CONTENT BASED SATELLITE IMAGE RETRIEVAL SYSTEM FOR BIG DATA USING MACHINE LEARNING APPROACH

Hitakshi, M.Tech. Research Scholar,
Department of CSE,
SBBS University, Jalandhar
(vermpadam686@gmail.com)

Dr.Jagdeep Kaur, Associate Professor,
Department of CSE
SBBS University, Jalandhar
(sbbsuuietcse@gmail.com)

Dr. Vijay Dhir, Professor, Department of CSE,
SBBS University, Jalandhar
(drvijaydhir@gmail.com)

Abstract:

The CBIR for satellite images using feature extraction techniques like color, shape texture etc. remains a open problem. The images can be retrieved but with poor efficacy. The problem of the above problem is to bridge this gap by using a high level feature extraction technique. The high level descriptors are also used but still retrieved the results with lower results. By considering this point in mind, the objective of the work, is to bridge the above define gap and provide an accurate and efficient solution. In this research, a novel CBIR system for satellite image using AI approach is developed. This system used Speeded Up Robust Features (SURF) as feature descriptor, and Particle Swarm Optimization (PSO) as feature optimization approach. These approaches help to provide better and higher quality satellite images, which improves the future classification of the uploaded image. For classification, Artificial Neural Network (ANN) is used to retrieve the satellite images. Similarity matching plays very important role to attain accurate results. Similarity technique is applied to identify the best possible image among the rest of the similar images. The experiment shows better results in terms of kappa coefficient and overall accuracy compared to existing work.

Keywords: Content Based Image Retrieval, Speeded Up Robust Features, Particle Swarm Optimization, Artificial Neural Network**.**

DOI Number: 10.14704/nq.2022.20.8.NQ44231 **NeuroQuantology 2022 ;20(8):2118-2126**

1. Introduction

In recent year, the count of satellite launch has been increased day by day. Satellites are used to extract essential information of manmade and natural sources. These images are also helpful for vegetation monitoring, cover mapping, time series analysis etc. presently, more than 200 satellites are capturing images

both multi-spatial and multi-temporal, and collected information is send on earth surface on hourly, daily, or weekly basis for data processing [1]. Due to which, count of image data received from satellitesalso increases.The widespread availability of the captured satellite images not only increase the data volume but also increase the order of magnitude to store the amount of details within the satellite images. Collectively, this is known as "Big satellite data" [2]. The images contain very complex and high resolution data, which is very difficult to store that data into a large memory. Also, if it stored anyhow, then it is very tough to process or analyse the stored information. The process of doing this entire task is also a time consuming especially for multispectral images. The information in terabytes is generated by each satellite in a day so it is very difficult to extract useful information from them [3].

Machine learning approach is one of the most useful approaches that can be used to extract the useful information from large and complex data. But before applying machine learning approach data should be presented in a proper form [4]. The large image data collected by the satellite sensors are stored into a big data repository. This repository holds data in different forms that are structured, unstructured, and semi-structured. Data normalization is performed on big data, which frame data into rows and columns. This framed data is considered as data warehouse for further data analysis [5]. To further process this framed data techniques such as ensemble analysis [6], machine learning [7] etc. can be used. In this research Content Based Image Retrieval (CBIR) method is adopted to extract content information from the image. This is really agreat research as the extraction of information manually is not possible. Satellite images holds information like weather forecasting, ocean and coastal environment, atmospheric disturbances etc. to extract these information CBIR is the best option. Using this technique actual content from a million of images is find out instead of the metadata like as keywords, tags etc [8]. here, the word Content means to search different key parameters that are color, shape, texture etc. within the image. The presented CBIR technique is performed on the images collected from "name of dataset" that contains images of various categories including land, forest, human population, crop etc. Initially, pre-processing is performed to remove unwanted data from the image. Thereafter color histogram method is applied to represents the distribution of colors presented in the image along with its pixel count. The features of these images are extracted using SURF approach. Thereafter, Particle Swarm Optimization (PSO) is applied to optimize these pixel values. At last system is trained using Artificial Neural Network (ANN) as machine learning approach.

2. Related Work

Srivastava and Ahmed (2021) have presented aFeature-Based Image Retrieval (FBIR) system for earth surface monitoring satellite images. The images are initially pre-processed; thereafter to extract true color combination of image radiometric calibration has been applied. This extracts three primary colors that are Red, green and Blue. Along with color, radiometric calibration technique also represents the shape and sizes of different image features. Next step is the feature extraction, which is extracted from the image pixel intensities. For this color histogram method was applied, which calculate the average and standard deviation of the color intensity of each color component. For training and classification three classification techniques, maximum likelihood (ML), support vector machine (SVM), and artificial neural network (ANN) have been applied on various vegetation indices. The results using ANN technique was found to be better compared to other two techniques [9].**Kavitha and Vidhya (2021)** have presented a fuzzy based multi-classification system, which find region

immediately from a large database. It minimized the use of image retrieval algorithms. Gaussian filter is used to process the images [10].**Cao et al. (2020)** have proposed a new CBIR technique by using the Convolutional neural network (CNN) technique for remote sensing images. A triple network has been created by learning process using CNN approach. The features were also extracted in a semantic space by using the Euclidean distance as a similarity measure [11].**Xiong et al. (2019)**have developed a CNN based CBIR system for satellite images. CNN extract the silence features from the query image while suppressing the unwanted features. Further, relation between image features distances that is inter and intra class features have been detected. Using this the author achieved better accuracy [12].**He et al. (2018)**have used SIFT as feature descriptor to extract each image and then stored each features. When new image is uploaded its features are extracted and then compared with the available features in library. After comparing visual similarity images results are computed and final or the best similar image is provided as an output image. In addition to SIFT, the authors have also used distance ratio as an reference to control the numerous matched feature points in an image [13]. **Bouteldja and Kourgli (2015)** have presented a CBIR based system for the retrieval of high resolution satellite images. Steerable pyramid scheme has been applied for image texture identification. Two color spaces RGB and CIElab have been used for the extraction and generation of feature database. The features like translation, rotation and illumination properties of images have been taken into consideration. The researchers concluded that both RGB and CIElab techniques on an average perform equally. The researchers have considered rotation retrieval technique in future shape feature can be taken into consideration [14].**Jiexian et al. (2014)** have presented a multi-scale distance vector scheme for CBIR for satellite image. This technique was used to remove noise in those pictures that are having same extracted features and distance. Initially Gaussian function has been used to design an image contour curve. The researchers have considered other features like rotation translation scaling etc constant [15].

3. Methodology

To verify the efficiency and accuracy of proposed CBIR system for satellite image retrieval using Particle Swarm Optimization (PSO) as metaheuristic approach and the Artificial Neural Network (ANN) as a classifier; we have perform several experiments with this procedure on several satellite images. In the proposed CBIR system, there are several steps used for retrieving the satellite images. The methodology of proposed work is given below:

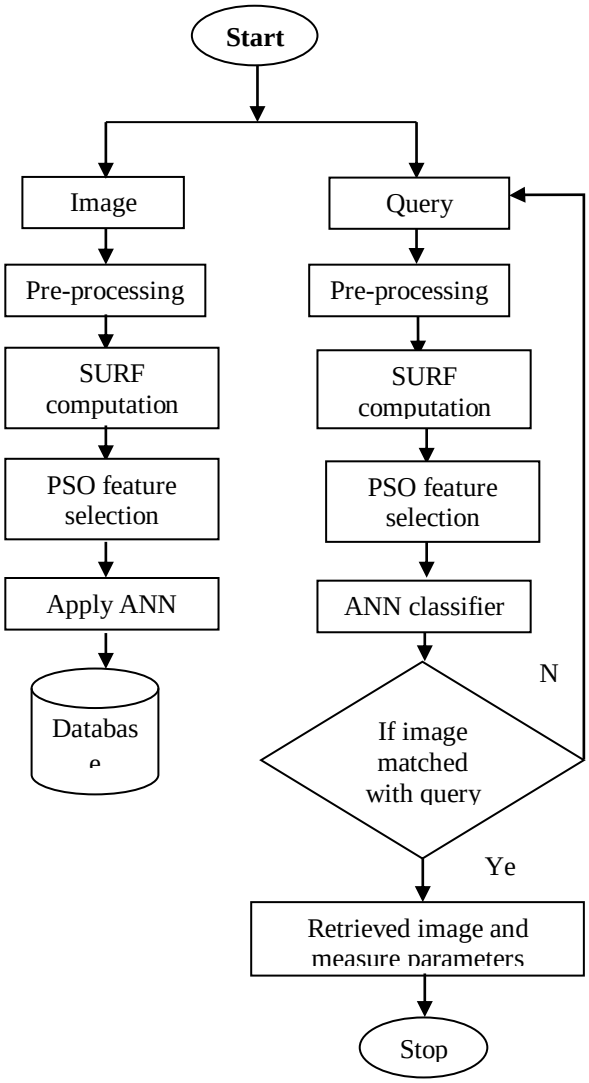

Fig. 1 Flow of Proposed CBIR based satellite image retrieval system

Initially, GUI for the proposed CBIR system for satellite image retrieval has been designed. The designed GUI consists of various buttons like as Upload query image, image retrieval, calculate parameters, create database, load database reset, and exit. Now, next step is to create database using high quality satellite images. After creating it, system is trained. But before, training, the images are processed to obtained high quality images. So that the data can be extracted easily and desired image can be identified. The uploaded query image is shown in Figure below.

*e*ISSN1303-5150 www.neuroquantology.com

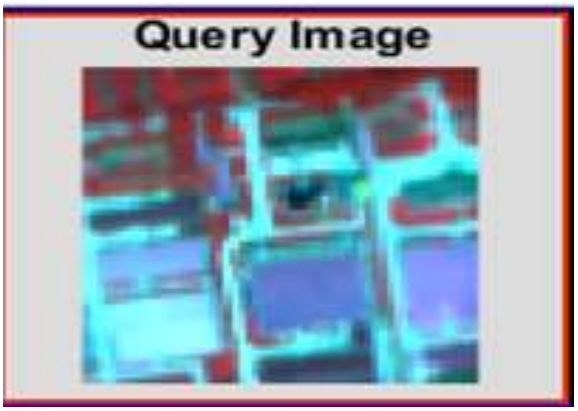

Fig. 2 Query Satellite Image

The foremost step on image that needs to be done is pre-processing. In pre-processing the undesired part of the image is removed or filtered to make small and high quality data. Color histogram of image has been measured. Using this, the digital images are converted into a graphical representation, which represents the number of pixel for each value. The color histogram for the same image is shown in Figure below.

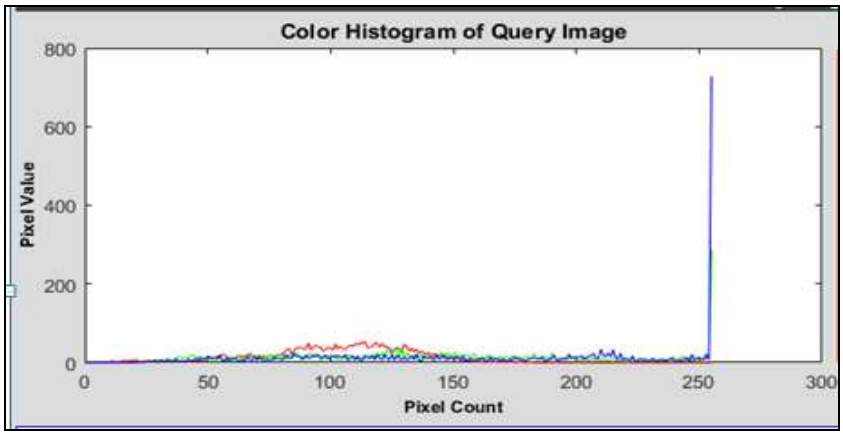

Fig. 3 Color Histogram of Satellite Image

Apply speeded up robust features (SURF) Descriptor for the feature extraction of the pre-processed images. It descript images with high speed and hence time of extraction is small. It mainly offers the point of interest detection, which has been performed in three steps. First one is to compute the integral image, then analyze its bob structure to obtained interest points, at last, the best interest points have been selected and localized. After feature extraction, the feature points are optimized using PSO as metaheuristic algorithm. Using this algorithm a threshold value is decided for the interest points. The points that are having values above this threshold value are considered as optimized image and then passed that image for training and the query image for classification.

Apply Artificial Neural Network on optimized data to train the created database. Also in the testing phase, query image is compared with the images stored into the database. If features of the uploaded query image are matched, that is having smaller similarity distance, shows highest matched image. The similarity measures of different images are shown in figure below.

*e*ISSN1303-5150 www.neuroquantology.com

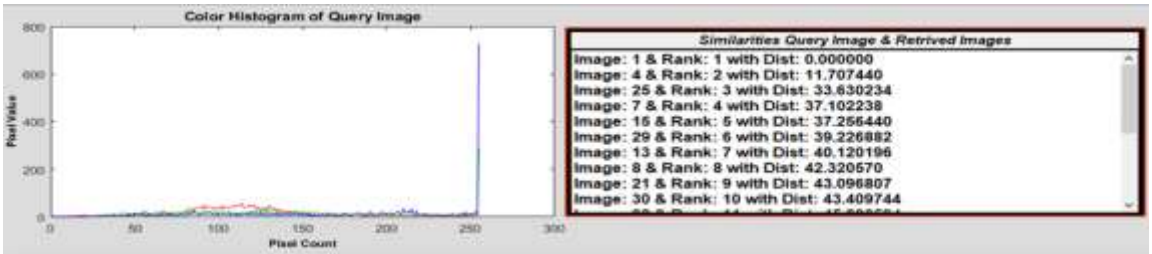

Figure 4: Retrieved Images

Then the performance parameters are measured. Otherwise upload the query image again, and repeat the entire process.

4. Result and Discussion

Several experiments has been performed by uploading different image on the designed satellite image based ANN CBIR system. The designed CBIR system is tested using different categories of satellite images. A set of database is prepared by collected multiple satellite images manually. The images are available in RGB format having size of 256 * 384 pixels.

Table 1 Performance parameters

Image ID	Kappa coefficient	Overall Accuracy (%)
Image:1	0.795	87.465
Image:2	0.801	89.397
Image:3	0.785	88.495
Image:4	0.813	83.937
Image:5	0.827	85.834
Image:7	0.795	88.193
Image:8	0.804	86.395
Image:9	0.785	84.402
Image:10	0.794	86.144
Image:11	0.814	83.198
Image:12	0.792	84.497
Image:13	0.816	85.395
Image:14	0.793	86.019

Figure 5 illustrates the graph between the kapp coefficient and the image ID's. A total of 13 Image ID's the kappa coefficient is calculated and presented in Table 1. The kapp coefficient value for 13 different satellite images varied from 0.785 to 0.827 for image ID: 9 and Image ID: 5 respectively. The average kappa coefficient of the designed CBIR system using PSO with ANN was determined as 0.8010.

The overall accuracy of the proposed CBIR based AI system for satellite image retrieval is presented in Table 1. Graphically, it is presented in Figure 6. The x axis represents the Image ID's, y axis represents the Kappa coefficient of the respective uploaded image. The range of overall

accuracy was determined as 83.198 % to 89.397 %. The average overall accuracy was measured as 86.105 %.

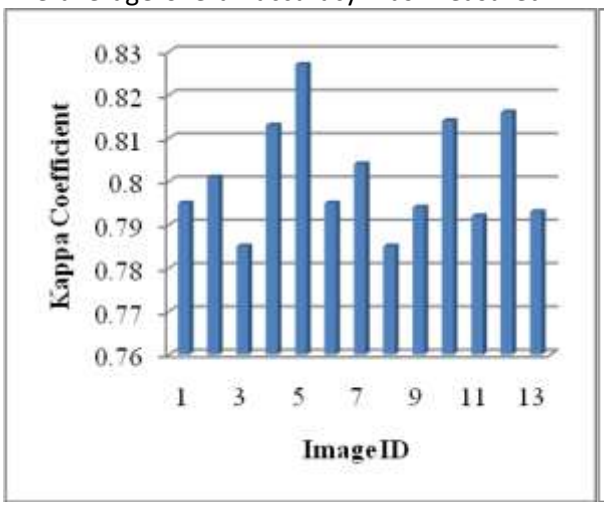

Figure 5: Kappa coefficient

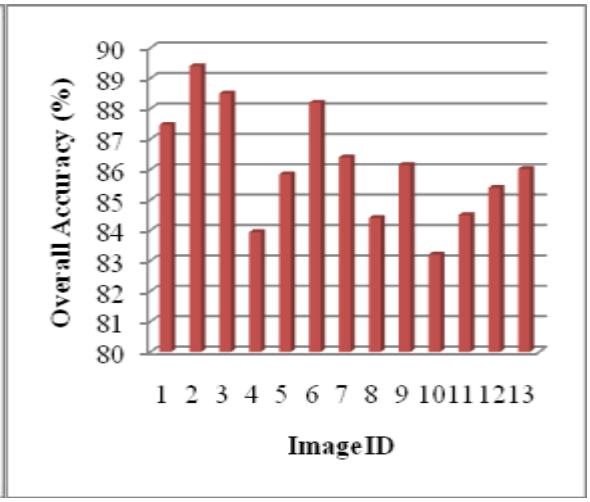

Figure 6 Overall Accuracy

To show the effectiveness of the proposed work with the existing work, comparison between the measured parameters is compared with the existing work **Srivastava, and Ahmed (2021**). The comparison average values are listed in Table 2.

Table 2 Comparison of proposed with existing work

	Kappa Coefficient	**Overall Accuracy (%)**
Proposed using PSO with ANN	0.8010	86.105
Existing using SVM	0.7842	85.1

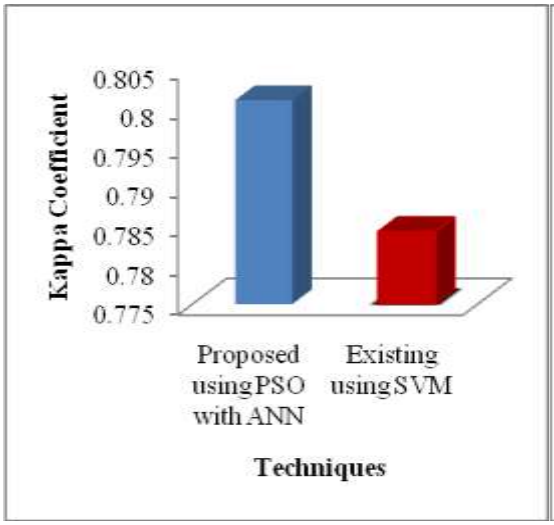

Figure** Error! No text of specified style in document..**1 Comparison of Kappa Coefficient

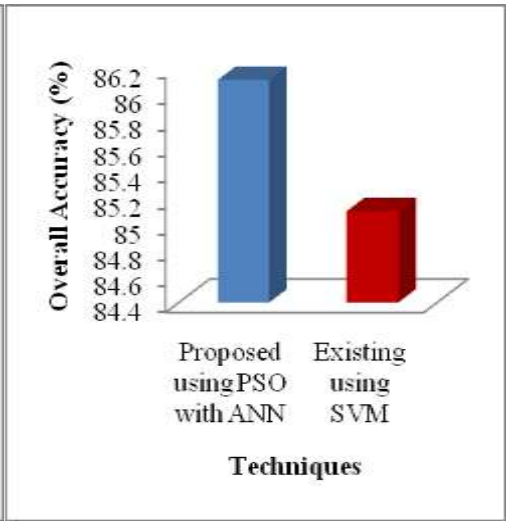

Figure** Error! No text of specified style in document..**2 Comparison of Overall Accuracy

*e*ISSN1303-5150

Table 2 represents the comparison of kappa coefficient observed for the proposed work and existing work. In proposed work, parameter is evaluated by apply PSO as feature optimization approach and ANN as classification approach to retrieve the uploaded image. On the other side, in existing work researcher (**Srivastava, and Ahmed**) have retrieve images using SVM as classification approach. From the graph the observed average value of kappa coefficient for proposed and existing work are observed as 0.8010 and 0.7842 respectively. Therefore, there is a percentage improvement of 2.14 %.

5. Conclusion

In this research, a novel method for CBIR system for satellite images is proposed. The main objective of this research is to identify or retrieved the most similar image from the big data of satellite images as per the uploaded query image. The image identification has been performed for different satellite images. In this research, metaheuristic approach has been implemented along with feature descriptor on multispectral satellite images. The images have been processed initially to obtained better image. For successful retrieval of images, feature extraction technique along with feature optimization technique play a crucial role. There are many systems being developed for CBIR, but they have not provided promising results due to the semantic gap. In this design, we have tried to bridge the semantic gap by using feature extraction technique with AI approach. The proposed system provides better results in terms of kappa coefficient and overall accuracy. Following points are examined from the research are:

- The average kappa coefficient of the designed CBIR system using PSO with ANN was determined as 0.8010.
- The range of overall accuracy was determined as 83.198 % to 89.397 %. The average overall accuracy was measured as 86.105 %.
- the observed average value of kappa coefficient for proposed and existing work are observed as 0.8010 and 0.7842 respectively. Therefore, there is a percentage improvement of 2.14 %.
- The average value of accuracy observed for both proposed PSO with ANN and existing SVM technique to retrieve satellite image using CBIR system was observed as 86.105 % and 85.1 % respectively. Thus, there is an improvement of 1.18 % has been attained.

References

[1]. Kumar, M. S., Rajeshwari, J., & Rajasekhar, N. (2022). Exploration on Content-Based Image Retrieval Methods. In *Pervasive Computing and Social Networking* (pp. 51-62). Springer, Singapore.

[2]. Ashraf, R., Ahmed, M., Jabbar, S., Khalid, S., Ahmad, A., Din, S., & Jeon, G. (2018). Content based image retrieval by using color descriptor and discrete wavelet transform. *Journal of medical systems*, *42*(3), 1-12.

[3]. Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., ... & Khalil, T. (2019). Content-based image retrieval and feature extraction: a comprehensive review. *Mathematical Problems in Engineering*, *2019*.

[4]. Dai, O. E., Demir, B., Sankur, B., & Bruzzone, L. (2018). A novel system for content-based retrieval of single and multi-label high-dimensional remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(7), 2473-2490.

[5]. Li, P., Zhang, X., Zhu, X., & Ren, P. (2018). Online hashing for scalable

*e*ISSN1303-5150

remote sensing image retrieval. *Remote Sensing*, *10*(5), 709.

[6]. Hamreras, S., Boucheham, B., Molina-Cabello, M. A., Benitez-Rochel, R., & Lopez-Rubio, E. (2020). Content based image retrieval by ensembles of deep learning object classifiers. *Integrated Computer-Aided Engineering*, *27*(3), 317-331.

[7]. Alrahhal, M., & Supreethi, K. P. (2019, February). Content-based image retrieval using local patterns and supervised machine learning techniques. In *2019 Amity International Conference on Artificial Intelligence (AICAI)* (pp. 118-124). IEEE.

[8]. Bapu, J. J., & Florinabel, D. J. (2020). Automatic annotation of satellite images with multi class support vector machine. *Earth Science Informatics*, *13*(3), 811-819.

[9]. Srivastava, S., & Ahmed, T. (2021). Feature-Based Image Retrieval (FBIR) System for Satellite Image Quality Assessment Using Big Data Analytical Technique. Psychology and Education, 58(2), 10202-10220.

[10]. Kavitha, P. K., & Vidhya Saraswathi, P. (2021). Content based satellite image retrieval system using fuzzy clustering. *Journal of Ambient Intelligence and Humanized Computing*, *12*(5), 5541-5552.

[11]. Cao, R., Zhang, Q., Zhu, J., Li, Q., Li, Q., Liu, B., & Qiu, G. (2020). Enhancing remote sensing image retrieval using a triplet deep metric learning network. International Journal of Remote Sensing, 41(2), 740-751.

[12]. Xiong, W., Lv, Y., Cui, Y., Zhang, X., & Gu, X. (2019). A discriminative feature learning approach for remote sensing image retrieval. *Remote Sensing*, *11*(3), 281.

[13]. He, T., Wei, Y., Liu, Z., Qing, G., & Zhang, D. (2018, January). Content based image retrieval method based on SIFT feature. In *2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)* (pp. 649-652). IEEE.

[14]. Bouteldja, S., & Kourgli, A. (2015, September). Multiscale texture features for the retrieval of high resolution satellite images. In 2015 International Conference on Systems, Signals and Image Processing (IWSSIP) (pp. 170-173). IEEE.

[15]. Jiexian, Z., Xiupeng, L., & Yu, F. (2014). Multiscale distance coherence vector algorithm for content-based image retrieval. *The Scientific World Journal*, *2014*.

***e*ISSN**1303-5150