International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST229516

A Novel Classification Performance Approach for Remotely Sensed Multispectral Image Data by Using Data Mining Techniques

Dr. Bodla Kishor[1], Dr. Y Sowmya Reddy[2], Dr. K. Rubenraju[3], Dr. Mahesh Kotha[4]
[1]Assistant Professor, CSE Department, CMR Engineering College, Hyderabad, India
[2]Associate Professor, CSE Department, CVR College of Engineering, Hyderabad, India
[4]Assistant Professor, CSE Department, CMR Institute of Technology, Hyderabad, India
[5]Associate Professor, CSE (AI & ML) Department, CMR Technical Campus, Hyderabad, India

Article Info
Volume 9, Issue 5
Page Number : 92-100

Publication Issue
September-October-2022

Article History
Accepted : 01 Sep 2022
Published : 12 Sep 2022

ABSTRACT

In order to combine multimedia imagery and multispectral remote sensing data to analyze information, preprocessing becomes a necessary part of it. As one of the most important branches in the field of data analysis, it is widely used in many fields such as classification, regression, missing value filling, and machine learning. As a lazy algorithm, this method requires no prior statistical knowledge and no additional data to train description rules and is easy to implement. This study compares classification algorithm performances of data mining clustering algorithms for remotely sensed multispectral image data using WEKA data mining software. Clustering algorithm selection is very important for data mining classification method-based clustering. The class attribute for remotely sensed multispectral image data is obtained from six different clustering algorithms for classification. Classification algorithm performances computed depending on the data labeling of six different clustering algorithms in terms of correctly classified instances and kappa statistics for seven different classification algorithms. A strategy is developed for selecting the best unsupervised clustering algorithm, among different clustering algorithms, giving the highest supervised classification accuracy in terms of correctly classified instances and kappa statistics for semi supervised classification of remotely-sensed multispectral image data. The performances of seven semi-supervised classification methods assessed depending on six different unsupervised clustering algorithms for supervised classification of remotely sensed multispectral image data. This study determines data free clustering algorithms for classification.

Keywords: Data mining, remotely-sensed multispectral image data, clustering, classification

I. INTRODUCTION

Remote sensing image fusion is a technology that combines multisource remote sensing images through advanced image processing. It makes full use of the different characteristics of a variety of data, so that the image has a higher spectral and spatial resolution at the same time, and improves the vision of the image. The effect and accuracy of image feature recognition and classification accuracy Remote sensing image fusion is a hot research topic in the international remote sensing community in recent years. In the method of image fusion, there are some classic algorithms, such as HIS transformation method, COS transformation method, HIS transformation method, and HSV transformation method. In recent years, with the introduction of wavelet transform into the field of image processing, image fusion methods based on wavelet transform have attracted people's attention. Remotely sensed multispectral image data is big (stream) data. Data mining unsupervised clustering algorithms applied for labeling remotely sensed multispectral image data. With the rapid development of data collection and storage technology, there are plentiful unlabeled data but very few and often expensive labeled data in real-world applications.

Guerra et al.[1] studied the comparison between data mining supervised and unsupervised classifications of neuronal cell types. The increased developments of high-spatial resolution multispectral images improved the level of complexity required in data processing. To overcome this problem, several approaches investigated on both per-pixel and per-field based classifications of the remotely sensed multispectral image data.

Semi-supervised classification performances of data mining unsupervised clustering algorithms for remotely sensed multispectral image data is examined in this study to determine data free clustering algorithms. Unsupervised clustering algorithm selection is very important for labeling data in data mining. Selecting the best unsupervised clustering algorithm, among different clustering algorithms, giving the highest semi supervised classification accuracy in terms of correctly classified instances and kappa statistics is developed for semi supervised classification of remotely-sensed multispectral image data using WEKA data mining software [5]. The performances of semi-supervised classification methods assessed depending on different unsupervised clustering algorithms for supervised classification of remotely sensed multispectral image data. Data mining method applied is a semi-supervised classification method since it uses supervised classification based on unsupervised clustering of data. The results of supervised classification methods assessed and the performances of different unsupervised clustering algorithms compared for supervised classification of remotely sensed multispectral image data.

II. LITERATURE SURVEY

In the last years, a large interest has been devoted to the development of novel methodologies for multitemporal information extraction and analysis. This is demonstrated by the sharp increase in the number of papers published in the major remote sensing journals, the increased number of sessions in international conferences, and the increased number of projects related to multitemporal images and data.

The main reasons for this are: (i) the increased number of satellites with shorter revisit time that allow the acquisition of either long time series or frequent bitemporal images, (ii) the new open policy for data distribution of archive data that makes it possible a retrospective analysis on large scale (e.g., the Landsat Thematic Mapper archive), and (iii) the policies for the free distribution of many new satellites' data (e.g., Landsat 8, ESA Sentinels).

Multitemporal information extraction methodologies differ because of both the specific investigated application and the kind of data available. The most widely addressed applications are related to products obtained through multitemporal change detection, detection of land-cover transitions, and trend analysis of time series of data (for long terms change identification or forecasting/ prediction).

According to an information theory perspective, the information in multitemporal data is associated with the dynamic of the variables that are measured, which is linked with the changes occurring between successive acquisitions. Thus, the most interesting applications are related to the classification/integration of multitemporal data/images for the detection of changes. We can distinguish among abrupt changes, which occur in a short time (e.g., the ones caused by forest fires, floods, and earthquakes) and medium/long term changes, which can be appreciated only by comparing long time series of images (e.g., desertification, urban growth, and vegetation monitoring). The abovementioned applications can be addressed by using images acquired at different times by: (i) the same sensor, (ii) different sensors with similar properties (multisensor images), or (iii) different sensors with different properties (multisource images).

The main methodological approaches proposed in the literature for the analysis of changes in multitemporal remote-sensing images can be categorized by the use of bitemporal images or image time series. Accordingly, the following three groups of methods for multitemporal image analysis will be reviewed: (i) unsupervised bitemporal image analysis methods, (ii) supervised/semi supervised bitemporal image analysis methods, and (iii) image time series analysis methods.

1. Unsupervised bitemporal image analysis includes algorithms where change information is extracted by analyzing multitemporal features/images. Multitemporal information is associated with differences in the spectral signatures (or the backscattering coefficient) of the land-covers. After multitemporal comparison, the separation/classification between changed and unchanged areas (i.e., each pixel is associated with one of two possible classes: the class of changed patterns or the class of unchanged patterns) is performed by unsupervised approaches. Sometimes land-cover transitions can be distinguished, but without explicit labelling.

2. Supervised/semi-supervised bitemporal image analysis includes algorithms that elaborate the multitemporal signature using classification techniques. Approaches in this category are mainly supervised or semi-supervised. They explicitly identify land cover classes at each considered time and land-cover transitions are labelled accordingly (these methods can also be used when there are no changes between images for generating land-cover maps).

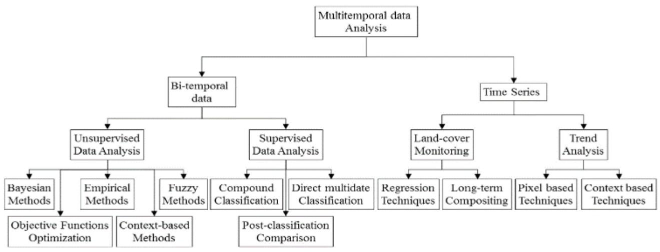

Fig. 1 Multitemporal data classification approaches.

3. Image time series analysis algorithms depend on the application goal and the temporal scale. At very high temporal resolution, changes such as those which occur on agricultural or urban areas become interesting and can be studied, whereas at lower temporal resolution, changes such as desertification/revegetation can be investigated. In both cases, a proper sampling of time series is required. Nevertheless, factors like atmospheric conditions or revisit period of the satellites can lead to

the availability of irregular and noncontinuous time series. Therefore, techniques that aim at building/studying time series by means of gap-filling, curve-fitting, and warping methods, and by using multisensory and multisource data can be found in the literature. Once proper time series are available, the analysis of different trends, at pixel or region levels, can lead to the classification/detection of interannual (seasonal) changes and inter-annual changes.

The first preprocessing step aims at making the multitemporal images radiometrically comparable. Ideally, a ground object should show the same brightness values if no change has occurred. In reality, measured intensity/backscattering values are sensitive to differences in acquisition geometry and environmental conditions (e.g., soil moisture for SAR systems, atmospheric contusions for optical passive systems). Radiometric conditions can be influenced by many factors such as imaging seasons, incidence angles, meteorological conditions, etc. Acquisition geometry, such as sensor viewing angle, local incident angle, and solar orientation have strong effects on the acquired images.

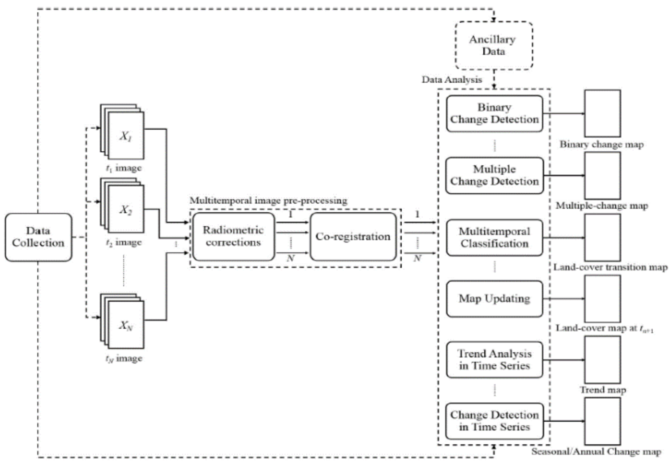

Fig. 2 General high level block scheme of multitemporal data analysis.

Atmospheric conditions have a serious impact on the measured radiance when using optical remotely sensed images. Absolute or relative normalization is often used to reduce this impact and make the multitemporal optical images comparable. Absolute normalization converts digital numbers to scaled surface reflectance and requires information about the atmospheric condition during image acquisition, which is not always easy to obtain. Relative normalization consists of linear transformation of spectral characteristics of the image in order to correct them to match those of a reference image. SARs are less affected by atmospheric condition.

III. RELATED WORK

A. Remotely Sensed Multispectral Image Data Analytics Remotely-sensed multispectral image data, used in this study, taken by Landsat Thematic Mapper [6]. There are three attributes (variables): first variable, taken as, X1, consists of spectral bands values of band 3, second variable, taken as X2, consists of spectral bands values of band 4 and third variable, taken as X3, consists of spectral bands values of band 5. There are 39600 instances (observations) or pixel values surrounding 198×200 data matrix.

Calis and Erol [2] applied a mixture of normal distribution models to determine the homogeneity of the per-field data structure. They classified remotely-sensed multispectral image data of an agricultural region on per-field basis using mixture discriminant analysis.

There are 5 crop types as agricultural features which were assigned into 24 homogeneous classes. Our analysis of data involved exporting the data from Excel worksheet to a Comma Separated Value (CSV) format, which was converted into Attribute Related File Format (ARFF) for use in WEKA data mining software. The original data labels thus, class attribute for remotely-sensed multispectral image data exists.

The class attribute consists of 24 homogeneous clusters.

Histograms for spectral brightness of spectral bands values of three attributes (variables): first attribute X1, second attribute X2, third attribute X3 and the class attribute obtained from original data are shown in Figure 1.

Data mining supervised classification methods based on selecting the best unsupervised clustering algorithm giving the highest supervised classification accuracy in terms of correctly classified instances and kappa statistics are applied for supervised classification of remotely-sensed multispectral image data. In unsupervised clustering algorithms, the aim is to discover groups of similar instances within the multispectral image data. There is a prior information about the class labels of data. There are 24 homogeneous clusters in the class attribute. Homogeneous clusters are supervised classified by using spectral signatures of crop types.

The best unsupervised clustering algorithm leading the best supervised classification of multispectral image data with highest classification accuracy in terms of correctly classified instances and kappa statistics [7] is selected and determined. B. Unsupervised Clustering Algorithms Applied For Supervised Classification Of Remotely Sensed Multispectral Image Data

Remotely sensed multispectral image data are classified by applying seven different classification methods: (1) Naïve Bayes [8] from Bayes group; (2) Multi Layer Perception from Functions group; (3) SMO from Functions group; (4) KStar from Lazy group; (5) Bagging from Meta group; (6) Decision Table from Rules group and (7) J48 from Trees group using WEKA data mining software [5]. The classification methods used in this study are all supervised classifications.

The class attribute for each classification method should be obtained from remotely sensed multispectral image data thus, actual big data using a clustering algorithm for comparing the performances in WEKA data mining software [5]. The class attribute for each classification method obtained from different unsupervised clustering algorithms. These unsupervised clustering algorithms are: (1) Class attribute obtained from Canopy clustering; (2) Class attribute obtained from EM clustering; (3) Class attribute obtained from Farthest First clustering; (4) Class attribute obtained from Filtered clustering; (5) Class attribute obtained from Make Density Based clustering; (6) Class attribute obtained from Simple K means clustering and (7) Class attribute obtained from original data.

Table.1. The Results for Classification of Remotely Sensed Multispectral Image Data Using Different Classification Methods Based on the Class Attribute Obtained from Canopy Clustering.

Number	Classifier	Correctly Classified Instances %)	Kappa Statics
1	Naïve Bayes	87.55	0.8609
2	Multi-Layer Perception	88.19	0.8511
3	SMO	89.25	0.8910
4	KStar	90.10	0.8308
5	Bagging	96.55	0.9504
6	Decision Table	92.88	0.9102
7	**J48**	**98.85**	**0.9815**

Results for classification of remotely sensed multispectral image data using different classification methods: (1) Naïve Bayes; (2) Multi-Layer Perception; (3) SMO; (4) KStar; (5) Bagging; (6) Decision Table and (7) J48; based on the class attribute obtained from EM clustering [9] is given in Table 2.

Table 2. The Results for Classification of Remotely Sensed Multispectral Image Data Using Different Classification Methods Based on the Class Attribute Obtained From EM Clustering.

Number	Classifier	Correctly Classified Instances %)	Kappa Statics
1	Naïve Bayes	96.55	0.8611
2	Multi-Layer Perception	89.19	0.8521
3	SMO	88.25	0.8901
4	KStar	90.10	0.8305
5	Bagging	98.55	0.9513
6	Decision Table	92.88	0.9111
7	**J48**	**99.45**	**0.9915**

Results for classification of remotely sensed multispectral image data using different classification methods: (1) Naïve Bayes; (2) Multi Layer Perception; (3) SMO; (4) KStar; (5) Bagging; (6) Decision Table and (7) J48; based on the class attribute obtained from Farthest First clustering is given in Table 3.

Table 3. The Results for Classification of Remotely Sensed Multispectral Image Data Using Different Classification Methods Based on The Class Attribute Obtained from Farthest First Clustering.

Number	Classifier	Correctly Classified Instances %)	Kappa Statics
1	Naïve Bayes	85.23	0.8210
2	Multi-Layer Perception	93.24	0.9421
3	SMO	91.27	0.9104
4	KStar	95.10	0.8708
5	Bagging	98.55	0.9832
6	Decision Table	94.68	0.9302
7	**J48**	**99.62**	**0.9913**

Results for classification of remotely sensed multispectral image data using different classification methods: (1) Naïve Bayes; (2) Multi-Layer Perception; (3) SMO; (4) KStar; (5) Bagging; (6) Decision Table and (7) J48; based on the class attribute obtained from Filtered clustering is given in Table 4.

Table 4. The Results for Classification of Remotely Sensed Multispectral Image Data Using Different Classification Methods Based on The Class Attribute Obtained from Filtered Clustering.

Number	Classifier	Correctly Classified Instances %)	Kappa Statics
1	Naïve Bayes	95.23	0.9420
2	Multi-Layer Perception	93.29	0.9402
3	SMO	91.25	0.9131
4	KStar	95.10	0.9210
5	Bagging	98.55	0.9812
6	Decision Table	93.68	0.8905
7	**J48**	**99.15**	**0.9911**

Results for classification of remotely sensed multispectral image data using different classification methods: (1) Naïve Bayes; (2) Multi-Layer Perception; (3) SMO; (4) KStar; (5) Bagging; (6) Decision Table and (7) J48; based on the class attribute obtained from Make Density Based clustering [10] is given in Table 5.

Table 5. The Results for Classification of Remotely Sensed Multispectral Image Data Using Different Classification Methods Based on the Class Attribute Obtained from Make Density Based Clustering.

Number	Classifier	Correctly Classified Instances %)	Kappa Statics
1	Naïve Bayes	95.55	0.9406
2	Multi-Layer Perception	92.29	0.9423

3	SMO	94.25	0.9108
4	KStar	95.10	0.9224
5	Bagging	98.55	0.9810
6	Decision Table	89.68	0.8908
7	**J48**	**99.16**	**0.9910**

Results for classification of remotely sensed multispectral image data using different classification methods: (1) Naïve Bayes; (2) Multi-Layer Perception; (3) SMO; (4) KStar; (5) Bagging; (6) Decision Table and (7) J48; based on the class attribute obtained from Simple K means clustering is given in Table 6.

Table 6. The Results for Classification of Remotely Sensed Multispectral Image Data Using Different Classification Methods Based on The Class Attribute Obtained from Simple K Means
Clustering.

Number	Classifier	Correctly Classified Instances %)	Kappa Statics
1	Naïve Bayes	95.23	0.9420
2	Multi-Layer Perception	93.29	0.9402
3	SMO	91.25	0.9131
4	KStar	95.10	0.9210
5	Bagging	98.55	0.9812
6	Decision Table	93.68	0.8905
7	**J48**	**99.10**	**0.9908**

Results for classification of remotely sensed multispectral image data using different classification methods: (1) Naïve Bayes; (2) Multi-Layer Perception; (3) SMO; (4) KStar; (5) Bagging; (6) Decision Table and (7) J48; based on the class attribute obtained from original data clustering is given in Table 7.

Table 7. The Results for Classification of Remotely Sensed Multispectral Image Data Using Different Classification Methods Based on The Class Attribute Obtained from Original Data Clustering.

Number	Classifier	Correctly Classified Instances %)	Kappa Statics
1	Naïve Bayes	95.23	0.9420
2	Multi-Layer Perception	93.29	0.9402
3	SMO	91.25	0.9131
4	KStar	95.10	0.9210
5	Bagging	98.55	0.9812
6	Decision Table	93.68	0.8905
7	**J48**	**99.20**	**0.9918**

IV. CONCLUSION

Supervised classification algorithms are used to inferring a function from labeled training data, of which an algorithm analyzes the training data and inferred function are produced, thus creating for new entries to be mapped using a defined set of rules. The best model will allow for the algorithm to correctly determine the class labels for unseen instances. The classification algorithms of WEKA data mining software [5] were used for the supervised classification of remotely sensed multispectral image data. The classification algorithm organizes observations into groups, by splitting the data based on homogeneity, where the classification is based on similarity until or purity is achieved. The K-means algorithm was used as a precursor to reduce the multispectral image data into much smaller space (k=24), for finding meaningful structure in the multispectral image data. Seven different unsupervised clustering algorithms: (1) Class attribute obtained from Canopy clustering; (2) Class attribute obtained from EM clustering; (3) Class attribute obtained from Farthest First clustering; (4) Class attribute obtained from Filtered clustering; (5) Class attribute obtained from Make Density Based clustering; (6) Class attribute obtained from Simple K means clustering and (7) Class attribute obtained from original data; were used for labeling the proposed data into the 24 known clusters [2]. This approach was succeeded by classification of features into the known clusters using the proposed seven different data mining classification methods for classification of remotely sensed multispectral image data using WEKA data mining software. The percentages of correctly classified instances and kappa statistics for each of the proposed seven different classification algorithms were explained. In the semi supervised classification approach, the J48 classifier from trees group of WEKA implementation performed better than other classifiers with highest classification accuracy in terms of correctly classified instances and kappa statistics for the best supervised classification of multispectral image data. It is concluded that using the true number of clusters for data labeling and constructing the class attribute for big data is very important in clustering algorithms. This study determines that J48 classification algorithm, among seven different classification algorithms, is data free classification algorithm and it is independent from clustering algorithms for classification of remotely sensed multispectral image data. The data free classification algorithm means that which clustering algorithm, among six different clustering algorithms, were used for data labeling for obtaining the class attribute is not important or semi-supervised classification using J48 classifier algorithm, among seven different classification algorithms, for classification of remotely sensed multispectral image data in this study. J48 classifier algorithm gives the highest classification accuracy in terms of correctly classified instances and kappa statistics for the best supervised classification of multispectral image data for all of six different unsupervised clustering algorithms.

V. REFERENCES

[1]. Bala Mikat, "Classification Performances Of Data Mining Clustering Algorithms For Remotely Sensed Multispectral Image Data", 978-1-5386-5150-6/18,©2018 IEEE.

[2]. Yingxin Sun, "Multispectral Remote Sensing Data Analysis Based on KNNLC Algorithm and Multimedia Image",Journal of Sensors Volume 2022.

[3]. Ravindra Changala, "Development of Predictive Model for Medical Domains to Predict Chronic Diseases (Diabetes) Using Machine Learning Algorithms and Classification Techniques", ARPN Journal of Engineering and Applied Sciences, VOL. 14, NO. 6, March 2019, ISSN 1819-6608.

[4]. R. Jin, Y. Wang, C. Niu, and W. Song, "A novel high precision and low consumption indoor positioning algorithm for internet of things," IEEE Access, vol. 7, pp. 86874–86883, 2019.

[5]. Ravindra Changala, "Evaluation And Analysis Of Discovered Patterns Using Pattern Classification Methods In Text Mining", ARPN Journal of Engineering and Applied Sciences, Vol3,Issue 11.

[6]. J. Ren, Y. Wang, C. Niu, W. Song, and S. Huang, "A novel clustering algorithm for wi-fi indoor positioning," IEEE Access, vol. 7, pp. 122428–122434, 2019.

[7]. V. Torra, Y. Narukawa, A. Honda, and S. Inoue, "Modeling decisions for artificial intelligence," 14th International Conference, MDAI 2017, 2017, pp. 77–88, Kitakyushu, Japan, 2017.

[8]. V. Subramaniyaswam and R. Logesh, "Adaptive KNN based recommender system through mining of user preferences," Wireless Personal Communications, vol. 97, no. 2, pp. 2229– 2247, 2017.

[9]. Ravindra Changala, "A Survey on Development of Pattern Evolving Model for Discovery of Patterns In Text Mining Using Data Mining Techniques" in Journal of Theoretical and Applied Information Technology in 31st August 2017. Vol.95. No.16, ISSN: 1817-3195, pp.3974-3987.

[10]. Z. Qin, T. Kirubarajan, and Y. Liang, “Application of an efficient graph-based partitioning algorithm for extended target tracking using gm-phd filter,” IEEE Transactions on Aerospace and Electronic Systems, vol. 56, no. 6, pp. 4451–4466, 2020.

[11]. N. Bhaskar and P. M. Kumar, “Optimal processing of nearest neighbor user queries in crowdsourcing based on the whale optimization algorithm,” Soft Computing, vol. 24, no. 17, pp. 13037–13050, 2020.

[12]. Ravindra Changala “A Survey1 on Clustering Techniques to Improve Energy Efficient Routing in Wireless Sensor Networks” in International Journal of Applied Engineering Research, 10(58), pp.-1-5,2015.

[13]. Calis N., and Erol H., (2012). A new per-field classification method using mixture discriminant analysis, Journal of Applied Statistics, vol.39, pp.2129-2140.

[14]. Xi Chen (2014). Semi-Supervised Classification Method Based on Spectral Clustering. Journal Of Networks, VOL. 9, NO. 2, 384-392.

[15]. Ravindra Changala, "Data Mining Techniques for Cloud Technology" in International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Volume 4, Issue 8, Pages 2319-5940, ISSN: 2278-1021, August 2015.

[16]. Duda, R. and Hart, P. (1973), Pattern Classification and Scene Analysis, Wiley.

[17]. Ravindra Changala “A Survey on Development of Pattern Evolving Model for Discovery of Patterns in Text Mining Using Data Mining Techniques” in Journal of Theoretical and Applied Information Technology, August 2017. Vol.95. No.16, ISSN: 1817-3195, pp.3974-3987.

[18]. Russell G. Congalton (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote Sensing of Environment 37(1):35-46.

Cite this article as :

Dr. Bodla Kishor, Dr. Y Sowmya Reddy, Dr. K. Rubenraju, Dr. Mahesh Kotha, "A Novel Classification Performance Approach for Remotely Sensed Multispectral Image Data by Using Data Mining Techniques", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 5, pp. 92-100, September-October 2022. Available at doi : https://doi.org/10.32628/IJSRST229516

Journal URL : https://ijsrst.com/IJSRST229516