Efficient Image Mining in Crop Diseasesusing Svm Classification and K-Means with Colourcorrelogram

Dr.K.Venkatasalam,
Associated Professor,
Mahendra Engineering College ,
Namakkal

Dr.P.Ramya,
Associated Professor,
Mahendra Engineering College,
Namakkal

Abstract:-Agriculture is one of the main economic activities of Tamilnadu for its livelihood. Agricultural cultivation in the state of Tamilnadu contributes considerably to the state's gross national income. These different crops are influenced by the attack of harmful insects. Therefore, the identification of harmful insects in the field is an important task for sustainable agricultural development in Tamilnadu. Plant tract / disease studies refer to studies of visually observable models of particular crops. Nowadays, cultures face many diseases. Insect damage is one of the main diseases. Insecticides are not as efficient as insects are. It also damages the natural food chains of animals. A common practice for the plant is due to the damage of the plant (leaf, stem). This translates into subjectivity and low performance.

Today, agricultural organizations work with large amounts of data. It is necessary to process and retrieve significant data in this abundance of agricultural information. The use of information and communication technology automation to extract meaningful data in an effort to obtain knowledge and trends, which enables data mining as well as crop information, allows farms to predict trends in his behavior, which is obtained by analyzing data from different perspectives and finding connections and relationships in apparently unrelated data. The raw data of farms are very large and diversified. It is necessary to collect and store them in an organized way and their integration allows the creation of an agricultural information system. Data mining in agriculture offers many opportunities to explore hidden patterns in these data collections. Content-based image retrieval technique plays an important role in the concept of image recovery, but requires a manual annotation of each image in a database, which is prohibitive in a large image database. In textual annotation it is inadequate and ambiguous to search for image databases. This text annotation depends on the language. Sometimes it is not compatible with the CBIR technique.

To overcome these problems, we propose one of the RBIR techniques . It supports both query by region of interest and keyword query, as the system segments an image into regions and extracts low-level regions and high-level regions of an image. From this, high level characteristics are obtained. The queries entered by the user are analyzed based on the analysis of the content and the analyzed queries are retrieved from the database with the help of the indexing method. From this, the spatial information on the color image is managed with the help of the color correlogram technique and, finally, the image center is located from the database with the help of the k-means algorithm.

This document provides progress in various methods used to study disease crop traits through image processing and data extraction. The methods studied are to increase the yield and reduce the subjectivity that derives from human experts in detecting the diseases of the cultures that are stored. The different characteristics of the crop are grouped into an "extraction of associative rules".

Keywords:- RBIM, SVM classifier, Content Analysis and Indexing, k-Means Clustering Algorithm, agricultural enterprises.

*****

I. **Introduction**

In recent years there has been a large area of infection from crop diseases and pests, in addition to the degree of severity, which has caused huge economic losses to farmers. Harvest losses due to pests and diseases are quite considerable, particularly in the semi-arid conditions of India .In the current scenario, agricultural data are virtuallybe collected together with the crops and are collected stored in the databases. As the data volume increases, thegap between the amount of data stored and the amount ofthe analyzed data increase this data can be used inproductive decision making if appropriate data mining (DM)the techniques are applied. DM allows you to extract the maximumimportant information from many data and to be discoveredpreviously unknown models and hidden relationshipswithin the data that could be relevant to the current agricultureproblems. With the increasing amount of informationin their farms, farmers not only collect in terms ofagricultural production, but also a large amount of data. Thesedata must be used for optimization [33].

Critical threshold ofmeteorological elements for incidence, propagation andintensification of pests and diseases determined inlaboratory conditions have little relevance to the fieldcondition. Therefore, they must be determined esupervised in simultaneous field conditions throughobservation of micrometeorological parameters relevant data on pests and diseases [34].The simple definition of data mining in marketing is the extraction previously unknown, understandable and suitable for great information of data files and their use in business decisions in order to support them are carried out by formulating marketing initiatives tactics and strategies and measure their success. This definition emphasizes the following aspects of "data mining": information extraction, large data repository and development of initiatives. "Data Mining" simplifies the correlation finding in the data. "Data mining" is an extremely interdisciplinary field. It covers database fields, expert systems, information theory, statistics, mathematics, logic and a number of related fields. The fields in which data mining can be successfully applied are diverse, for example, business organization, economics, engineering, medicine, genetics. In general, data mining is applicable in all areas where there are certain regularities, links and rules on the basis of large data. It can be said "data mining" is finding the rules in the data. Data extraction technology is closely related to data storage and is related to management system databases. Data mining involves the process of finding a large number of previously unknown data and, consequently, its use in important business decisions. The key phrase is "unknown data", which means that data is hidden in a lot of operational data, if analyzed, providing relevant information to decision makers within the organization.

The raw data is displayed as a single table required by the mining algorithms included in WEKA, an open source system that provides a collection of visualization tools and algorithms for data analysis and predictive modeling. The table is then transformed into an attribute value table that includes header information based on attribute data types. This requires the cleaning of data, the treatment of lost values, the discovery of false values, etc. once the data mining algorithms process this data, they are transformed into a module that will produce a readable and accurate data model.One or more versions of clean data are now processed by data mining schemas. Now it is decided which groups are accurate or interesting enough to justify future research and which are common knowledge for that field. The ability to search for images continues to be more important as the number of available images increases dramatically. However, it is not practical for humans to label every image available and to dictate which images are similar to other images. Therefore, there is the urge to teach computers how to correlate images while minimizing human work. The normal approach is to use the low-level features of images such as color, textures, and edges to try to predict similarities between images[8][9]. Unfortunately, there is a semantic gap between what shows low-level characteristics and high-level features that represent what a human understands in the image. Content-based image retrieval (CBIR) only attempts low-level functions to overcome the problem with region-based image recovery[18]. The RBIR acts as a bridge between these categories.[1,2] It focuses primarily on the high-level functionality of an image, from which the user can obtain relevant comments of the query required by an image database. This document describes a construction methodology for image-based restoration of regions and features of an image by automatically extracting and querying information about color images [26].

Georgios et al (2014) developed an image recovery approach based on the region, where introduction of a new set of look features to perform prediction of the assessment of user relevance at the regional level and the design of an effective and efficient object-based RF framework for image recovery through the application of fixed-line signal and based on objectsovercome the main limitation of affordable and portable RF-based approaches in the region.Feng Jing et al. (2004) proposed an image recovery framework that integrates an efficient representation based on regions in terms of storage and complexity and an effective ability to learn and compare online, indexing using modified inverted files, relevant feedback and weighting of learning regions in Earth Moving Distance are presented.Simone A. Ludwig (2015) has the management and analysis of big data through the use ofK-means algorithm (FCM). The algorithm is parallel using the MapReduce paradigm which describes how the primitives Map and Reduce are implemented.

MAO et al (2015) deals with cluster centers with a particle swarm optimization algorithm (PSO) and FCM, a parallel optimization algorithm that uses an improved k-means method combined with particle swarm optimization (AF-APSO)To get the best physical performance and grouping significantly.Mohsen Zand et al. (2015) have the RBIR Gabor wavelet and the curvature transformations, a method for encoding the information of the sub-bands in the polynomial coefficients to create a vector of plot characteristics with maximum discriminating power. The task of classifying the plots with the ImageCLEF and Outex databases demonstrates the effectiveness of the proposed approach.Keng-Pei Lin et al (2011) has a classifier with SVM that preserves privacy and does not reveal the private content of support carriers. the classification accuracy of PPSVC is comparable to the original SVM classifier

Lorenzo Bruzzone et al (2009) has a semi-supervised vector classifier machine (CS4VM) sensitive to the context by exploiting the contextual information of the pixels belonging to the proximity system of each training sample in the learning phase to improve the robustness of a possible training model wrong.Atony Fierro-Radilla et al (2014) have recovered image content based using descriptor Correlogram dominant color (DCCD) and shape characteristics using pyramid gradients oriented histogram (Phog) having the task of image recovery as, medium accuracy recovery (ARP), mean recovery rate (RPC) and Modified Scale average normalized recovery (ANMRR) and mean recovery curve (R) - Medium precision (P).Ahmed Talib et al (2014) recovered image recovery based on content using dominant colors of an image to avoid sequential search in the database, reducing the search space to less than 25% without degrading accuracy[7].

Shahana N Youseph et al (2014) have MVR security image, different trial process, the detection of inconsistencies between different images can be identified with the help of the estimated based illuminating color pixels and border areas of the exploited images as a method of detection of the falsification and how they are provided for decision making with minimal user interaction.Samuel Barrett (2007) has Retrieval Based Learning Content Images from short and long term machine learning Support Vector blurred semantic concepts depicted in images using relevance feedback and the effectiveness of semantic clustering.Bassam (2012) deals with the middle C cluster of Fuzzy (FCM), the FCM algorithm to allow the generation of equal sized clusters. In addition, scattered points that are far from all clusters are grouped into groups. Another change is to locate specific points that have the ability to be in more than one group

J. C. Dunn (1973) has grouped the signals of the "fuzzy" algorithm of the presence or absence of CWS groups in X, the mathematical arguments and the numerical results are offered in support of the declarations mentioned above.Byoungchul ko et al (2008) deals with image-based restoration in the region using flexible metrics for the query image, local data densities can be sufficiently exploited, improved performance. Ying Liu et al (2007) has an image recovery based on semantics using a decision tree based learning algorithm called DT-ST that contributes to the problem of discretization, improves recovery performance and has the simplification method of the hybrid tree to manage the problems of noise and noise fragmentation of the trees thus improving the classification performance of the tree . Dr. Fuhui Long (2003) deals with the fundamental theories of image-based image recovery used for the development of content-based image recovery techniques, widely used methods for describing visual content. Srinivasa Kumar Devireddy (2009) has automatic and effective techniques for content-based image recovery systems (CBIR). The main idea is to represent each image as a vector of characteristics and to measure the similarity between the images with the distance between their vectors of corresponding characteristics according to a certain metric. Finding the right features to represent images, as well as the similarity metric that groups together visually similar images together, are important steps in building any CBIR system.

Debashis Debnath et al (2011) deals with the recovery of mobile images and proposes a new effective and efficient method for image recovery from mobile devices that applies a weighted combination of color and texture using spatial color and second-order statistics. The system for searching for mobile images is performed in real time on an iPhone and can be easily used to find a specific image.Niket Amoda et al (2013) has image recovery based on the region by methodDiscrete Wavelet Transform (DWT) and a k-means clustering algorithm, Reducing manual image annotation is a costly and expensive job for a large image database. Yossi Rubner et al (2001) has a performance evaluation through the use ofEMD, multivariate histograms, marginal histograms for higher classification and recovery performance. Shows how the selection of a measure is based on a large-scale evaluation.Fei Li, Qionghai Dai et al (2008) has a framework based on the propagation of multilabel neighborhoods is proposed for more accurate RBIR weighted graph for label propagation and can be calculated more meaningful labels of high level to describe the images.Yong Rui (1999) wrote a comprehensive survey on image retrieval, the application for real-world applications, open search problems have been identified and promising future research directions have been suggested.

B.MiloviC and v. RadojeviC (2015) wrote a application of data mining in agriculture, Raw agricultural data the companies are very large and diversified. it is necessary to collect and archive in an organized way and their integration allows the creation of an agricultural information system. Data mining in agriculture offers many opportunities to explore hidden patterns in these data collections. These models can be used to determine the conditions of customers in agriculture organizations.Thomas M. Deserno et al (2009) Long wrote a Gaps in content-based image retrieval, CBIR applications are incorporated into routine clinical medicine or medical research. The cause is often attributed without sufficient analytical reasoning for the inability of these applications to overcome the "semantic gap". The the semantic gap divides high-level analysis of the human scene from the analysis of low-level computer pixels.A.S Deokar et al (2016) wrote provide a system that can unintentionally obtain significant characteristics of the plant affected by the disease and calculate the image of the diseased plant loaded. Stands for Easily help the plant biologist diagnose plant disease and give farmers an initial precaution measures. Recognizing the portance and dominance of the agricultural sector, a system based on content Image recovery techniques and the k-means algorithm for the diagnosis of plant disease are proposed.

Dr.Hari Ramakrishna et al (2012) wrote The sheet with force suggests the need to use a method similar toInformation technology for the agricultural sector.Barbora Zahradnikova et al (2015) wrote image mining: review and new challenges, Current image mining approaches and techniques that aim to expand the possibilities of facial image analysis. This document aims to review the current status of the IM, as well as in the description of the challenges and the identification of the addresses of future research in the field.Jaganathan Palanichamy et al (2014) wrote It isour opinion that effective techniques can be developed and adapted to solve complex agricultural problemsusing data mining At the end of this study, we provide suggestions for future research directions in agriculturefields.M. Yasmin et al (2014) wrote an efficient content based image retrieval using classification and color features , the results showing the proposed method an excellent balance between precision and recovery in a minimum recovery time, the results achieved are between 66% and 100% Accuracy rate and 68% -80% to remember.

2. Content Based Image Retrieval

CBIR technique that uses visual content to search for large-scale images of image databases has been an active research area over the past decade. Imagebased image retrieval is the recovery of images based on visual characteristics such as color and texture [22]. The reasons for its development are that in many large image databases, traditional methods of image indexing have proved to be insufficient, laborious and extremely slow. These old methods of indexing images, ranging from storing a database image and associating it with a keyword or number, to the association with a categorized description, have become obsolete [17]. This is not in the CBIR. In CBIR, every image that is stored in the database has its characteristics extracted and is compared with the characteristics of the image of the query. It involves two steps:

2.1. Content gaps

This group of gaps is about modeling, understanding and using images from the user's point of view. As a result, two gaps seem relevant.

2.1.1. Semantic gap

The similarity of images defined by a human observer in a particular context is based on a high level of semantics, which is usually dealt with by assigning meaningful labels to the concepts represented[25]. In contrast, the computational analysis of the image content is based simply on gray pixel values[6]. In our definition, the semantic gap is connected if a relationship is established between the image structures and the medical meaning. This vacuum in a system is therefore:

- Not addressed: significant terms are not assigned to images or ROI.
- Manual: significant terms are assigned manually.
- Computer-assisted: a semi-automatic process is used to assign meaningful terms.
- Automatic: significant terms are assigned automatically.

2.1.2. Context gap

The context in which a CBIR system can be used is generally limited. CBIR medical systems are often designed to allow consultations in a specific imaging modality or within a specific clinical context, such as the protocol used or diagnoses. This restriction allows the use of a priori medical knowledge of the imaging modality or context, which might otherwise be difficult to formulate to be calculable. It may be desirable that the system supports widespread use with minimal or no limitations on the user. As such, to close the context gap, a system can be classified as onewhere this feature is:

- Not addressed: the system is a specific context and the gap in the context is wide.
- Limited: restrictions apply only to modality, protocol or diagnosis.
- General: no restrictions apply, either to the modality or to the protocol or to the diagnosis.

2.2. Characteristic gaps

The gaps related to the characteristics derive from the computational point of view. The spaces correspond to the inadequacies of the numerical characteristics chosen to characterize the image content.

2.2.1. Extraction gap

Not all CBIR medical systems automatically extract features. Some are based on manual indexing of images, which is accompanied by considerable efforts and the potential for errors. This gap is connected via computer or automatic methods of extracting the characteristics are obtained from the input data

- untreated: completely interactive or manual, for example, shapes that are manually defined.
- Assisted computer: partially interactive, for example, forms segmented with the "livewire" algorithm -Automatic: non-interactive.

2.2.2. Space structure

The extraction of global parameters describing the complete image is often insufficient for medical applications. Therefore, regions of interest (ROI), which describe only a certain part of an image, must be identified and characterized by the appropriate parameters. To fill this gap, the assignment of the image features is

- not direct: for the whole or global image.
- local: for a single ROI.
- Relational: for a specific composition of ROI or individual objects.

2.2.3. Scale gap

Because an appropriate size of the ROI or scenes still depends on query activity and context, and therefore is variable, dedicated multi-scale approaches for describing the image content must be developed. To fill this gap, the scale of image analysis is- Not addressed: a single fixed scale is used.

- Multi: a multi-scale approach is applied.

2.2.4. Dimension Gap

A system has this shortage if features are extracted and used in a size smaller than the size of the original data. For example, three-dimensional data is often processed as individual 2D sections. However, for 1D biomedical signals and 2D medical images, this interval does not exist. The gap in the system is identified as:

- Not addressed: the system manages only 1D or 2D data.
- Full range: for example, color characteristics are used for color images.
- Complete domain: for example, volumes are used as ROIs for 3D data.
- Complete both: in other words, neither a domain nor a range is opened.

2.3. Performance gaps

Not all systems found in the literature are fully implemented and are executable for performance evaluation. For those that can be tested, performance criteria include the quality of integration and evaluation, as well as other classical performance measures.

2.3.1. Application gap

In the scientific literature there is a huge gap between the conceptual level of CBIR systems and their implementation or institution. Often, the concepts are published. The application gap is reduced if a CBIR medical application is present

- Not addressed: not mentioned at all.
- Mentioned: in the project description, but no proof is provided.
- Documented: screenshots are shown in the publication to test the implementation of the system.
- Offline: available for download and installation.
- Online: directly accessible and executable via the Internet.

2.3.2. Integration gap

If there is a system for medical CBIR, another space is opened. Such a system is usually independent and not sufficiently integrated into the clinical routine. The integration gap is based on the level of workflow integration. These levels are

- Not addressed: the application is not interconnected with the clinical software.
- Date: the application can access clinical data.
- Function: the application can come from other clinical software.

- Context: the patient / actual image information is passed to the CBIR application.

2.3.3. Indexing gap

The performance of a CBIR medical system also depends on response time and indexing of image descriptions on multiple scales for efficient data access. This indexing is not retrieval. Simple strategies like A * -trees or reverse files cannot be applied directly, and extensive research is needed to target large image archives as they are generated in health care.

The indexing gap is linked if the calculation of the similarities is performed using the following approaches:

- Not addressed: the system is based on a brute-force approach, in which all the characteristics of each image are compared.
- Parallel: the calculation of distances is brute force but distributed.
- Indexed: quick access to the cluster of relevant features or cluster trees is provided.
- Both: the CBIR application uses clustered forests with distributed computing.

2.3.4. Evaluation gap

In large databases, the gold standard or the truth of the terrain is not known, that is, it is impossible to determine the correct answer for a test query. In other words, an expected result of the system answering a given question is not available. Therefore, the comparison of competitive approaches for the extraction of global and local features and distance measurements is difficult and inaccurate. Instead of the error measures calculated by the experiments that come out one, the precision, the recovery and the measure F are calculated, where the number of correct answers is not used. The experiments are carried out

- Not addressed - xxx: no experiment has been described, but the database contains xxx images.
- Qualitative - xxx: no expected result or underlying truth based on the xxx images.
- Quantitative - xxx: with expected output, based on xxx images.

2.4. Usability gaps

This group of gaps concerns the usability of the system. While performance gaps are concentrated on the area in which the system is used, the usability gaps describe the ease of use of the system from the end user's point of view.

2.4.1. Consultation gap

Using the QBE paradigm, in which a visual example is presented to the recovery system, specialized mechanisms and interfaces are needed. Currently, there is a lack of appropriate tools to help the user draw or compose a research model, and the QBE is difficult and time consuming. The query gap in the system is identified as:

- untreated: the alphanumeric text is used without taking into account the QBE paradigm.
- Characteristic: certain intervals of vector characteristics or vector components are supplied by the user.
- Pattern: said pattern can be an example image or part of an image (ROI).
- Composition: the user selects interactively and positions the structures of a given set.
- Sketch: the system allows the insertion of models created individually and interactively, including the options above.

2.4.2. Gap Feedback

The result of a CBIR query is usually presented by displaying the most similar images found in the file. However, it is difficult to understand why the images presented are similar and how the query should be modified to improve recovery.

To close the feedback gap, the CBIR system provides a justification for the results obtained. This can be - not addressed: the results returned by the system are not commented at all.

- Base: a similarity number is provided for each returned item.
- Advanced: the system provides more sophisticated explanations.

2.4.3. Refining gap

CBIR systems should provide the user with options to repeat and edit a query. Sometimes, they also follow refinementprocess to learn from user preferences. To close the refining gaps, the query refinement is

- Not addressed: only one request is given.
- Returns: a rudimentary option is provided to refine the query.
- Backwards: in the improvement cycle, the user can go back if the results get worse.
- Complete: a complete history of the interactive session is available to restore any intermediate phase.
- Combination: several consultations can be made and the results can be combined.
- Learning: during use, the system adapts to the user's needs. [27]

3. Region-Based Image Retrieval

Segmentation based on the region: these methods focus on an important aspect of the segmentation process omitted with point-based techniques. There, a pixel is classified as an object pixel judging only by its gray value regardless of context. This means that individual points or small areas can be classified as pixels of objects, regardless of the fact that an important feature of an object is its connectivity. The procedure is to limit the size of the mask on the border to the points on the object or background. This can only be achieved if we can distinguish the object and the background after the characteristic calculation.

Obviously, this problem cannot be solved in a single step, but only one procedure is used iteratively in which the calculation and segmentation of the features are performed in an alternative way. In the first phase, the characteristics are calculated without taking into account the limits of any object. Then a preliminary segmentation is performed and the functionalities are recalculated, now using the segmentation results to limit the neighborhood operation masks to the edges of the object relative to the object or pixels of the background, depending on the position of the central pixel. To improve the results, it is possible to repeat the calculation and segmentation of the characteristics until the procedure converges into a stable result [29].

Region-based image retrieval is an attractive approach that works as follows: (i) Images are segmented into different regions; (ii) characteristics are extracted from each region; and (iii) the set of all features is used to represent the image content of an image database. At the time of the query, features are first extracted from the query image, a user-supplied sketch, or a segmented image. These extracted features are compared with the features [23] representing the database images. However, the RBIR generally grants high premiums for the quality of segmentation, which is often difficult to achieve in practice due to several factors, such as changing environmental conditions and occlusion. Therefore, RBIR offers challenges and opportunities to develop effective interactive mechanisms that use imperfect information to perform visual queries [14]. In our system, all images in the database are processed offline. The system first segments each image from different regions and then extracts the low-level color and structure characteristics of each region. Using Content Analysis and Indexing, C-Means Clustering Algorithm [12], the low-level characteristics of each region are associated with one of the high-level image concepts defined for the database. During recovery, the proposed system allows users to provide a query keyword or specify a region of interest in the query image [15]. In this subsection, we first define the similarity functions of two regions. Thus the similarity function for two images is defined on the basis of similarity in the regions [19].

Similarity between regions

The degree of similarity of the regions is calculated based on the values of the three properties that we register for each region. We define the similarity functions of these properties. In the following equations, R1 and R2 are two regions of two different images. Moreover, Ratio (Ri) and Size (Ri) indicate respectively the ratio and the size of the region i.

Shape:

$$Sim_{ratio}(R_1, R_2) = 1.0 - \frac{|Ratio(R_1) - Ratio(R_2)|}{Max(Ratio(R_1), Ratio(R_2))} \qquad (1)$$

Size:

$$Sim_{size}(R_1, R_2) = 1.0 - \frac{|Size(R_1) - Size(R_2)|}{Max(Size(R_1), Size(R_2))} \qquad (2)$$

3.1 Contrast enhanced image: (Pre-processing)

Image with data, images and functions already existing to get effective data, we manipulate through the use of software. Factor affecting image clarity noise, exposure range, color accuracy distortion, vignetting, lateral chromatic aberration, lens flare, color moiré. Preprocessing is a collection of operations for image detection, image classification and image retrieval[5]. Geometric grinding images resample or modify the pixel grid, so they look for the detected pixels. The previous processing collects the image, improves the image contrast and moves on the image fuses. Adapthisreq enhances the gray scale of image parts using adaptive equalization of contrast-restricted histograms (CLAHE)

3.2 COLOUR CORRELOGRAM IMAGE:

A chromatic correlogram (hereafter referred to as a correlogram) expresses how the spatial correlation of color changes with distance. A color histogram (hereafter, histogram) only captures the distribution of color in an image and does not include any spatial information. Therefore, the correlogram is a type of spatial extension of the histogram [18].The probability that the color will occur in the vicinity of the color j at the distance k

$$r_{i,j}^{k} = Pr\ (i,\ j/k) \qquad (3)$$

A chromatic correlation (as we follow from a correlative) expresses the spatial correlation of color changes with distance. A color histogram (hereafter, histogram) only captures the distribution of color in an image and does not include spatial information. Therefore, the correlogram is a type of spacial extension of the histogram [20, 24].

The probability of the color is to verify near the color at the distance k

$$_{o<j<1}\ |^{p1-p2|}\ r_{i,j}^{(k)} = \sum_{0\leq i\leq C(i)} [^{p2\text{€}ic(i)} = k]$$

Where i, j ∈ {1, 2, ..., N}, k∈ {1, 2, ..., d} and | p1 - p2 | is the distance between pixels p1 and p2.The correlogram is a variant of the histogram accounts for the local spatial correlation of colors. Find a pixel color at distance x, pixels i, in the image. The color histogram describes the color level and the corresponding pixels. Including global and local information requires the histogram based method . Table indexed by color i where the input d-th shows the probabilities of finding a pixel i from the same pixel at distance d. The correlogram with a small amount of color and distance will give good results by reducing the computational cost.

Let [D] be the set of fixed distances D {d1, d2. . . . . . . dD}. Thus, the correlogram ɤdci cj (I) of image I is defined for the pair of colors (ci, cj) at a distance d,

$$P^{d}_{CiCj}(I) = P1€ I_{Ci}^{Pr} P2 € I^{[p}{}_{2}€I\ Cj1/P1- \quad P2/=d] \quad (5)$$

A is a weighting matrix, whose elements aij correspond to the similarity of ci and cj h (Q) denotes the histogram of the image (Q) with the quantized color set (C). When one considers HSV, the similarity of colors is calculated as an inverse distance of three-dimensional color points resized in the chromatic space.

3.3 CORNERS IN THE IMAGE

3.3.1 FAST (Features from Accelerated Segment Test)

It is used to identify the point of interest in the images. Points of interest have a high content of local information.

Algorithm:

Input: image, radio, threshold, intensity

Output: image by marking the intersection points of the corners.

Step 1: Consider c as an interest

Step 2: Acquire 16 pixels with a radius of 3 pixels around c

Step 3: Check all 16 pixels for the intensity level

Step 4: Initialize the count and count 1 out of 0

Step 5: Check if the control value below 16 precedes more, otherwise go to step 10

Step 6: check if the intensity is greater than c added with the previous additional threshold; otherwise, continue with step 8

Step 7: Increase the count by 1 and go to step 10

Step 8: check if the intensity is less than c added with the previous additional threshold; otherwise, continue with step 10

Step 9: Increase Count1 by 1

Step 10: If the count and count 1 are added together, greater than or equal to 12, C is an angle

3.3.2 Harries

Recognizing the point through a small window, changing the window in any direction should make changes in intensity.

Algorithm:

Input: image, standard deviation, threshold, region radius (1-3)

Output: angles of image marking, coordinates of the row column.

Step 1: derived mask

Step 2: derived from the image

Step 3: Gaussian filter Noise-> w (x, y) = exp (- (x2 + y2 / 2σ2)

Step 4: derivatives of the polished square image.

Step 5: Measure the angle of Harries

Step 6: deletion and non-maximum threshold

3.4 NORMALIZATION IMAGE:

Normalization image used to change the pixel intensity value. It is also called contrast stretching or bargraph stretcher. Normalization transforms the dimensional gray scale.

Image I: {X ≤Rn} -> {min, max}

In a new image

IN: {X ≤Rn} -> {new minimum, new maximum}.

Linear normalization for the grayscale digital image

IN = (I-min) (new maximum - new min / maximum-minimum) + new Min (7)

Non-linear normalization for grayscale digital images

IN = (new max -new min) 1/1 + e-i-β / ɤ + new min (8)

The gradient function is a function used to derive a dimension of several dimensions. In image filtering, each pixel in the image takes the sum of the products. Each product we obtain is the color value of the current pixel and the adjacent pixel. In soft zero filtering, the masking sum in the digital image will be zero to be compared to the original image.

$$MASK=\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & -16 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

3.5. K-Means Clustering Algorithm:

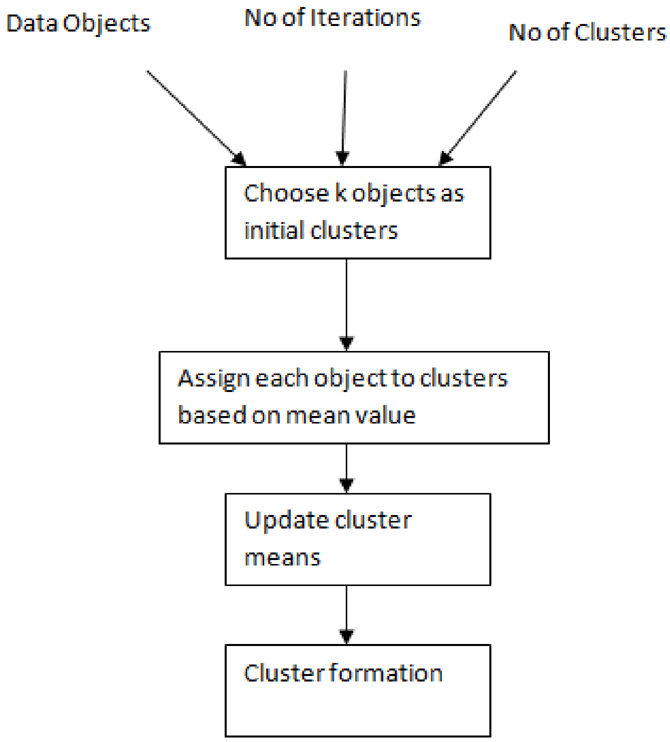

Fig.1. Cluster formation using k-means

K-means selects (k) algorithm the number of random objects, each object represents a group means. Each object is assigned, based on the distance between the object and the cluster average, to the cluster it is closest to. Then calculate a new average for each group. the algorithm repeats itself until reaching its goal. the example is shown inFig2.Defining grouping is about dividing or dividing input data points into groups so that data points within the same group have similar properties to each other, while data points in different groups have different properties. Different clustering methods are used for image clustering to sort, display and digitize images and improve the performance related to CBIR clustering techniques, annotation images, applications etc. The grouping algorithms can be divided into two methods: hierarchy and partitions. Hierarchical clustering algorithms are groups repeatedly as clustering where each data point in the cluster combines most of the pairs of similar clusters to form a cluster hierarchy or division mode, which is a top-down approach, in which all data points in a cluster it is subsequently divided into small groups. Partition clustering algorithms are different hierarchical clustering algorithms where all clusters are as a partition and do not form hierarchical way. A single link and a complete link are examples of hierarchical grouping algorithms. K-means is a type of non-supervised learning algorithm that is used to solve the grouping algorithm. K-means follows a simple procedure in which the data set provided in a fixed number of clustered (cluster k) classified. The first step is to establish k centers. The centers must be formed in such a way that the different position of the group can change the final result. The best option is to keep them away from each other. The next step is to assign data points to the nearest center. When the data points are finished, cluster formation is completed. At this stage, we must reassign k centroids as the central point of the groups formed by the previous step. After obtaining the k-centroids, the same data points are assigned to their nearest new data centers. The cycle is formed. The result of the cycle causes the centers to change their current position after each step until no further changes are made or the centers cannot be moved[3].

K-media clustering algorithm: 1. determines the initial centroids coordinate. 2. Calculate the distance of each data point in the centroids. 3. Group data points based on their minimum distance. (Find the nearest centroids) The K-means cluster is a very efficient and powerful algorithm for managing large data sets. It helps to restore images faster and also allows you to search for more relevant images in large image databases[4]. The K-means clustering algorithm is an expert in generating correct results for image recovery problems Using k-means, the user can capture the nearest group of images so that the result output is fast. [28]
Randomly assigning each object sample to one of the k groups S (j), 1 <= j <= k

calculate the center c (j) for each group S (j)
while (clusters are not stable)
for each sample Sample (s)
calculate the distances between Sample (s) and all the centers c (j)
find j * such that c (j *) is closest to the sample (i)
Assign the sample (s) to the S group (j *)
recalculate the modified cluster centers
final for
finish while

Fig. 2.K-media clustering algorithm

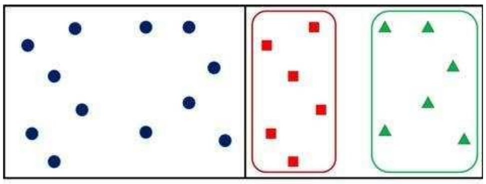

Fig. 3 An optimal partition in groups of a set of points in a Cartesian space: (a) the points are not yet assigned to any cluster; (b) points belonging to the same group are marked with the same symbol.

Many variations of the basic k-means algorithm have been proposed over the years. The most popular implementation of this algorithm is the Lloyd algorithm. The k-means and h-means? The algorithms are variants of the k-means and h-means algorithms, respectively, in which empty clusters are not allowed.

In effect, k-means and h-means can provide solutions containing empty clusters as partitions. This is usually avoided if it is known that the desired number of clusters is exactly equal to k. The algorithm of the means Y is designed to be independent of the value of the parameter k. In this variant, the value of k is adjusted during the execution of the algorithm. Other variants of the basic algorithm include the Jmeans algorithm and the k-means genetic algorithm, in which the algorithm's performance is improved.

Note that the k-means algorithm does not estimate cluster covariance; only considerthe average of conglomerates. In we considered a difficult assignment version of the Gauss mixture model with general covariance matrices, known as elliptic k-means algorithms.

The nearest close k approach

The near neighbor k (k-NN) is a classification technique. A training set is known, used to classify samples of unknown classification. The basic assumption in the k-NN algorithm is that similar samples should have a similar classification. As in the k-means approach, the similarities between samples are measured using appropriate remote functions. A schema of the k-NN algorithm is shown inFig3. Parameter k shows the number of similar known samples used to assign a classification to an unknown sample. Given an unknown sample, its distances are calculated from all samples of the training set and the nearest k samples are localized. Hence, the most frequent classification among the known nearby samples is assigned to the unknown sample (see Fig. 4).

The k-NN method provides a very simple classification rule, but it can be quite expensive to perform. For each unknown sample, it is necessary to calculate their distances with respect to all known samples and this procedure can have a high computational cost. The k-NN method uses the information in the training set, but does not extract

for all unknown samples UnSample (i)
for all known samples Sample (j)
calculate the distance between UnSamples (i) and Sample (j)
final for
find the smallest distances k
identify the corresponding samples Sample (ji ... .jk)
assign UnSample (i) to the class that appears most frequently
final for

Fig. 4. The k-NN algorithm

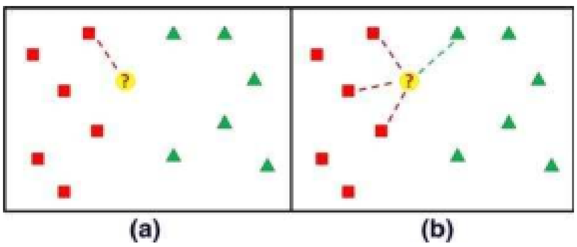

Fig.5 The point marked by the symbol? it is classified according to the classification of its nearest neighbors: (a) k = 1 and the unknown point is classified as belonging to the class marked by squares; (b) k = 4 and the unknown point is classified as belonging to the class marked by squares, as well as extracted from any rule or law to classify other data.

Rather, it reuses the entire training set whenever it is necessary to classify an unknown sample. For this reason, many techniques have been developed with the aim of reducing the training set to the minimum number of samples that maintains the quality of the classification performed by k-NN.For example, a Neighbor Condensed rule was introduced with the aim of reducing the computational cost of the k-NN algorithm. Let T be a set of available training. Instead of using T NN, one of its subsets, T CNN NN, can be used to decrease the number of distances calculated in each step of the algorithm. In the event that T CNN can correctly classify each sample in the TT CNN set, then it is considered a consistent subset of T. In order to keep the classification quality high, when a training set is replaced by a subset T NN CNN, the coherence T. In other words, all known samples in CNN T that are not in CNN T NN must be correctly classified by k-NN when the training set is T. Similarly, the Neighbor rule The smaller can smaller, in general, generates reduced training sets T CNN RNN smaller than condensed T sets. Other strategies for this purpose have been proposed in the literature, and examples are provided. Finally, keep in mind that another way to speed up k-NN is to speed up the sample matching process, avoiding distance calculations. [31]

3.6. Content Analysis and Indexing:

The user enters the image query in the database and the restore is performed with the help of an indexing scheme. The indexing scheme provides an efficient way to search the image database [16]. This can be seen in a form of loud presentation. This shows the analysis and indexing scheme.

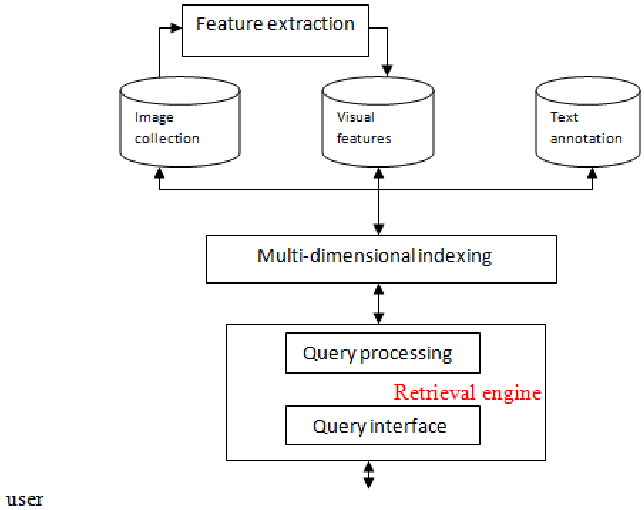

Fig.6.Example of multimodal retrieval system for image mining

Image indexing and image retrieval In order to be able to recover images from databases efficiently, proper indexing is required. Relational databases provide indexes based on primary and secondary keys. This approach is not applicable when image databases are extracted, since image recovery is generally based on similarities. KDB tree, tree R, tree R *, tree R +, tree SR, TV tree, tree X and iMiniMax are the indexing methods most commonly used as described Query for associated attributes - Image restorebased on the attributes stored as metadata _ Query for Description - Description of the contextkeywords assigned to images (for example, in file names) _ Content queries - Organizes images based on their visual content (based on the characteristics detected, such as texture, shape, color, based on similarity, etc. Many subsequent applications focus on the combination of the above methods to allow a more specific and convenient search for particular data.

Multimodal recovery has been proposed to handle different types of unstructured data simultaneously, including image, video, audio and text [22]. The proposed algorithm allows data to be recovered based on visual characteristics and text patterns. [23] The framework that implements deep learning architecture has been proposed by [23] as a tool to improve the accuracy of image recovery in the management of medical image data has also developed a text-based indexing system for image recovery and mammography classification. The extraction of accurate information from a large amount of data is obtained by relying on the Naïve Bayesian classifier. [30]

4. Experimental Results and Discussion

In this the RBIR plays the main role in the concept of image recovery. With the help of RBIR, the color correlogram and the analysis and indexing of the content managed all the problems created by the CBIR [11]and the color histogram. Our proposed techniques help to address the problems of our existing techniques. In this sense, the RBIR supports both the keyword query and the query by region of interest, from which the text annotation deals with and successfully manages. The figures below show the experimental results of RBIR using C-means clustering.

India is an agricultural country; where about 70% of the population depends on agriculture. Farmers have a wide range of diversity to select suitable crops. However, cultivation of the crop for optimal yield and a quality product is highly technical. It can be improved with the help of technological support. The management of perennial crops requires rigorous monitoring, especially for the management of diseases that can significantly influence production and subsequently life after harvest.

Image processing can be used in agricultural applications for the following purposes:

1. Detect sick leaves and stems
2. Quantify the area affected by the disease.
3. To find the shape of the affected area.
4. Determine the color of the affected area
5. To determine the size and shape of the crop. Etc

Fig.7. Symptoms of bacterial pod disease for crops.

1. Bladder disease of the vessel usually appears in the last stages of plant growth.
2. The initial symptoms usually develop as the lower leaf pods lesions near the waterline with the plants are being delayed or early internodes (about 10 to 15 days after flooding) other. The fungus affects the crop from the start-up stage. The initial symptoms are observed in the leaf sheaths near the water level. In the leaf sheath are greenish oval or elliptic or irregular gray spots. As the points get bigger, the center becomes a blackish brown irregular border brown purple or grayishwhite. Injuries in the upper parts of the plants quickly spread fusion together with covering whole farmers from floating to the flag leaf. The presence of numerous large lesions on a leaf sheath usually causes death of the entire sheet and, in severe cases; all the leaves of a plant can be affected.

Fig.8. Symptoms of Crop Crops shell-eating disease (close-up view)

The fungus produces the sepal mycelium which is hyaline when young, yellowish brown when it was old. It produces a large amount of spherical brown clerotia.

Fig.9. Symptoms of bacterial pod disease in ricecrops

Fig.10. Symptoms of the bacterial pod disease in Paddy crops.

The rice stem rot (picture 4) the other disease that acts as an indicator of potassium deficiency. We also believe that a low level of potassium in the soil may play a role in increasing the severity of the bacterial plague of rice panicles. Previous research has indicated that Cercospora leaf, downy mildew, putrefaction and bursting can be observed more frequently in fields with excessive nitrogen fertilization.

Fig.11. **A Study on Paddy crops Disease Prediction Using k-means RBIRTechniques**

The mean and standard deviation are important factors for analyzing the basic information of the pixel of the image[10], when we consider the base value of the pixel, the variation of the pixel value will be high if we consider that the solution of the original value will not be effective, therefore you should find the average value obtained by summing the total number of pixels and divided by the total number of pixels. the standard deviation is detected by the calculated mean value subtracted from the observed value, then squared of each difference and adding that the column divides by the total number of samples subtracted from a resulting value is a variance that assumes the square root of the variance we obtain the use deviation standard to find the average propagation of data. Here the meaning of the pixel can be effective.

Query image	**Paddy**	**Tomato**	**Cotton**	**Corn**	**Onion**
Mean value	0.1329	0.3933	0.3322	0.1286	0.3227
Standard value	0.2427	0.1724	0.3170	0.1780	0.2292
Mean Color count	31.2031	31.3334	32.4766	28.3984	31.985
StandardColour count	31.7982	30.9845	30.2084	32.6116	33.873

Table 1: Mean and standard deviation of pixel values

Above table reprsent various mean value and standard valve has entered from that value colour mean and colour standard deviation has been calculated from the graph we can calculate the relationship between same and different group of images . If colour between various region is standared and colour in between the region is same then deviation value will be high.

Contrast level is mainly used for have standared visual effect of the image, contarct level is low then image will appear darken then if contrast level is high then we will have light image have effect of white screen , so contrast level should be maintained to increase visibility of the image result in good result.center of region has meaning of interest point which has similar point in an region which help us to find an common area. Correlation used to prove the availability of required data if we know the similarity then availability can be easily checked if two pixel vaue is correlated we can predict one after other.energy leads to know the strength of the image , the image strength can be found how effective the data is used in real world if strength is higher then effect of data is good so the retrival data will be effective.

Homogeneity means same can be define as local homogeneity and total homogeneity. Local homogeneity is used to find the similarity in between the region; image has been divided in to various regions to find the homogeneity of the image pixel level comparison is made if pixel level comparison is same then region level comparison is performed if both are same then we can saw the local homogeneity has obtained, after homogeneity checked between all region of the image if all or maximum region is same then we can obtain the total homogeneity.

Query image	Paddy	Tomato	Cotton	Corn	Onion
Contrast	0.4137	0.3740	0.2793	0.3265	0.4734
CoR	0.9624	0.9179	0.9727	0.9699	0.9905
Correlation	0.8282	0.7321	0.8276	0.8275	0.9706
Energy	0.1734	0.2462	0.2950	0.2664	0.2166
Homogeneity	0.8924	0.8730	0.9204	0.9410	0.9170
Entropy	6.9203	6.3285	6.3107	6.1683	6.1805

Table 2. Training data

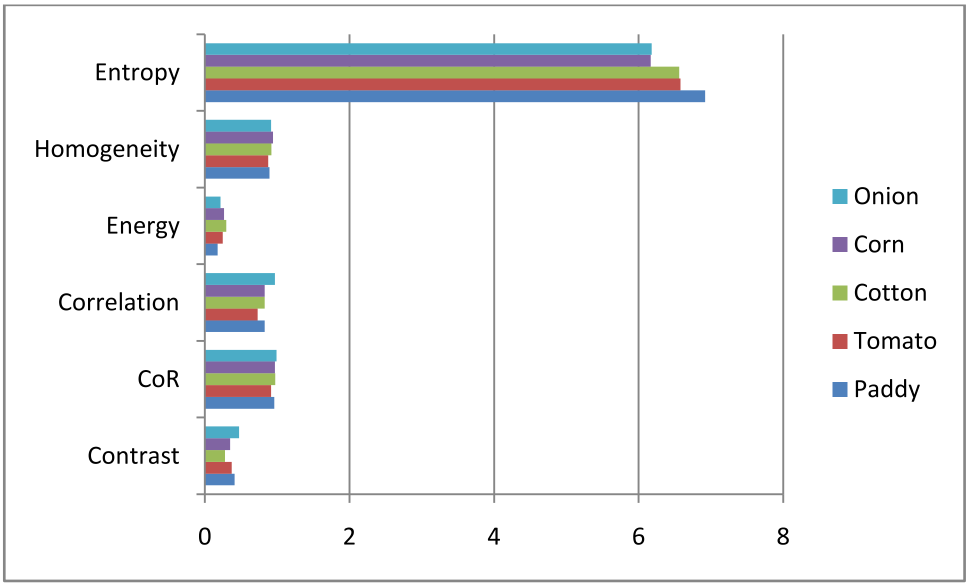

Fig.12.Relationship between various pixels and inter pixel strength

When considering the contrast, the center of the region (COR), the correlation, the energy and the homogeneity of an image, we can know the complete information of an image, both the low-level and the high-level characteristics. level. Having the low-level function, we can predict the similarity between different regions, which will help to increase the recovery rate.

The training data considered infigure4 have COR as a similar value, the correlation is good for analyzing neighboring data and the value has been analyzed in a cluster that has less homogeneity and good entropy for an image[13].

The cluster is nothing more than grouping data in a relationship and has a link to all surrounding data. Here the whole image is taken and stored in a database, then a grouping with the image is created. This similar image will be grouped by common reference name when the image of the query is checked by similarity; therefore, a particular group with similarity will check before increasing performance.

Figure 7 shows the characteristics, successes and failures of the various images grouped together for the recovery of the k-means algorithms that are connected to all the grouped images. During image recovery, the characteristics of the entire cluster were searched and images of similar characteristics were retrieved. Then share the result of the similar image for the image of the queryFig2.

3rd National Conference on Innovative Research Trends in Computer Science and Technology (NCIRCST 2018) ISSN: 2454-4248
Volume: 4 Issue: 3 126 – 140

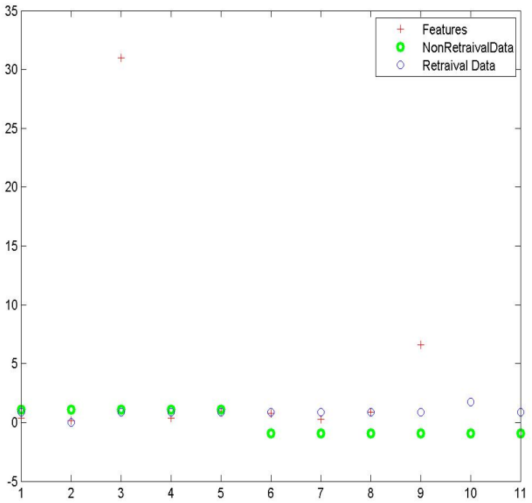

Fig.13. Cluster Retrieval and Non Retrieval Representation

Query image	Total Retrieved Images	Relevant Retrieved Images	Irrelevant Retrieved Images	Precision	Recall
Paddy	75	71	4	.94	.72
Tomato	75	68	7	.90	.69
Cotton	75	75	0	1	.75
Corn	75	32	18	.76	.34
Onion	75	69	16	.92	.70

Table.3. image Retrieval Results for Different Query Images

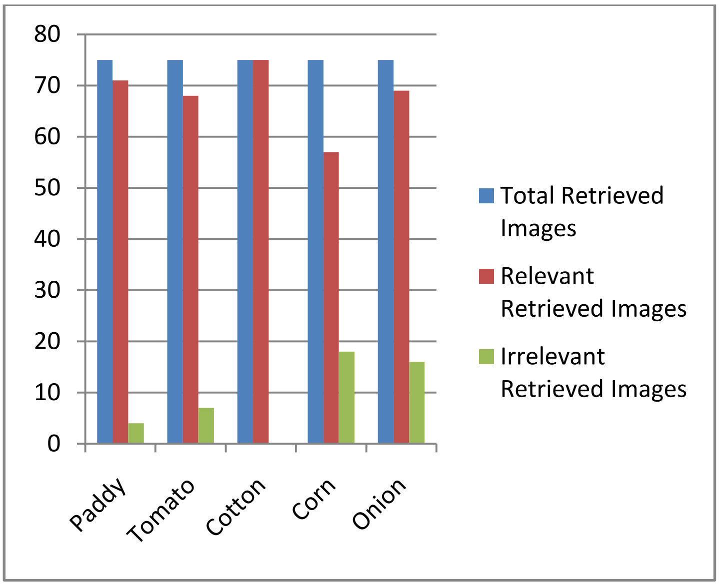

Fig.14. Accuracy of relevant and irrelevant retrieved from total image by K-means RBIM

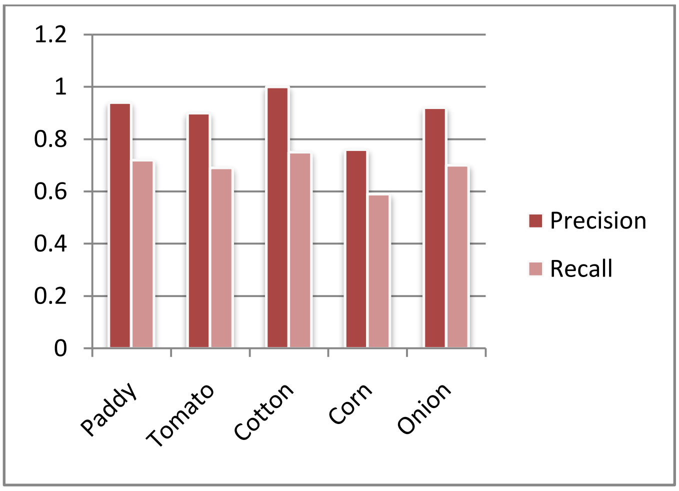

Fig.15. Relationship between Precision and Recall

Query image	Total relevant images	Total retrieved images	Computational time (sec)
Paddy	80	75	.836
Tomato	80	75	.748
Cotton	80	75	.482
Corn	80	75	.731
Onion	80	75	.831

Table .4. Retrieval time for different query images

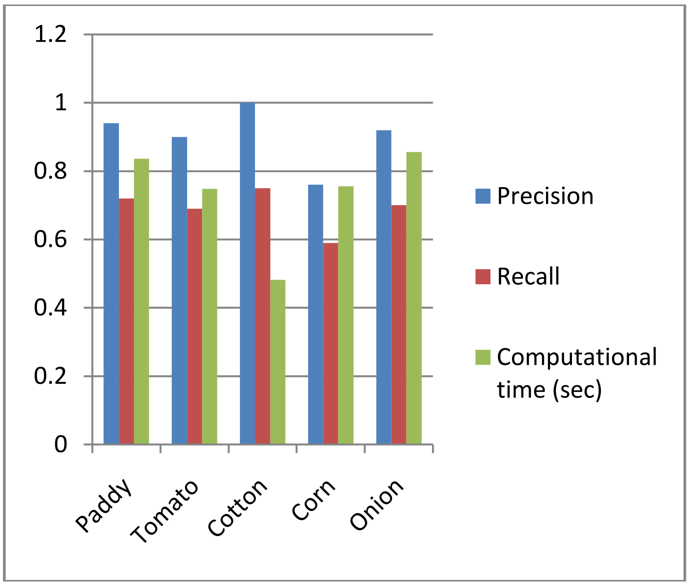

Fig.16. Comparison of precision, recall with computation time

Query image	RBIR Based SVM Classification using K-means algorithm		CBIR Based SVM Classification using K-means algorithm		Based on colour+ Edge Features	
	Precision	Recall	Precision	Recall	Precision	Recall
Paddy	.94	.72	.93	.70	.50	.20
Tomato	.90	.69	.80	.73	.45	.39
Cotton	1	.75	.72	.33	.30	.12
Corn	.76	.34	.82	.76	.36	.18
Onion	.92	.70	.91	.65	.48	.26

Table .5 .Comparison of Proposed Method with Retrieval Results of Image Based on colour+ Edge Features[32]

Performance	**Access time**	**Sensitivity**	**Specificity**
Paddy	98.8146	71.8172	84.783
Tomato	98.2	71.682	83.7779
Cotton	98.7	72.2234	84.264
Corn	98.2648	72.1693	83.8107
Onion	97.4005	71.7372	82.6922

Table 6.Performance of RBIR with Sensitivity and Specificity.

Conclusion

The CBIR requires manual annotation of the text of each image in a database, but in textual annotation it is inadequate and ambiguous to search for image databases and is also a dependent language. These problems can be managed with the help of techniques such as RBIR, Content Indexing and Analysis, Color Correlogram, and C-Means Grouping Algorithm. In this sense, the RBIR supports both the keyword query and the query by region of interest, from which the text annotation deals with and successfully manages. The queries entered by the user are analyzed based on the analysis of the content and the analyzed queries are retrieved from the database with the help of the indexing method. From this, the spatial information of a color image is managed with the help of a correlogram color technique and, finally, the image center is located from the database with the help of the k-means algorithm.

The study is done in this paper provides a new insight into the detection of plant diseases. The purpose of the research in this field is as follows: There are two main characteristics of the detection of plant diseases using automatic learning methods that need to be achieved: speed and accuracy. Therefore, there is a margin to work on developing innovative, efficient and rapid interpretation algorithms that will help the plant scientist to detect diseases. You can work to automatically estimate the severity of the detected disease. The main result of this project is an software tool to find and cure crop diseases. Benefits for farmers,in the

technique, two new algorithms for image analysis will be proposed and the contrast, COR, correlation, energy, homogeneity, entropy between database attributes will be searched. Easy access to agricultural services specialized in rural, sub-urban, semi-urban and remote areas, Early diagnosis and rapid treatment, Visits reduced to the fields, Reduction of travel expenses
and Reduction of the burden of the disease.

References:

[1] Georgios Th. Papadopoulos, Member, IEEE, Konstantinos C. Apostolakis, and Petros Daras, Senior Member, IEEE "Gaze-Based Relevance Feedback for Realizing Region-Based Image Retrieval " IEEE Transactions On Multimedia, Vol. 16, No. 2, February 2014.

[2] Feng Jing, Mingjing Li, Member, IEEE, Hong-Jiang Zhang, Fellow, IEEE, and Bo Zhang , "An Efficient and Effective Region-Based Image Retrieval Framework", IEEE Transactions On Image Processing, Vol. 13, No. 5, May 2004.

[3] Simone A.Ludwig' "Map reduce – Based k-means clustering algorithm: implementation and scalability" Springer 20 April 2015.

[4] MAO li, song yi-chun, Liyin, Yang hong, Xiaower, "Research of improved K-means Algorithm based on new metric norm", Springer 2015.

[5] Mohsen Zand , Shyamala Doraisamy, Alfian Abdul Halin, Mas Rina Mustaffa , "Texture classification and discrimination for region-based image Retrieval " J. Vis. Commun. Image R. 26 Elsevier (2015) 305–316.

[6] Keng-Pei Lin and Ming-Syan Chen, Fellow, IEEE , "On the Design and Analysis of the Privacy-Preserving SVM Classifier" , IEEE Transactions On Knowledge And Data Engineering, Vol. 23, No. 11, November 2011.

[7] Lorenzo Bruzzone, Senior Member, IEEE, and Claudio Persello, Student Member, IEEE , "A Novel Context-Sensitive Semisupervised SVM Classifier Robust to Mislabeled Training Samples " IEEE Transactions On Geoscience And Remote Sensing, Vol. 47, No. 7, July 2009.

[8] Atony Fierro- Radilla , Karina perez-Daniel , Z Nakano-Miyatakea Hector Perez-meana, and Jenny Benosis-Pineau' "An Effective visual descriptor Based on Color and Shape features for image retrieval " springer 2014.

[9] Ahmed Talib Mussudi Mashmoddin, Husmiza Husni and Loay E.George , "Dominent colour based indexing method for Fast Content –Based image retrieval" . Springer 2014.

[10] Shahana N Youseph ,Rajesh Roy Cherian, "Pixel and Edge Based Illuminant Color Estimation for Image Forgery Detection" Elsevier (2014).

[11] Samuel Barrett "Content-Based Image Retrieval: A Short-Term and Long-Term Learning Approach", Stevens Institute of Technology, 2007.

[12] Bassam M. El-Zaghmouri and A. Abu-Zanona "Fuzzy C-Mean Clustering Algorithm Modification and Adaptation for Applications", ISSN: 2221-0741, Vol. 2, No. 1, 42-45, 2012.

[13] J. C. Dunn "A Fuzzy Relative of the Isodata Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics 3: 32-32, 1973.

[14] byoungchul ko, Jing Peng, Hyeran Byun "Region based image retrieval using probabilistic feature Relevance Learning", 2008.

[15] Ying Liu, Dengsheng Zhang, Guojun Lu "Region-based image retrieval with high-level semantics using decision tree learning", 2529 – 2320, 2007.

[16] Dr. Fuhui Long, Dr. Hongjiang Zhang and Prof. David Dagan Feng ,"Fundamentals of Content-Based Image Retrieval", 2003.

[17] Srinivasa Kumar Devireddy,"Content Based Image Retrieval", Computer Science and Telecommunications 2009.

[18] Debashis Debnath, Dr. Ranjan Parekh, "Content Based Image Retrieval Using Directional Colour Correlogram", ISSN : 0975-2962, Vol. 3 No. 6 June 2011.

[19] Niket Amoda and Ramesh K Kulkarni,"Efficient Image Retrieval Using Region Based Image Retrieval" , SIPIJ Vol.4, No.3, June 2013.

[20] Yossi Rubner,∗Jan Puzicha,† Carlo Tomasi, and Joachim M. Buhmann† " Empirical Evaluation of Dissimilarity Measuresfor Color and Texture Computer Vision and Image Understanding" 84, 25–43 Elsevier (2001)

[21] Fei Li, Qionghai Dai, Senior Member, IEEE, Wenli Xu, and Guihua Er ," Multilabel Neighborhood Propagation for Region-Based Image Retrieval" IEEE Transactions On Multimedia, Vol. 10, No. 8, December 2008

[22] Yong Rui and Thomas S. Huang, "Image Retrieval: Current Techniques, Promising Directions, and Open Issues", Journal of Visual Communication and Image Representation 10, 39–62 (1999)

[23] .B. MiloviC , B. MiloviC and v. RadojeviC, "Application of DATA Mining in Agriculture", Bulgarian Journal of Agricultural Science, 21 (No 1), 26-34 (2015)

[24] Thomas M. Deserno, Sameer Antani and Rodney Long, " Gaps in content-based image retrieval" , Aachen University of Technology (RWTH), 27032 Aachen, Germany (2009)

[25] A.S Deokar, Akshay Pophale , Swapnil Patil , Prajakta Nazarkar, "Plant Disease Identification using Content Based Image Retrieval Technique Based on Android System" , International Advanced Research Journal in Science, Engineering and Technology Vol. 3, Issue 2 (2016)

[26] Dr.Hari Ramakrishna, T.S.Praveena, Prof. N.Rama Devi, "segmentation based image mining for agriculture", Chaitanya Bharathi Institute of Technology Gandipet-Hyderabad, Vol. 1 Issue 6 (2012)

[27] Barbora Zahradnikova, Sona Duchovicova and Peter Schreiber , "Image Mining: Review and New Challenges", International Journal of Advanced Computer Science and Applications, Vol. 6, No. 7, (2015).

[28] Dr P Jaganathan, S.Vinothini & P.Backialakshmi, "A Study of Data Mining Techniques to Agriculture , International Journal of Research in Information Technology, Volume 2, Issue 4, Pg: 306- 313 (2014).

[29] M. Yasmin, M. Sharif, I. Irum, S. Mohsin, "An Efficient Content Based Image Retrieval using EI Classification and Color Features" Journal of Applied Research and Technology , Vol.12 , 877-885 (2014)

[30] Georg R., Rudolf K., Martin S., Peter W. Data Miningwith Neural Networks for Wheat Yield Prediction. P. Perner(ed.), Springer LNAI, Vol. 5077, pp. 47-31 (2008).

[31] Shekh, A.M., Patel, H.R., and Pandey, V. (1999), Predictingthe outbreak and spread of pests and diseases using weatherparameters. Dept. of Agri. Meteorology, Gujarat Agriculture University Anand, Annual Research Report, Vol.2, pp. 249230.