RESEARCHARTICLE OPENACCESS

Using Image Processing and Big Data Mining Methods jointly to Retrieve Images

Ch Aruna Reddy[1], B Nirmala[2],G Vidyulatha[3] , Dr.T.Charan Singh[4]

[1,3] (Assistant Professor, Dept of CSE, Sree Dattha Institute of Engineering and Science)

[2](Assistant Professor, Dept of CSE,Holymary Institute of Technology and science)

[4](Assoc. Professor, Dept of CSE, Sri Indu College Of Engineering & Technology (A))

Abstract:A novel method in the realm of image processing is mining. The process of extracting latent data from photos, combining image data, and identifying excess patterns that are hidden in images is known as image mining. This includes image processing, data mining, and other related processes. All significant patterns can be created without warning. The main topic of this research study is knowledge extraction from a large database. Information is conveyed by either direct or indirect means. Neural networks, grouping, correlation, and association are some of these techniques. This essay describes the applications of data mining in the industries of marketing, manufacturing, fraud detection, telecommunication, and education. We may employ an image's size, texture, and dominating color characteristics by using this technique. A feature called the Gray Level Co-Occurrence Matrix (GLCM) [1] determines an object's texture.

Keywords: Cloud computing, knowledge discovery databases, data mining, feature extraction, image retrieval, clustering, and gray level co-occurrence matrices.

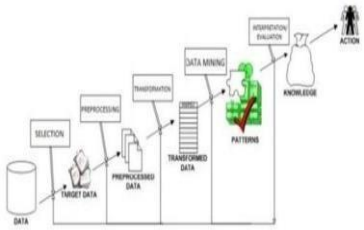

Fig.1.Knowledge DataMining

I. INTRODUCTION

Massive amounts of data are found in the real world in many different fields, including education, business, and medicine. These statistics might provide insight and information to help with decision-making. For example, we may find the sales information in the shopping databases and identify the dropout rate among students at any institution or university. To meet the problems, these data might be analyzed, condensed, or comprehended. Discovering remarkable patterns from vast amounts of data and understanding the data held in different databases, such as warehouses [2], the World Wide Web, and external sources, are the key concepts of data mining [3]. The idea of the pattern is to comprehend potentially legitimate and unknown data. One type of sorting method used to uncover hidden patterns is called data mining. Their objectives have passed.

- **Selection:**Selectdatafromvariousresourceswhere operationtobeperformed.
- **Preprocessing**:Alsoknownasdatacleaninginwhi chremove theunwanteddata.
- **Transformation:**Transform/consolidateintoane w formatfor processing [4].
- **Datamining:** Identifythedesireresult.
- **Interpretation/Evaluation:** Interpret the result /query to give meaning full information.

Knowledge discovery from databases is intended to be accomplished by a variety of methods and approaches, including classification, clustering, regression, artificial intelligence, neural networks [5] , association rules, decision trees, genetic algorithms, nearest neighbor method, etc.[6].This paper's primary goal is to learn about data mining, and the remaining portion of Section 2 addresses models and methods for data mining. Data mining's application is examined in Section 3. In Section 4, we finally bring the paper to a clos

IMAGEMINING

The main goal of this is to look for and find the valid concealed data. The various image mining system processes are displayed in the above diagram (Fig. 1). Other techniques are also employed to acquire information. They are artificial intelligence [7], data mining, image processing, and image retrieval. Two distinct techniques to image mining are made possible by the methods. Extracting from databases or photos is the initial step. The second step is mining the pictures or alphanumeric data. The feature extraction diminishes dimensionally in this instance. The input data will be transformed [8] into a smaller set of features if it can be accessed more often and isn't likely to be repetitive. It reduces the amount of resources required to find a lot of data in an understandable way. Numerous other characteristics are employed.

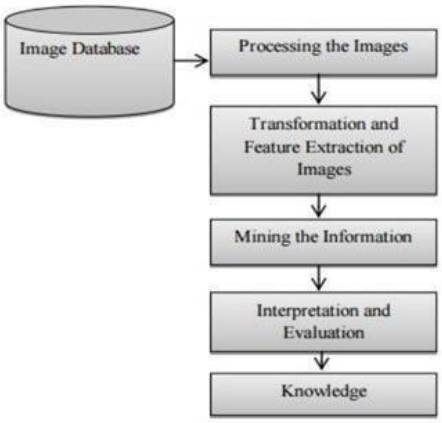

Fig.2.Image Mining Process

II. FEATUREEXTRACTION

Typically, feature extraction faces significant challenges in object detection; however, the Genetic Algorithm (GA) [13] provides a simple, robust framework for identifying superior feature sets. Lower detection error rates are the result. Zehange Sun et al. argue in favor of applying Principle Component Analysis (PCA) and Support Vector Machines (SVMs) for classification. GA can therefore eliminate the detection of undesirable aspects. The two challenging items detection challenges with the approaches are face and vehicle detection. They improve both systems' performance by classifying data using Support Vector Machines. Patricia G. Foschi states that feature extraction and selection constitute the image pre-processing stage.

A. Mining [9] It's an important step. Mining involves taking pictures and using them to derive patterns. Finding the best ones is thegoal. According to Broun, Ross A. et al., , the design of a mining system is supported by the necessity of digital picture forensics. A hierarchical SVM [10] can be used to teach it to recognize things. In general, image mining is concerned with the research and development of novel technologies. Not only should pertinent photos be found again, but new image patterns should also be created. Mining costs can be decreased by leveraging a natural source of parallelism, as demonstrated by Fernandez.J et al.,]. To enhance their quality, the pre-processed photos are first taken from the database.

B. Color Features

Because of the richness of the data displayed, image mining produces distinctive qualities. The performance metrics are necessary for the evaluation outcome. An evaluation for comparing the function by color is cited by Aura Conci et al., The technique is demonstrated by color affinity mining experiments using quantization [11] on color space and similarity measures. A quick and efficient way to index picture metadatabases was suggested by LukazKobylinski and Krzysztof Walczak An index is created based on their color attributes. A meta database index can be created using the Binary Thresholded Histogram (BTH) color feature description approach. It has been demonstrated that the BTH is an adequate technique for displaying picture databaseproperties.

C. For picture mining, Ji Zhang, Wynne Hsn, and Mong Li Lee [8] suggested an efficient information-driven approach [12] They separate]d into four stages.

D. TextureFeature

The color histogram texture is the basis for how humans see images. Human neurons contain 1012 bits of information, and the brain, which receives information from the sense organs such as the eyes and interprets it, is the source of all knowledge. Rajshree S. Dubey et al.state that the color histogram and image texture serve as the foundation for mining photos. According to Janani, M. and Dr. Manicka Chezian,, image mining is a crucial technique for mining

knowledgefromImages.Thisisbasedonthecontentbase dImageRetrievalsystem.Color,texture,pattern and shape of objects are the basis of visual content.

E. ShapeFeature

According to Peter Stanchev, a novel technique for using image mining to extract low level color, shape, and texture into high level semantic features has been proposed. A theory by Johannes Itten is presented for obtaining high level form features. visual retrieval, according to Harini D.N.D. and Dr. Lalitha Bhaskari D., is just revealing basic level pixel representation in order to identify high level visual objects and their associations [22, 23].

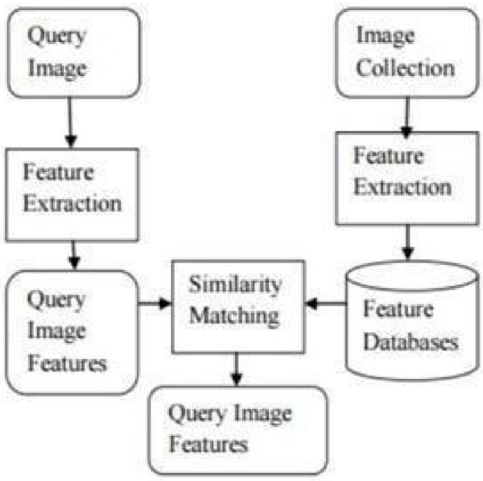

Fig.3.Content Based Image Retrieval System Architecture

III. METHODOLOGY

The gray-levelco-occurrence matrix (GLCM) [14] considers the relationship of pixels. This calculateshow often the pairs of pairs of pixel with specificvalues and in a specified spatial relationship in animage.

UnderstandingaGray-LevelCo-OccurrenceMatrix

To create a GLCM, we utilize the gray comatrix function. By figuring out how frequently a pixel with the intensity (Gray Level) value i appears in a preset, it generates GLCM [15]. Every element (i,j) in the input image is the sum of the pixels with values i and j that occur in the designated spatial relationship. Scaling is used by Graycomatix to minimize
The number of intensity values.The Num levels and The Gray Limits control this scaling of gray level.Let us understand the process through the Following diagram. The following figures explain show gray comatrix calculates the first three values in a GLCM.

The following figure demonstrates how to compute a graycomatrix for the first three values in a GLCM. Because there is just one occurrence in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively, element (1,1) in the output GLCM carries the value 1. Because there are two occasions where two horizontally adjacent pixels have the values 1 and 2, glcm(1,2) includes the value 2. Because there are no examples of two horizontally adjacent pixels with the values, element (1,3) in the GLCM has the value 0. additionally 3.Graycomatrix keeps analyzing the input image, looking for further pixel pairings (i, j), and logging the sums in the relevant GLCM elements..

Process Used to Create the GLCM

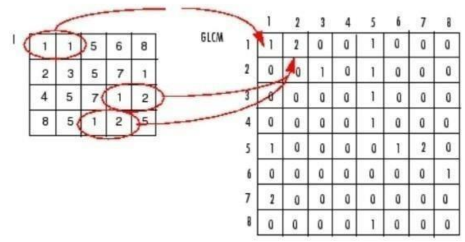

Fig.4.Processused to create the GLCM

SpecifyOffsetUsedinGLCMCalculation

By default, the graycomatrix with offset as two horizontally adjacent pixels produces a single GLCM. It's possible that a single GLCM is insufficient to capture the input image's texture features. It's possible that a single offset is not texture-sensitive. As a result, graycomatrix can create several GLCM from a single input image. Multiple GLCM to graycomatrix functions are produced by the offsets. They primarily describe four distances and the pixel relationships of three different directions (horizontal, vertical, and two diagonals). 16 GLCMs display the input image in this manner. We are able to compute statistics from these GLCMs and determine the average.

WeightedEuclideanDistance

The standardized Euclidean distance betweentwoJ-dimensionalvectorscanbewrittenas:

$$=(\sum_{j=1}(i- \frac{}{j} \frac{j}{j})^2$$

Where s_j is the sample standard deviation of the j-thvariable. Notice that we need not subtract the j-th mean from $x_j$andy_j because they will just cancel out in the differencing. Now (1.1) can berewritteninthefollowingequivalentway:

$$, = (\sum_{j=1} \frac{1}{2}_j (\,{}_j -)^2{}_j)$$

$$= (\sum_{j=1} w(_j{}_j -)^2{}_j)$$

Where $w_j = 1/s_{j2}$is the inverse of the j-th variance $w_j$as a weight attached to the j-th variable: in otherwords

IV. DATAMININGTECHNIQUES

Retrieving pertinent information from disorganized data is known as data mining. Thus, it aids in achieving particular goals. Its sole objective is to develop a predictive model or a descriptive model. A prediction enables the data miner to forecast unknown (often future) values of a particular or target variable, while a description relates to the primary attributes.[18] Their objectives are essentially to employ a range of data mining methods, as seen in figure 5[8].

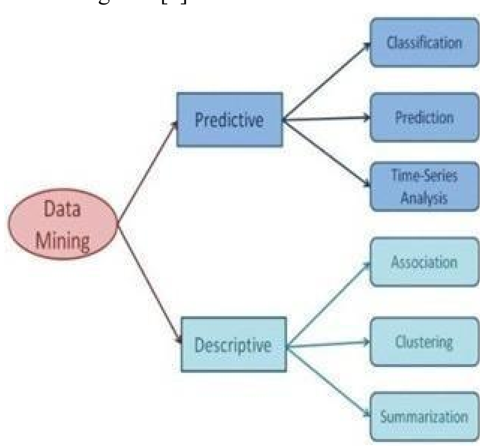

Fig.5. Data MiningModels

3.1Classification: Discrete and unordered are its foundations. This is predicated on the intended result. The training set and values are the basis for the data classification. Neural networks, classification rules (If-Then), and decision trees are used to accomplish these objectives. For example, we can examine the former academic records of the students who departed for university using this rule. This aids in determining the kids' performance.

3.2 Regression: It is utilized to map a data part to are al valued prediction variableItcanbeusedforpredictiontoo.Here,thetargetvaluesareknown,forexample,wecanpredictthechildbehaviorbasedonfamilyhistory.

3.3 TimeSeriesAnalysis: Thisprocessusedthe statisticaltechniquestomodel.Itexplainsatime dependent series of data points. It is a method ofusing a model to create prediction (forecasts) for future happenings based on known past events.[16] Stock market is a good example

.

3.4 Prediction: With the use of this technique, one may ascertain the link between independent and dependent variables.[17] Continuous or ordered value is the foundation of this concept.

3.5 Clustering: This is used in the relatively new field of "Education Data Mining." This facilitates comprehension of behavior, course selection, dropout rates, and student achievement. Higher education is a field that heavily utilizes it.[10, 22]

3.6 Summarization: This is data abstraction. It is composed of several connected tasks. It offers a summary of the data. For instance, a long-distance running event can have its overall time cut down to minutes or seconds. An additional well-known method for data mining is the association rule. The most frequent item set is found. It finds associations between items in the same transaction by pattern-matching data. Because of the way it links the sets and items, it is also known as a "relation technique."[6,26]

3.7 Sequence Discovery: This sequence discloses the relationships among data.[19] It is a set of object associated with its own timeline of events. Natural disaster and analysis of DNA sequence and scientific experiments are best examples.

V. DATAMININGAPPLICATIONS

Data mining is used to extract reliable information quickly from vast amounts of data. Some of the primary applications are listed below. Its primary areas of application include marketing, fraud detection, banking, telecommunication, the education sector, the medical industry, etc..

4.1 Data Mining inEducationSector: This is applied in the recently developed field of "Education Data Mining.[20]" This aids in understanding student performance, dropout rates, behavior, and course selection. The field of higher education makes extensive use of it.

4.2 DataMininginBankingandFinance: Itis

Used largely in the Banking and Financial market.It mines the credit card fraud,estimate risk and trend and profitability.Infinancial markets, it plays as a neural networks in stock forecasting price prediction etc.,

4.3 DataMininginMarketBasketAnalysis: The shopping database serves as the foundation for these methods. Finding out what goods and what people buy is their aim. The store can make use of this information [21] by increasing the visibility and client accessibility of these products.

4.4 Data Mining in Earthquake Prediction: This uses satellite maps to anticipate the earthquake. The abrupt release of tension from a geologic fault in the interior of the Earth causes an earthquake, which is the abrupt shifting of the crust. There are two ways to do this: forecasts, which are made months or years in advance, and short-term predictions [22] , which are made hours or days in advance.

4.5 Data Mining in Bio informatics: Bioinformatics created a huge amount of biological data [23]. This is a new field of inquiry to generate and integrate large quantities of proteomic, genomic and other data.

4.6 Data Mining in Telecommunication: This field has large amount of data consisting of hugecustomers.Soitisneedtominethedatatolimitthefraudulent,improvetheirmarketingeffortsandbettermanagementofnetworks .

4.7 Data Mining in Agriculture: The primary purpose of this is to increase crop yields. Four factors are taken into account: the year, the amount of rainfall, the output, and the planting area. By using it, the yield from the prediction data is improved. Data mining methods like K Means, K closest neighbor (KNN), Artificial Neural Network, and support vector machine (SVM) can be used to promote it .

VI. CONCLUSION

This article discusses how image mining is growing. It provides an analysis of the previously measured picture approaches. This review reveals the difficulties and responsibility of several prospects. This mostly focuses on data mining methods for different kinds of tasks. Obtaining information through current data is its goal. Techniques for association, grouping, prediction, and classification can be used by people in a variety of fields.

VII. REFERENCES

[1]. Janani M and Dr. ManickaChezian. R, “A SurveyOnContentBasedImageRetrievalSystem”, InternationalJournalofAdvancedResearchinComputer Engineering & Technology, Volume 1,Issue5,pp266,July2012.

[2] .Bhushan, P. Vinay, et al. "An Efficient System for Heart Risk Detection using Associative Classification and Genetic Algorithms."

[3] .Anil K. Jain and AdityaVailaya, “Image Retrievalusingcolorandshape”,InSecondAsianConferenceon ComputerVision,pp 5-8.1995.

[4]. ChNarsimhaChary,INTERNATIONAL JOURNAL OF SCIENTIFICRESEARCH IN COMPUTER SCIENCE, ENGINEERING AND INFORMATION TECHNOLOGY, “DUO MINING TECHNIQUES IN KNOWLEDGE DISCOVERY PROCESS IN DATA BASE”, 2018/06, Volume 3, Issue 1.

[5] BHUSHAN, P. V., NITESH, V., CHARY, C. N., & GUPTA, K. G. Novel Approach for Multi Cancers Prediction system using Various Data Mining Techniques.

[6] .Brown,RossA.,Pham,BinhL.,andDeVel,Olivier Y, “Design of a Digital Forensics ImageMining System”, in Knowledge Based IntelligentInformationandEngineeringSystems,pp395-404,SpringerBerlinHeidelberg,2005.

[7]. GUPTA, K. GURNADHA, CH NARASIMHA CHARY, and A. KRISHNA. "STUDY ON HEALTH CARE LIFE LOG BY THE LEVEL OF CARE REQUIRED USING KEYGRAPH TECHNOLOGY IN TEXT DATA MINING.".

[8].. chnarasimha chary, INTERNATIONAL JOURNAL OF RESEARCH,, “CLASSIFICATION OF MACHINE LEARNING TECHNIQUES AND APPLICATIONS IN ARTIFICIAL INTELLIGENCE”, 2019/02,Volume 1, Issue 1

[9]. NarasimhacharyCholleti, “ANALYZING SECURITY OF BIOMEDICAL DATA IN CANCER DISEASE”, Journal of Critical Reviews,2020/5, Volume 7, Issue 7, Pages 150-156.

[10]. CHOLLETI, NARASIMHACHARY, and TRYAMBAK HIRWARKAR."BIOMEDICAL DATA ANALYSIS IN PREDICTING AND IDENTIFICATION CANCER DISEASE USING DUO-MINING." Advances in Mathematics: Scientific Journal 9 (2020): 3487-3495.

[11]. chnarasimhachary,JOURNAL OF CRITICAL REVIEW,"ANALYZING SECURITY OF BIOMEDICAL DATA IN CANCER DISEASE",Volume 9, Issue 1..

[12]. CHOLLETI, NARASIMHACHARY, and TRYAMBAK HIRWARKAR. "BIOMEDICAL DATA ANALYSIS IN PREDICTING AND IDENTIFICATION CANCER DISEASE USING DUO-MINING." Advances in Mathematics: Scientific Journal 9 (2020): 3487-3495..

[13]. Dr.NARASIMHA CHARY CH, The International journal of analytical and experimental modal analysis ,"Privacy Preserving Media Sharing With Scalable Access Control And Secure Deduplication In Mobile Cloud Computing", 2023/1, Volume 15,Issue 1, Pages 150-156..

[14]. J.Han and M. Kamber. "Data Mining, ConceptsandTechniques",MorganKaufmann,2000.

[15]. Ch, Dr. "Narasimha Chary,"." Comprehensive Study On Multi-Operator Base Stations Cell Binary And Multi-Class Models Using Azure Machine Learning"," A Journal Of Composition Theory 14.6 (2021).

[16]. PeterStanchev, "ImageMiningforImageRetrieval",InProceedingsoftheIASTEDConference on Computer Science and Technology,pp214-218,2003.

[17]. Dr.CH.NARASIMHA CHARY, A JOURNAL OF COMPOSITION THEORY,COMPREHENSIVE STUDY ON MULTI-OPERATOR BASE STATIONS CELL BINARY AND MULTI-CLASS MODELS USING AZURE MACHINE LEARNING.

[18]. Ranjith, D., J. Balajee, and C. Kumar. "In premisesofcloudcomputingandmodels."InternationalJournal of Pharmacy and Technology 8, no. 3, pp.4685-4695,2016.

[19] Dr.NARASIMHA CHARY CH,Journal of Engineering Sciences,"securing data with block chain and ai", , Volume 14, Issue 1, 2023/, Pages 218-223.

[20]. Dr.Narasimha Chary Ch, Dogo Rangsang Research Journal,"Diagnosis And Treatment Using Covid-19 Deep Learning Approaches And Artificial Intelligence", 2023/2, Volume 13, Issue2, Pages 206-212

.

[21]. Dr.Narasimha Chary Ch,The International journal of analytical and experimental modal analysis, "Privacy Preserving Media Sharing With Scalable Access Control And Secure Deduplication In Mobile Cloud Computing",,2023/1,Volume 15, Issue 1,. Pages 150-156

[22].Ravi, Chinapaga, et al. "Analysis of Concept Drift Detection–A Framework for Categorical Time Evolving Data.".

[23].Dr.NARASIMHA CHARY CH,INTERNATIONAL RESEARCH JOURNAL ,"Prematurely Prediction and Detection of Lung toxicants utilizing Big Data Mining", , 2023/June, Volume 10,Issue 6, Pages 947-951.