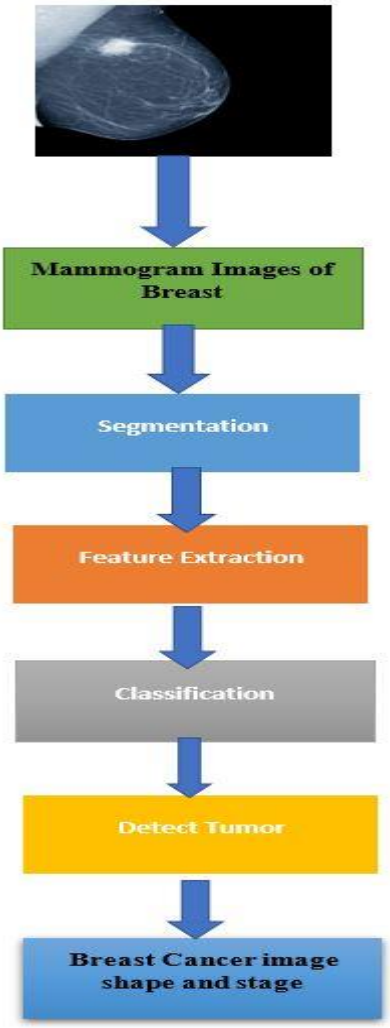

Figure 4. Block Diagram

VI. Analysis of Survey Studies

The measurement of training and testing groups, as well as the collection of exactness, revealed that the greater the training group, the more experience the picture classification system would have. When the accuracy of the tested groups is great, however, the overall training is little. This demonstrates the suggested system's great efficiency. As a result, Chauhan et al. (2017) employed k-NN of group training (60%) and accuracy of (86.6%).

Nonetheless, it is clear that the accuracy is low. At the same time, Kumar et al. used a hybrid strategy that used the discrete wavelet transform (DWT) and a genetic algorithm to reduce the number of features and support the vector machine (SVM) in order to achieve an accuracy of (85%). In, Z. Ullah and colleagues proposed combining the benefits of the two methods by combining the k-Nearest Neighbors (k-NN) and the Discrete Wavelet Transform (DWT) techniques to achieve an accuracy of (94.974%) with a group training of (70%). Mazin et al. later used the Artificial Neural Network (ANN) to achieve an accuracy of (94.07%). Kaur et al. employed the probabilistic neural network (PNN) to achieve group training of 60% and accuracy of (96.3%) in the same year.

Abiwinanda et al. (2019) proposed utilising a Convolutional neural network (CNN) to achieve 84.19% accuracy. Sultan et al. (2019) used the same Deep neural network (DNN) methods on two accessible datasets to achieve an accuracy of up to (96.13%).

Vigneshkumar et al. (2020) used a hybrid strategy of neural network (PNN) and genetic algorithm (GA) to achieve 96% accuracy. Ultimately, the newest approaches presented by Vigneshkumar and Sumathi (Deep convolutional neural network (DCNN)) achieved the best accuracy (98%) among the other experiments.

VII. CONCLUSION

To conclude, the present article highlighted the importance of adopting image mining for a variety of reasons, including the growth of picture databases. Moreover, we have covered the majority of the existing image mining methods, including object identification, image retrieval, image indexing, image classification and clustering, association rule mining, and neural networks. The dynamic and future perspectives of image mining applications have been thoroughly examined. Moreover, the present study includes a review of published research in the area of picture exploration, as well as a detailed reading of the researchers who employed image mining approaches in the identification and classification of Breast Cancer. Lastly, an examination of existing published research that employed image mining methods would aid in the development of future work in other sectors, such as expanding the usage of image mining to deal with video databases. Performing picture mining on worldwide webs as a source for image databases, as well as suggesting appropriate indexes and retrieving knowledge