

MULTI-VIEW POSITIONING HASHING, COMPONENT BASED FEATURE EXTRACTION ALGORITHM AND CLASSIFICATION IN IMAGE SEARCH ENGINE

Ripanjot Aneta¹, Mrs. Apoorva Arora²

M.Tech (Scholar), Assistant Professor

Department of Computer Science Engineering

Chandigarh University (Chandigarh-Ludhiana Highway, Mohali)

Abstract - Data mining tools predict future trends & behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining transfer beyond the analyses of past actions provided by retrospective tools typical of decision support systems. In recent time, there is a large growth of digital picture and video achieves. Some of these are very significant for user point of view. These collections of digital pictures most of the time might be for personal use and might be for public use. For searching pictures from dataset always require secure and efficient approaches, which occurs by the proper arrangement of data in the database. Proposed approach is using ECC approach for the efficient arrangement of data and a ICA algorithm for proper and efficient image feature retrieval, along with this there is also using BPNN for classification for secure communication between the user and database while data access. Parameters like FRR, FAR and Accuracy are calculated and compared with previous existing systems with better performance and high security using encryption technique.

Keywords - Data mining, digital images, encryption techniques, feature extraction and parameters.

I. INTRODUCTION

Data mining tools predict future trends & behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining transfer beyond the analyses of past actions provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that usually were too time consuming to resolve. They scour databases for secreted patterns [1], finding predictive information that experts may miss because it lies outside their expectations. Most companies already collect & refine massive quantities of data. Data mining methods can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, & can be integrated with new products & systems as they are brought on-line [2]. Data mining, also popularly known as Knowledge Discovery in Database, refers to eliminating or "mining" knowledge from great quantities of data. Data mining

techniques are used to operate on large volumes of data to discover hidden patterns and relationships useful in decision making. While data mining & knowledge discovery in database are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. Data mining and knowledge discovery applications have got a rich focus due to its significance in decision making and it has become an essential component in many organizations. Data mining methods have been introduced into new fields of Statistics, Databases, Machine Learning, Pattern Reorganization, Artificial Intelligence and Computation capabilities etc. Image processing is a method to perform some operations on an image, in order to get an enhanced picture or to extract some valuable information from it.

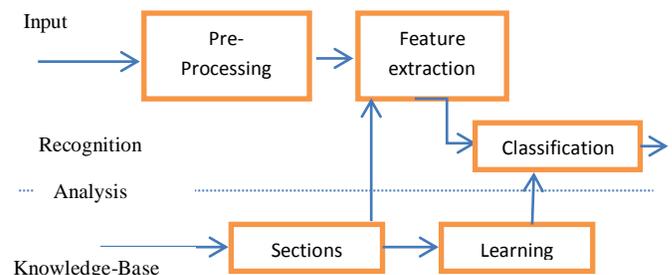


Fig. 1: Image Processing Flow Chart

It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is amongst rapidly growing technologies. It forms core examination area within engineering and computer science disciplines too. Image processing basically includes the following three steps [3]:

- Importing the image via image acquisition tools;
- Analyzing and manipulating the image[4];
- Output in which result can be altered image or report that is based on image analysis [5].

There are two types of procedures used for image processing specifically, analogue and digital image processing. Analogue image processing can be used for the hard copies like printouts & photographs. Image analysts use many fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in manipulation of the digital pictures by using computers. The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction [7,8].

II. RELATED WORK

Han and Kamber et.al 2000[9] describes data mining software that permit the users to analyze data from dissimilar dimensions, categorize it and summarize the relationships which are identified during the mining process. Pandey and Pal et.al 2011[10] conducted study on the student performance based by selecting 600 students from different colleges of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayes Classification on category, language & background qualification, it was creating that whether new comer students will performer or not. Hijazi and Naqvi et.al 2006 [11] conducted a study on the student performance by choosing a sample of 300 students (230 males, 70 females) from a group of colleges affiliated to Punjab university of Pakistan. The hypothesis that was stated as "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education are significantly related with student performance" was framed. By means of simple linear regression analysis, it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance. Galit et.al 2007 [13] gave a case study that use students data to analyze their learning behavior to predict the results & to warn students at risk before their final examinations. Al-Radaideh, et al 2006 [14] applied a decision tree model to predict the final grade of students who studied the C++ course in Yar-mouk University, Jordan in the year 2005. Three changed classification methods namely ID3, C4.5, and the Naive Bayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models. Yi Lou, Juqin Shen et.al 2016 [15] this system is to predict the dangerous risks issues in the hydraulic engineering migration, and conducts the visual output through the network system, allowing decision makers to better cope with the crisis.

III. ENCRYPTION ALGORITHM IN ELLIPTIC CURVE CRYPTOGRAPHY

ECC is the latest approach, and measured as a wonderful method with low key size for the user, and has a solid exponential time challenge for a hacker to break into the system [6]. In ECC a 160-bit key offers the similar security as related to

the out dated crypto system RSA with a 1024-bit key, thus lesser the computer power. Consequently, ECC gives significantly greater security for a assumed key size. Therefore, a key through small size makes it possible much more condensed executions for a given level of security which means faster cryptographic actions, running on lesser chips or more compact software.

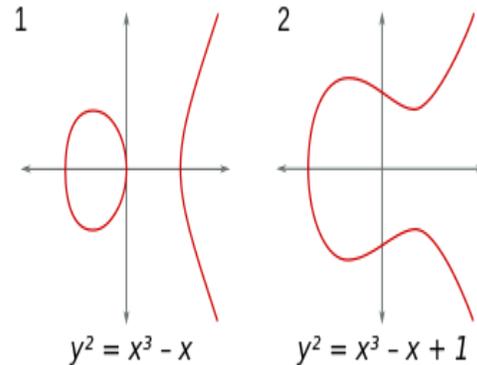


Fig. 2: Crypto-series: Elliptic Curve Cryptography

IV. FEATURE EXTRACTION AND CLASSIFICATION

A. Independence Component Analysis Algorithm

ICA (Independent Component Analysis) is a measurable and computational strategy [16]. ICA is a general model for multivariate information. In this model information variables are thought to be direct and other obscure variables are thought to be inert variables. The inactive variables are called autonomous segments of the watched information. ICA is to some degree identified with PCA. ICA is significantly more effective strategy. ICA is begun from numerous application fields, databases, monetary pointers and estimation and so forth. The two broadest meanings of freedom for ICA are

- Minimization of common data
- Maximization of non-Gaussianity

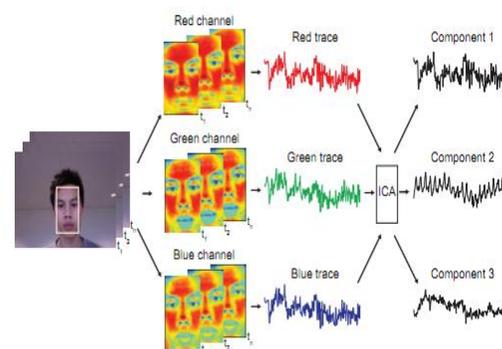


Fig. 3: Feature Extracted in Component based algorithm

B. Classification Approach

The Back Propagation neuron system is artificial neural network based on error back propagation algorithm. The Back Propagation Neural Network imitation consists of an enter layer, some hidden layers and an output layer. Each connection linking neurons has a characteristic weighting value. In training the network, the nodes in the BP neural network find input information as of exterior basis, & then go by to hidden layer which is an interior data processing layer & is responsible [17] for the information conversion, and then the nodes in the output layer supply the required output matter. Behind that, the anti-propagation of mistake is transported by distinct the actual output with wanted output. Every weight is reconsider & back propagated layer by layer from output layer to hidden layer and input layer. This procedure will be continued until the output fault of network is reduced to an acceptable level or the predetermined time of education is realized. The dispensation consequences of information are exported by output layers to the outside.

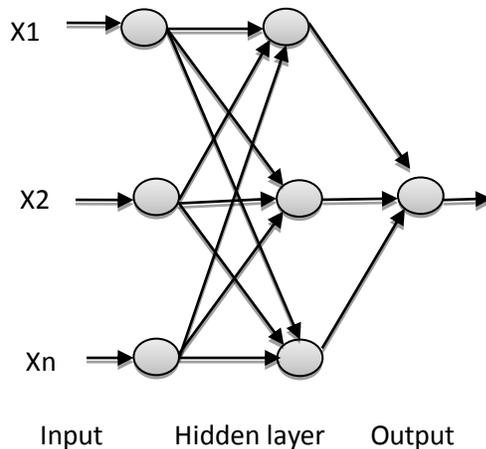


Fig. 4: Back Propagation Neural Network [5]

V. ISSUES IN IMAGE SEARCH ENGINE

The previous methods for image search suffered from the unreliability of the statement under where initial text-based image searches result [18]. However, such results containing a large number of figures and with quantities of unrelated figures. Image search-engines can apparently provide an effortless route but currently are limited by the poor precision of the returned images and also boundaries on the total number of the image provided by the data based image containing relevant and irrelevant image results. Which all of the existing algorithms require a prior assumption regarding the relevance of the figures in the initial, link based search result. When in search engines, those corresponding images are loaded in particular time, meanwhile among them there an uncategorized image also spotted. Anyhow producing such databases containing a large

number of figures and with high precision is still a manual task [19]. Generally, Image-search-engines apparently provide an easy route. Web navigation behavior is helpful in understanding what information of online users demand. Following that, the analyzed results can be seen as knowledge to be used in intelligent online requests, refining web site maps, web based personalization system and improving searching accuracy when seeking information. Nevertheless, online navigation behavior grows each passing day, and thus extracting information intelligently from it is a difficult issue. An efficient image search engine user saw the information publically and hack the information in the server side. We protect the information using ECC technique and classification (Back Propagation Neural Network) [20].

VI. SIMULATION MODEL

Proposed approach is able to provide secure communication between user and server. It provides a secure environment to the user to store their images in server and find with a unique key corresponding to user profile for search them. It restricts the outside applications to find the images and their links to show on search results. Only the authorized users can find and extract the link of image in the proposed system. It can be more secure to generate the visual record of search and other queries related to the images in a particular system. Proposed approach is using ECC encryption scheme to encrypt data over a web server repositories. These repos store user images and other data with a unique link so that they can provide more availability. The encryption algorithm encrypts the links over web server so that the original location of image can be hiding from the unauthorized users.

Proposed approach is working with feature extraction, feature selection and artificial intelligence. These make this approach more accurate and faster over a web server. Feature extraction process used to detect the unique properties of a particular image and train neural network for generate a knowledge base. Knowledge base used to classify the data real world data at the time of live server execution. When the server database has huge data sets, the artificial intelligence provide much batter speed for detect user query. Proposed approach compared with various parameters like FRR, FAR, and Accuracy to check the performance against existing systems.

All these parameters compared with them and as per performance the proposed algorithm working better than others. It is much secured than other due to hiding function of this algorithm from outside the world. This function performed with the ECC algorithms to decrease the probability of image access outside the application.

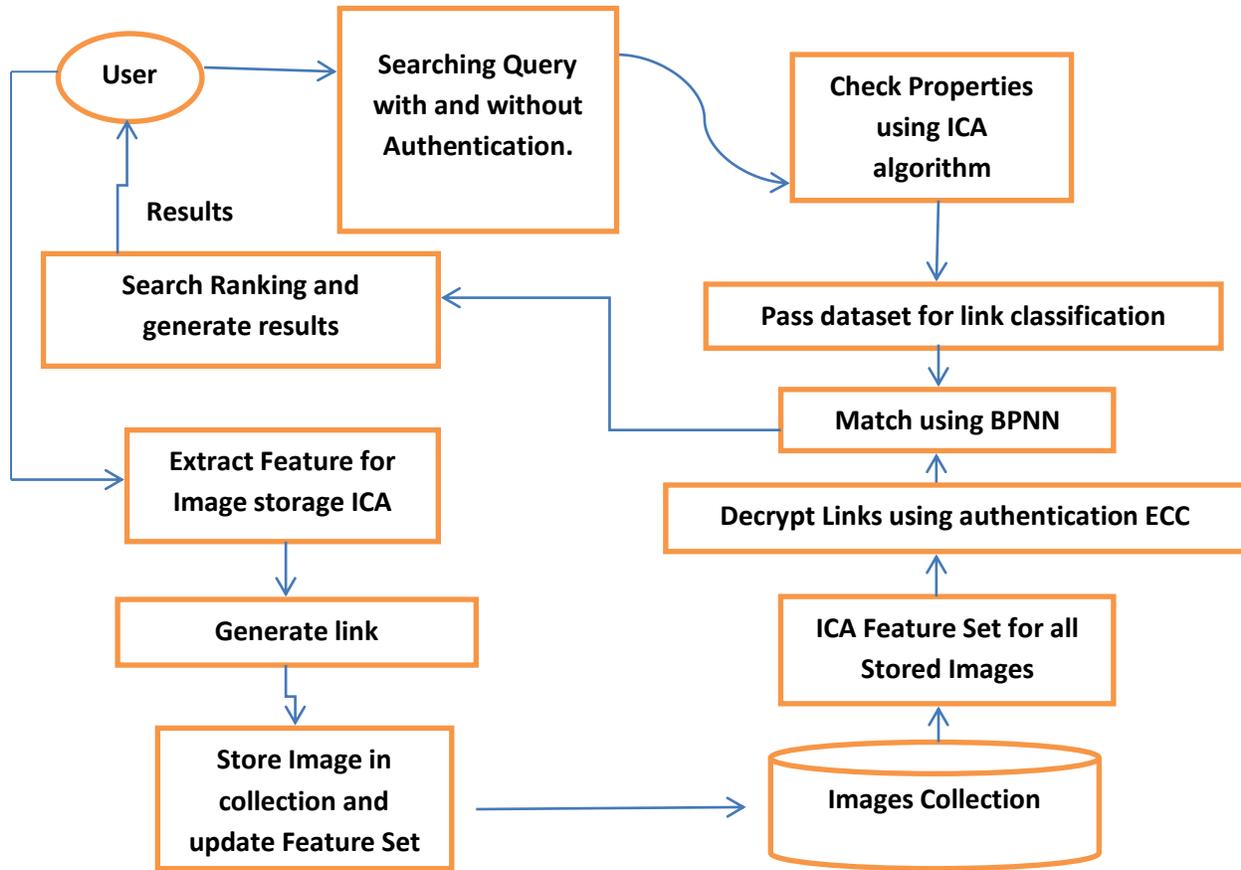


Fig. 5: Proposed Flow Chart

VII. RESULT AND DISCUSSIONS

In this section, we discuss the results an image search engine. It contains the entire module in this research. It used to bind them into a single panel. Various phases, like training and testing the dataset over a server shown here as upload and perform various operations.

User uploads their images after login from the login panel. Images stored with unique pattern into the dataset. It used to secure data over a huge dataset.

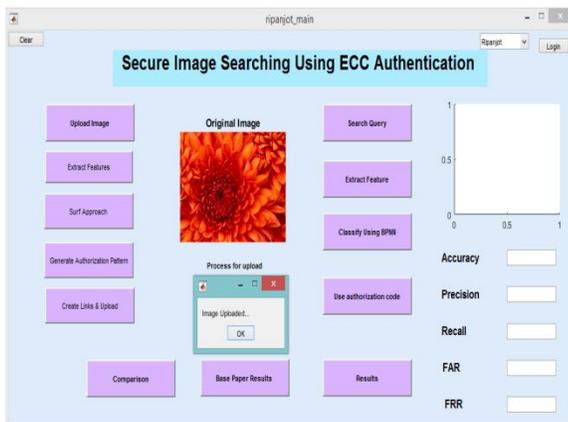


Fig. 6: Upload Image

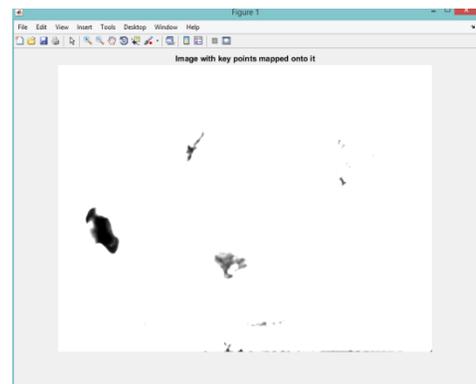


Fig. 7: SURF Image and Component based Extracted Feature graph

ICA used to extract features of image during upload over a web server. ICA extracts uniqueness of image for easy storage and searching. ICA converts image uniqueness into unique

numeric pattern. It helps classifiers to classify image over a large dataset.

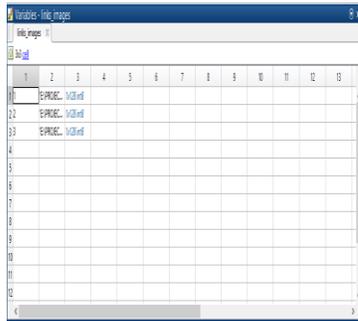


Fig. 8: Stored Links

Stored links used to check image location over a network. Proposed approach encrypt storage link for fast and secure retrieval. Proposed system encrypt links instead of files it make more fast for searching and decryption process.

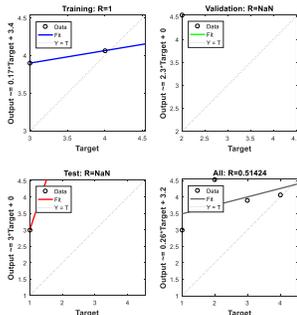
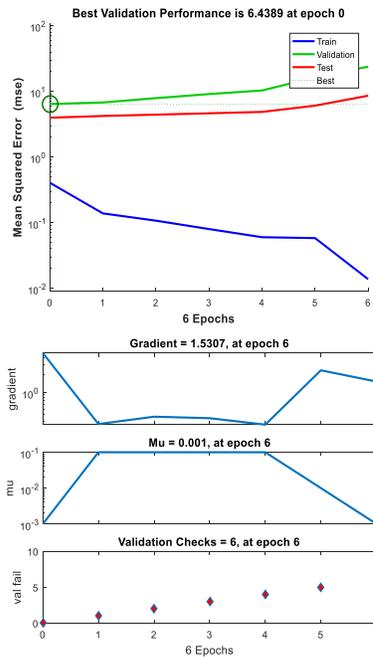


Fig. 9: Best performance, Training State and Regression

The above figure shows that the best performance of the Training set with the help of mean square error value. Our Best validation performance value is 6.4 . It calculates the training set with the gradient, mutation in performance according to the time. We calculate the gradient value is 1.530, mutation value is 0.001 and validation check used in 6 out of 6 epochs. It defined that the regression values i.e training set, testing set, validation check. We check the average of the whole scenario of the training and testing set. We found the average value is 0.514.

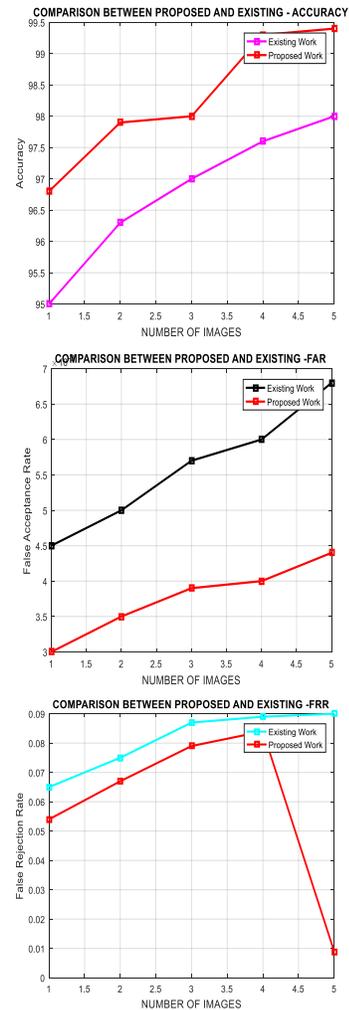


Fig. 10: Comparison between proposed and existing Work- accuracy , FAR and False Rejection rate

The above figure shows that the comparison between existing and proposed work accuracy. We improve the accuracy with the help of BPNN approach. In proposed work we attained the accuracy value is 99 % and existing work accuracy value is 96%. The comparison of existing and proposed in far. The false acceptance rate, or FAR, is the measure of the likelihood that the biometric security system will incorrectly accept an access attempt by an unauthorized user. A system's FAR typically is

stated as the ratio of the number of false acceptances divided by the number of identification attempts. The comparison between existing and proposed work in FRR. The FRR or False Rejection Rate is the probability that the system incorrectly rejects access to an authorized person, due to failing to match the biometric input with a template.

VIII. CONCLUSION AND FUTURE SCOPE

In today's technically and fashionable developed world there's a search of images on public and personal net-domain, which attain novel analysis develop strategies for commercial and ascend-able picture search. Where the normal picture program heavily relies on the matter-world connected/ hooked-up which pictures. Excluding providing higher-picture hunted for customary web portal-user, large scale similar contestable to be very helpful for locating sort of arduous issues in laptop vision and transmission like image division. Normally image explosion system embrace binary parts start might be trivial picture feature illustration and a quality search mechanism. By this binary part it's convenient for client to normally get valuable info. Form into vectorization some improved approaches of ECC (Ecliptic Curve Cryptography) and Classification to provide the authorized link. The independent component analysis, using the feature extraction of the original image. In existing Work, used the SURF feature extraction to identify the unique properties of the original image. Still there is a huge scope in the image retrieval process related to secure Image search. Global image signatures have been developed based on local gradient descriptors in this thesis. Often, for glossy or texture-limited surfaces, the shape and color of an object can be more discriminative cues than local intensity gradients. It will be interesting and useful to study the development of global image signatures based on other types of image features such as shape or color descriptors.

IX. REFERENCES

- [1] Heikki, Mannila, Data mining: machine learning, statistics, and databases, IEEE, 1996.
- [2] U. Fayadd, Piatetsky, G. Shapiro, and P. Smyth, From data mining to knowledge discovery in databases, AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0-262 56097-6, 1996.
- [3] Chitte, P. P., et al. "IRIS Recognition System Using ICA, PCA, Daugman's Rubber Sheet Model Together." International Journal of Computer Technology and Electronics Engineering 2.1 (2012): 16-23.
- [4] Shi, Jin-Xin, and Xiao-FengGu. "The comparison of iris recognition using principal component analysis, independent component analysis and Gabor wavelets." Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on. Vol. 1. IEEE, 2010.
- [5] Ramya, M., A. Muthukumar, and S. Kannan. "Multibiometric based authentication using feature level fusion." Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on. IEEE, 2012.
- [6] V. Miller, "Uses of elliptic curves in cryptography", Advances in Cryptology - CRYPTO'85, LNCS 218, pp.417-426, 1986.
- [7] Kani, Ernst. "The State Of The Art Of Elliptic Curve Cryptography." Queen's University.
- [8] Kapoor, Vivek, Vivek Sonny Abraham, and Ramesh Singh. "Elliptic curve cryptography." Ubiquity 2008.May (2008): 7
- [9] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000.
- [10] U . K. Pandey, and S. Pal, "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN:0975-9646, 2011.
- [11] S. T. Hijazi, and R. S. M. M. Naqvi, "Factors affecting student's performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.
- [12] Z. N. Khan, "Scholastic achievement of higher secondary students in science stream", Journal of Social Sciences, Vol. 1, No. 2, pp. 84-87, 2005..
- [13] Galit.et.al, "Examining online learning processes based on log files analysis: a case study". Research, Reflection and Innovations in Integrating ICT in Education 2007.
- [14] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, "Mining student data using decision trees", International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.
- [15] Yi Lou, Juqin Shen , " The Development and Application of Hydraulic Engineering Migration Risk Early Warning System Based on Data Mining ." Parassitologia 47, no. 1 (2016): 81-96.
- [16] Francis Rousseaux, and Paul M. Schwartz. " The Legal Debate about Personal Data Privacy at a Time of Big Data Mining and Searching" The University of Chicago Law Review 75, no. 1 (2016): 261-285.
- [17] AshokMV, Emine Uçar, and Dursun Delen. " Data Mining Approach For Predicting Student and Institution's Placement Percentage " Expert Systems with Applications 39, no. 10 (2016): 9468-9476.
- [18] Carolina Lagos, Katherine A. Connors, and Claudio Suárez. " Facing the data analysis complexity for the energetic efficiency management at great copper mining, in Codelco Chuquicamata, Chile " Natural Resources Research 16, no. 2 (2016): 171-190.
- [19] Colin Conrad, N., Fritz J. Scheuren, and William E. ELM: An Extended Logic Matching Method on Record Linkage Analysis of Disparate Databases for Profiling Data Mining. Springer Science & Business Media, 2016.
- [20] Žunić, Emir, Almir Djedović, and Dženana Đonko. "Application of Big Data and text mining methods and technologies in modern business analyzing social networks data about traffic tracking." In Telecommunications (BIHTEL), 2016 XI International Symposium on, pp. 1-6. IEEE, 2016.