ISSN: 2366-1313

PREDICTING STUDENTS FINAL GPA USING DECISION TREE

[1]VEGOLUM LAHARIKA, [2]Dr. S. VISWANADHA RAJU

[1] M.Tech Scholar, [2]Professor, Department of CSE,

JNTUH UNIVERSITY COLLEGE OF ENGINNERING, JAGTIAL, T.S., INDIA.

ABSTRACT:-

The paper deals with predicting grade point average (GPA) with supervised machine learning models. Based on the literature review, we divide the factors into three groups—psychological, sociological and study factors. Data from the questionnaire are evaluated using statistical analysis. We use confirmatory data analysis, where we compare the answers of men and women, university students coming from grammar schools versus students coming from secondary vocational schools and students divided according to the average grade. The differences between groups are tested with the Shapiro–Wilk and Mann–Whitney U-test. We identify the factors influencing the GPA through correlation analysis, where we use the Pearson test and the ANOVA. Based on the performed analysis, factors that show a statistically significant dependence with the GPA are identified. Subsequently, we implement supervised machine learning models. We create 10 prediction models using linear regression, decision trees and random forest. The models predict the GPA based on independent variables. Therefore, we recommend the use of a random forest as a starting model for modeling student results.

Keywords:- machine learning; prediction; statistical modeling; education; GPA; random forest; linear regression; student.

ISSN: 2366-1313

1. INTRODUCTION

Data Mining DM is the process of finding patterns, correlations and anomalies within data to extract knowledge from raw data Educational Data Mining (EDM) is one of the areas that elicited some interest in researchers. Several DM methods have been used within the area including classification, association rule mining and clustering. Our review of the area has shown that classification was the most popular DM method that has been used by researchers. That could be because of the high potential of using prediction in EDM, the availability of rich data stored in the systems and the diversity of classifications methods.

One of the classical prediction problems in the field is to predict student performance and discover students who are most probable to fail in order to provide help and support. Also, data mining can help instructors and administrators to make better-informed decisions when designing courses and programs [1]. For example, knowing which course has the most effect on students' performance can encourage educators to give more attention to such course and provide more resources like better instructors, more supportive materials, etc. Achieving such results can be done by using DM methods and since DM is one of Knowledge Discovery in Database (KDD) process' stages, research can apply KDD Process to find and extract knowledge from data.

KDD include many stages that help to have an accurate prediction and transform the raw data to information. These stages are preprocessing data, mining data and post processing data. Preprocessing data include selecting the target data, transforming data, etc. Mining data is to extract patterns. Postprocessing data include interpreting and evaluating these patterns. Every stage had its methods, techniques, algorithms and suitable tools. Selecting the right choice for each stage from these depends on the nature of the dataset. For instance, selecting the target data can be done by extracting the data from database or combining the data from several sources like XML files and databases.

In the preprocessing stage, the dataset could suffer from a missing data problem, and this problem could be handled by using several methods like if the

ISSN: 2366-1313

dataset is huge then deleting the records that have missing data is an option. However, If the dataset is not huge then there are other options like depending on the column type that have the missing data. For instance, if it is a categorical then replacing the missing with most frequent value but if it is a numerical then replacing the missing values with the average of that column. In addition, the dataset could need feature engineering by creating new features or deleting some of the features. Also, depending on the available attributes, sometimes we need to select the useful attributes manually based on our domain knowledge, and at other times we could use selecting feature techniques.

Machine learning is a subset of artificial intelligence (AI) that helps computers or teaching machines learn and make intelligent decisions from all previous data. The architecture for machine learning involves collecting and storing a rich set of information and turning it into a standardized knowledge base for various uses in different fields. Predicting student performance is a daunting issue faced every year by educational institutions such as colleges, schools, and training centers. As a result, forecasting the performance of students at an earlier stage would enable the educational institutions to find solutions to avoid negative student performance. Lectures should predict their student's success and find appropriate learning strategies to improve the performance of the students. It can also strengthen the institution's admission policies and help students resolve their grades.

ISSN: 2366-1313

2. RELATED WORKS

Much research has been done in the area of educational data mining where a predictive model is built to forecast the performance of students to identify the at risk students. This problem can be considered a hard problem because the performance depends on many characteristics related to the students. These characteristics can be categorized into student's GPA and grades, demographics, psychological profile, culture, academic progress, and educational background [2]. The student's GPA is the most important attribute used to predict the performance. The GPA can represent the real value for the future educational and career possibilities and progression. In addition, the academic potentials can be evaluated by the student GPA. The demographics information that consists of the family background, the gender, disability, and age is also considered an important attribute [3]. This research introduces two new attributes that focus on using descriptive features related to the internet and social network usage and their effect on the performance. On the other hand, many machine learning and data mining techniques have been used to predict the students' performance such as: Artificial Neural Network (ANN); K-Nearest Neighbor (KNN); Support Vector Machine (SVM); Linear Regression; Logistic Regression; Decision Tree (DT); Random Forest (RF); Principal Component Analysis (PCA); Naïve Bayes (NB); Neuro-Fuzzy classification (NF); Decision List (DL); Bayesian Network (BN); and Discriminant Analysis (DA).

Machine learning for education systems was extensively investigated [1], where five separate fields were defined: prediction, model discovery, data extraction for human judgment, clustering, and mining relationships. Most previous educational systems work, apply to universities or virtual learning [10]. The data collected in all previous works, either from surveys or from e-learning programs. Kapur et al. [11] used two different methods of machine learning (i.e. J48 Decision Tree, and Random Forest) to predict marks on students in the field of education. The data collected consists of 480 entries which relate to the enrollment of the student. Veracano et al. [12] used various methods of machine learning to estimate students dropping out for unbalanced data set. The authors collect 419 samples from a single high school in

Mexico. Saif and. Al [13] looks at a number of courses and discusses whether successful or bad performance can be expected. Saarela et al. [14] introduced a method for predicting the level of difficulty of various mathematical problems and for predicting whether or not the students could answer those questions.

In a study by Esmat et al. [3], the authors studied student success according to how a student progresses in their studies and stays in school. The aim of the study was to examine the admission process for the university program. They examined the students over a period of six years. Finally, authors found out that students who attended preparatory courses before the start of their studies had better prospects to succeed during their studies than students who had not attended the preparatory courses.

2.1 Related Works On HSI Classificatoin

In the early stage of the study on HSI classification, most methods have focused on exploring the role of the spectral signatures of HSIs for the purpose of classification. Thus, numerous pixel-wise classification methods (e.g., neural net-works [7], support vector machines (SVM) [8], multinomial logistic regression, and dynamic or random sub-space) have been proposed to classify HSIs. In addition, some other classification approaches have focusedon designing an effective feature-extraction or dimension-reduction technique, such as principle component analysis(PCA) , independent component analysis (ICA) [15],and linear discriminant analysis (LDA) [16]. However, the classification maps obtained by these pixel-wise classifiers are unsatisfactory since the spatial contexts are not considered. Recently, spatial features have been reported to be very useful in improving the representation of hyper spectral data and increasing the classification accuracies. More and more spectral-spatial features-based classification framework shave been developed, which incorporates the spatial contextual information into pixel-wise classifiers. For example, in, extended morphological profiles (EMPs) were used to exploit the spatial information via multiple morphological operations. Multiple kernel learning (e.g., composite kernel and morphological kernel was designed to explore the spectral-spatial information of HSIs.

ISSN: 2366-1313

3. THE PROPOSED SYSTEM

3.1 The System Components

The following diagram, figure 1, shows the main steps and components of the proposed machine learning system.

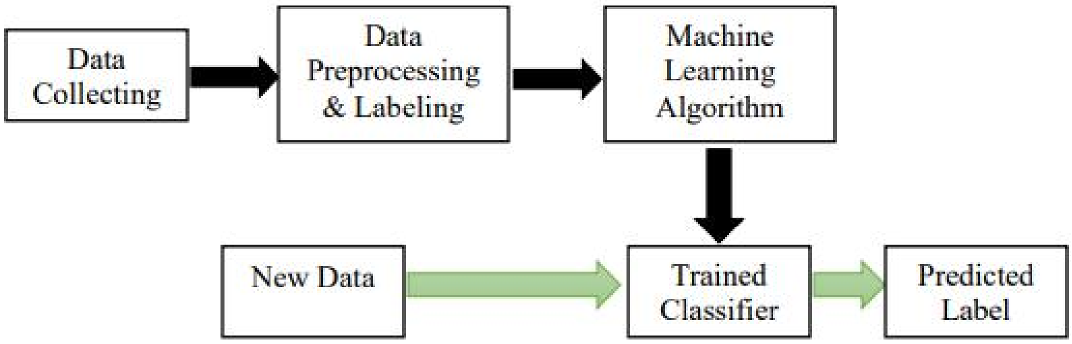


Figure 1-The Main Steps and Components of the Proposed System

The first step is collecting the data from the data sources. In our case, the data has been collected using a survey given to the students and the students' grade book. The second step is preprocessing the data in order to get a normalized dataset and then labeling the data rows. In the third step, the result of the second step, the training and testing dataset, is fed to the Machine Learning algorithm. The Machine Learning Algorithm builds a model using the training data and tests the model using the test data. Finally, the Machine Learning Algorithm produces a trained model or a trained classifier that can take as an input a new data row and predicts its label.

3.2 The Methodology

In this section, a brief review of the machine learning techniques that is used in this research is introduced. A decision tree model represents a tree structure that is similar to a flowchart. In this structure, each internal node represents a test on a dataset attribute while each tree branch represents the test outcome. In addition, each

ISSN: 2366-1313

leaf node represents a target feature label and the upper first node in the tree represents the root node. Decision trees can be a binary or a non-binary trees. Decision trees are popular classification techniques because using them does not need prior knowledge of the problem domain or a complicated setting of the classification parameters. In addition, they can be converted to classification rules easily and they can be understood easily. Decision tree classification technique has been used in many real word applications such as financial analysis, medicine, molecular biology, manufacturing production, and astronomy. During building the decision tree, the algorithm uses an attribute or feature selection measure which is used in selecting the attribute or the feature that best divides the dataset instances into distinct target classes. Such measures include the Information Gain, Gain Ratio, and Gini Index. Popular decision trees algorithms include ID3, CART, and C4.5[16].

A three layer fully connected feed forward ANN has been used in this research. The network consists of an input layer, two hidden layers, and the output layer. The input layer has twenty input unites, neurons, while the first hidden layer has six hidden unites. The second hidden layer has three hidden unites. The fourth layer is the output layer which has only one output unite. The Rectifier Linear Unit has been used as the hidden unites' activation function [15].

Data Preprocessing The raw dataset that we obtained was in the Arabic language, and it contained students' transactional data of computer and information college over 7 years (approximately from 2010 to 2017). The number of transactions was 301,078 rows and 16 attributes for 5566 students. Each transaction records student personal information like ID, gender, DOB and city as well as student academic information like academic status, major, semester GPA, cumulative GPA (CGPA), course name, code, term, grade, number or register students, and more.

ISSN: 2366-1313

4. THE EXPERIMENT

4.1 Dataset and Data Sources

The dataset used in this research is collected from the Archeology department and the Sociology department of the college of Humanities at Al-Muthanna University during the 2015 and 2016 academic years. Two data sources have been used, survey collected from the students and the students' grades data records. The dataset contains 161 student records, 76 male and 85 female. The dataset contains twenty attributes. The attributes can be divided into five categories which are personal and life style, studying style, family related, educational environment satisfaction, and student's grades. Table2 shows the attributes used in order to construct the dataset. Each student has been labelled as Weak or Good based on his/her final grade in the computer science subject. The weak student is the student who has a final grade less than sixty out of 100. On the other hand, the Good student is the student who has a final grade equal or greater than sixty. There are 75 students with Good status and 86 students with Weak status Identifying the weak status students is more important than identifying the good status students, therefore the weak status is considered a positive value of the target attribute. Computer Grade-Course1 attribute represents the average of the first two monthly exams in the computer science subject during the first course. The academic year contains two semesters or courses, midterm exam, and final exam. Each of the semesters has two monthly exams. To predict the students who need support as early as possible the grade of the first semester, the average of the first two exams, has been chosen as an attribute because it could be an indicator of the final student performance. By doing that, the weak students will have an early opportunity to enhance their academic performance. In addition, the faculty members can provide the appropriate support to the students as early as possible. Similarly, English Grade-Course1 attribute represents the average of two exams in the English subject during the first semester.

ISSN: 2366-1313

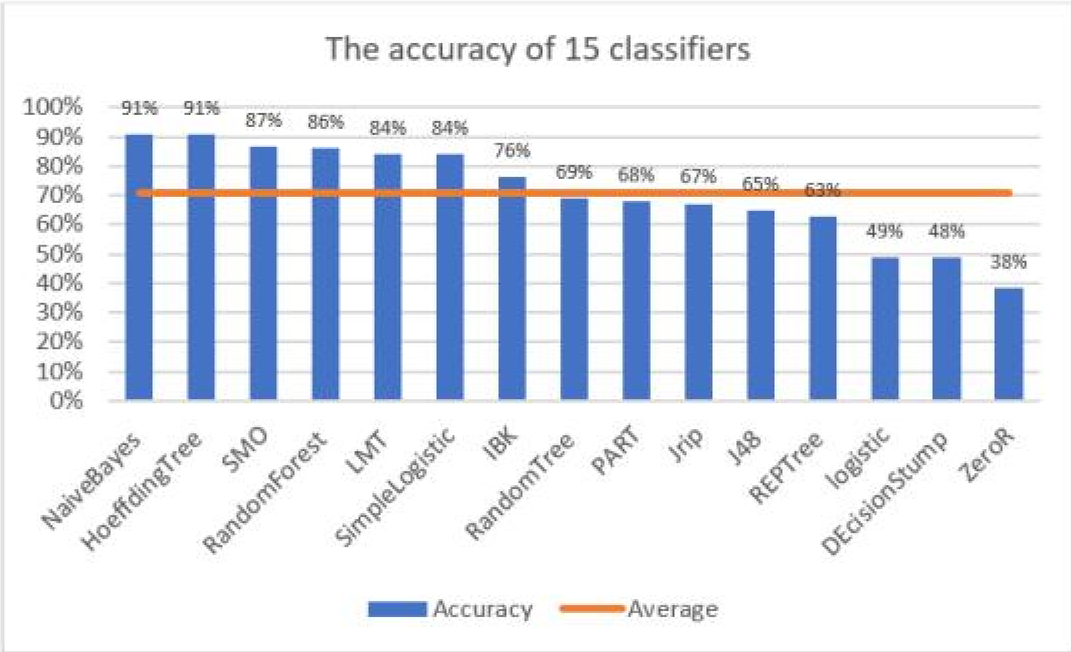


Figure 1: Comparing between the performance of the 15 classifiers and their average.

The English subject has been chosen as an attribute because of its relation with the computer science subject as most of the computer educational materials have been taught and presented in English.

3. DATASETS OF HYPERSPECTRAL IMAGES (HSIS)

HSIs captured by airborne and spaceborne sensors are very useful in many applications like remote sensing [3], land-cover [4], agriculture [5] etc. These sensors collect reflective portion of electromagnetic spectrum containing hundreds of narrow spectral bands and this reflective portion creates a unique spectral signature for an object. This allows the identification of various materials on earth surface by the unique spectral signature. As the sensors capture HSIs are very expensive, only few HSI datasets are publicly available [5]. Also, the task of creating groundtruth and pixel-labelling is exorbitant and time-demanding [1] and consequently only a few labelled HSIs samples are available for research work. Moreover, the DL-based models need an adequate number of training samples for parameter tuning during the training phase [2]. Thus, the availability of very less number of HSI samples makes the classification task very challenging [3]. Nevertheless, some of the HSI datasets are publicly available and in this section three most popular datasets i.e., Pavia University, Indian Pine, and Salinas are discussed.

3.1 Dataset of Indian Pine

Dataset of Indian Pine has been captured as first dataset by airborne-visible-infrared-imagingspectrometer (AVIRIS) sensor over a site in northwestern, Indiana, USA [8]. The spatial resolution of image is 20m with a spatial size of 145×145 pixels. The spectral range of the acquired image is 0.4- 2.5µm with a sum of 224 number of spectral bands. However, after removing water absorbed and noisy bands, only 200 spectral bands are utilized in experiments. Also, the image contains 16-classes with total 10366-samples in the dataset as depicted in Table 1.

ISSN: 2366-1313

Table 1. Samples and groundtruth classes of Indian Pine Dataset

Sr. No.	Class	Samples
1	Oats	20
2	Corn	237
3	Woods	1265
4	Wheat	205
5	Alfalfa	46
6	Grass-trees	730
7	Soybean-mintill	2455
8	Corn-mintill	830
9	Grass-pasture	483
10	Soybean-clean	593
11	Soybean-notill	972
12	Corn-notill	1428
13	Hay-windrowed	478
14	Stone-steel-towers	93
15	Grass-pasture-mowed	28
16	Buildings-grass-trees-drives	386
	Total	**10366**

3.2 Dataset of Pavia University

Reflective-optics-spectrographic-imaging-system (ROSIS) sensor was used to capture Pavia university dataset over the Pavia University, northern Italy [85]. The image contains a spatial resolution of 1.3m with a spatial size 610×340 pixels. The original image contains total 115 spectral bands in a spectral range 0.43-0.86μm. However, a total of 103 spectral bands are utilized in experiments after discarding noisy bands. Also, total of 9-classes having 42776-samples are available in the dataset as shown in Table 2.

ISSN: 2366-1313

Table 2. Samples and groundtruth classes of Pavia University Dataset

Sr. No.	Class	Samples
1	Trees	3064
2	Gravel	2099
3	Asphalt	6631
4	Bitumen	1330
5	Shadows	947
6	Meadows	18649
7	Bare soil	5029
8	Painted metal sheets	1345
9	Self-blocking bricks	3682
	Total	**42776**

3.3 HYPERSPECTRAL IMAGES CLASSIFICATION

The HSI classification process is shown in Figure. The classification process consists of the major steps of data input, data pre-processing, feature information extraction and feature map activation, classification model, accuracy evaluation and classification results.

The pre-processing of HSI mainly includes image format conversion, geometric correction, noise reduction, dimensionality reduction, etc. The purpose is to eliminate noise and reduce the complexity of HSI as much as possible to improve the operation efficiency and provide data for the subsequent classification model.

Feature extraction and feature selection is also essentially a dimensionality reduction, a process of finding the optimal solution, and commonly used methods include Principal Components Analysis (PCA), which uses linear transformations to extract features, but hyperspectral data is inherently nonlinear, so linear transformation methods such as PCA can lose a lot of useful information. Choosing a suitable feature extraction and classification model is the key to achieve high classification accuracy.

ISSN: 2366-1313

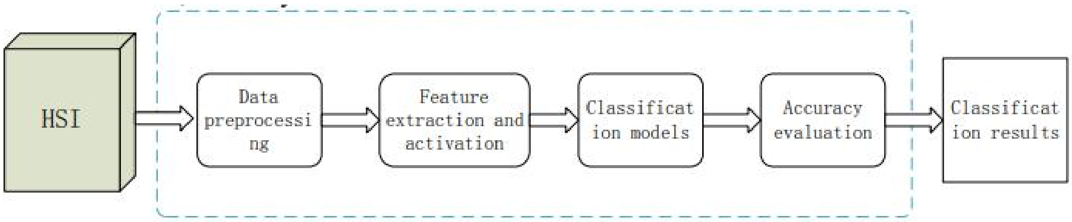


Figure1. Hyperspectral images classification process

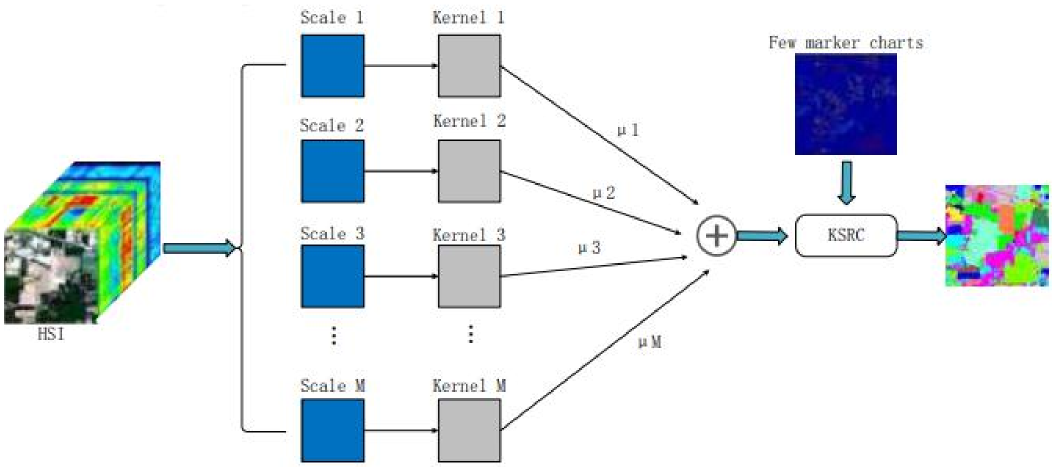


Figure2: Hyper spectral images classification model based on multi core fusion

Traditional methods can only extract limited spectral feature information, while the spatial spectral joint feature classification-based methods can extract not only spectral feature information but also spatial feature information and perform effective feature fusion, which can effectively fit the nonlinear relationship between the classification labels of high HSI and HSI data features for high dimensional data like HSI to obtain better classification results. On the other hand, the joint spatial spectral feature classification model integrates feature extraction and feature classification into one framework, which can achieve end-to-end training.

ISSN: 2366-1313

4. DEEP LEARNING

In recent years, hyper spectral image classification methods have introduced spatial information of hyper spectral images. This type of method is simply referred to as hyper spectral image classification methods based on spatial-spectral joint features. Deep learning originates from artificial neural networks. Compared with artificial neural networks, deep learning has a stronger pumping ability. Deep learning models have deeper layers, which also helps to extract feature information. This section mainly introduces Convolutional Neural Networks (CNN) in deep learning [9, 15], deep belief network (DBN), and Stacked Auto Encoder (SAE).

4.1. CNN Classification method based on spectral features:

Hyper spectral images have very rich spectral information and extremely high spectral resolution. Each pixel can extract one-dimensional spectral vectors. These vectors are composed of spectral information. Classification using only one-dimensional spectral vectors is called a classification method based on spectral information. In the classification method based on spectral information, generally, the pixel

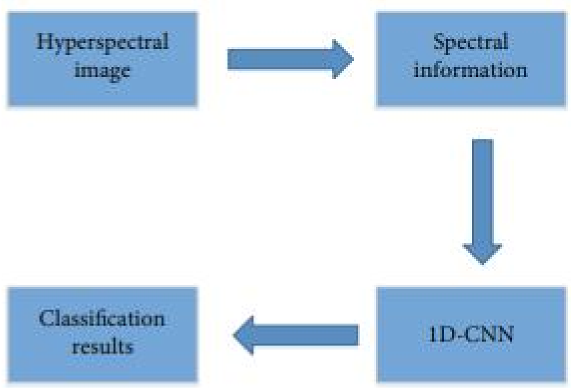


Figure1: Schematic diagram of 1D-CNN

is used to extract spectral information or to obtain certain specific features from spectral information through feature extraction to classify. Using CNN to classify spectral features of hyper spectral images is to use one-dimensional CNN (1DCNN) to extract spectral features and classify them. The process is shown in Figure 1.

CNN to extract spatial features. The classification is completed by combining the class prediction scores of both the streams. Similarly, in [13], another hybrid framework for multi-feature based HSI classification is proposed which uses PCA to reduce dimensionality, guided filters to get spatial features and a sparse AE to extract high level features. The authors in [14], have proposed a batch-based training scheme for AEs to exploit spectral-spatial features and these features are merged via a mean pooling scheme. Likely, a classification framework, exploiting spectral-spatial, is developed in [15] to utilize stacked sparse AE for feature extraction and random forest classifier for final classification. Furthermore in [16], the authors have used a threefold feature learning scheme proposed in [17] to implement an efficient multilayer extreme learning machine-based AE framework. In work [18], the authors have addressed the issue of high inter-class-similarity and high intra-class-variability.

Here a stacked model is used to learn discriminative features via imposing a local fisher discriminant regularization. In the work proposed in [19], extended morphological profiles are utilized to incorporate spatial information within the spectral information obtained from spectral segments. The proposed scheme has been very effective in terms of time complexity. Recently, [18], a k-sparse denoising AE is knitted with spectral-spatial features is employed for HSI-classification. In this work, the spatial features are obtained through restricted spatial information in order to reduce intra-class variability of spatial features.

ISSN: 2366-1313

5. EXPERIMENTS

In this section, we mainly conduct a comprehensive set of experiments from four aspects. Firstly, a series of experiments are designed to demonstrate the advantages of deep learning on HSI classification over traditional methods. Secondly, the classification performance of several recent state-of-the-art deep learning approaches is systematacially compared. Thirdly, we visualize the learned deep features and network weights to further explore the “black box”. Finally, the effectiveness of strategies included in Section IV is further analyzed. To complete our experiments, three benchmark HSIs are used, i.e., the Houston, University of Pavia, and Salinas images. The three images are introduced in the following subsection.

5.1 Experimental Data Sets

The Houston data was distributed for the 2013 IEEE Geoscience and Remote Sensing Society (GRSS) data fusion contest. This scene was captured in 2012 by an airborne sensor over the area of University of Houston campus and the neighboring urban area. The size of the data is 349×1905 pixels with a spatial resolution of 2.5 m. This HSI consists of 144 spectral bands with wavelength ranging from 0.38 to 1.05 µm and includes 15 classes.

5.2 Compared Methods

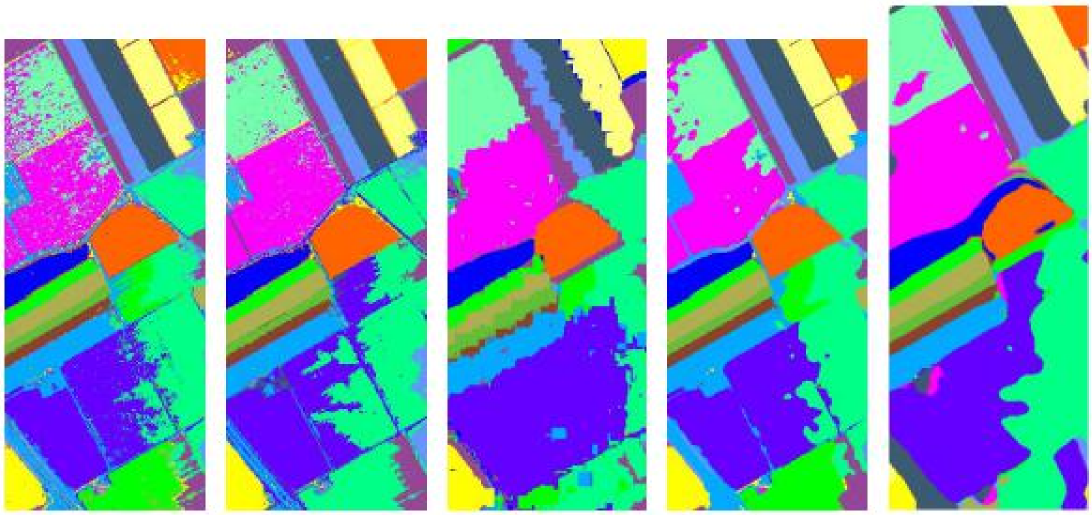
In this review, we have investigated several recent stateof-the-art deep learning-based approaches, including 3D-CNN [6], Gabor-CNN [9], CNN with pixel-pair features (CNNPPF) [6], siamese CNN (S-CNN), 3D-GAN , and the deep feature fusion network (DFFN) [7], for HSI classification. Specifically, the 3D-CNN exploits 3-D convolutional filters to directly extract spectral-spatial features from the original hyperspectral cube. However, the network architecture adopted in 3D-CNN is relatively simple and the correlation between different layers is neglected. The DFFN adopts the DRN [10], which can be considered a more powerful network, to extract more discriminative features. In addition, the features from different-level layers are

ISSN: 2366-1313

further fused to explore the correlation between layers. The main drawback of DFFN is that the optimal feature fusion mechanism depends on a hand-crafted setting with abundant experiments. In the GaborCNN, the Gabor filtering is first utilized as a preprocessing technique to extract spatial features of HSIs.

5.3 Classification Results

The first experiment was performed on the Houston data set. In this experiment, the training samples were given according to the 2013 GRSS data fusion contest. The amount of training and test samples per class is shown in Table I. The classification maps obtained by different methods. From this figure, we can see that the classification maps obtained by the SVM and JSR methods are not very satisfactory since some noisy estimations are still visible. By contrast, other methods perform much better in removing "noisy pixels" and deliver a smoother appearance in their classification results. By comparing two filtering-based methods, i.e., EPF and GaborCNN, we can see that the classification map of EPF seems to be over-smoothing, but the Gabor-CNN preserves more details in edges. Apart from visual comparison, Table IV gives quantitative results of various methods on the image, where three metrics, i.e., overall accuracy (OA), average accuracy (AA), and Kappa coefficient, are adopted to evaluate the c lassification performance.



(a) SVM (b) EMP (c) JSR (d) EPF (e) 3D-CNN

Figure3: The Configuration Of CNN Model

ISSN: 2366-1313

TABLE 3: THE CONFIGURATION OF CNN MODEL

Layer No.	Type	Size	Feature maps
1-3	Convolution	3×3	16
4-5	Convolution	3×3	32
6-7	Convolution	3×3	64
8	Pooling	3×3	64
9	Fully connected	-	200
10	Fully connected	-	16

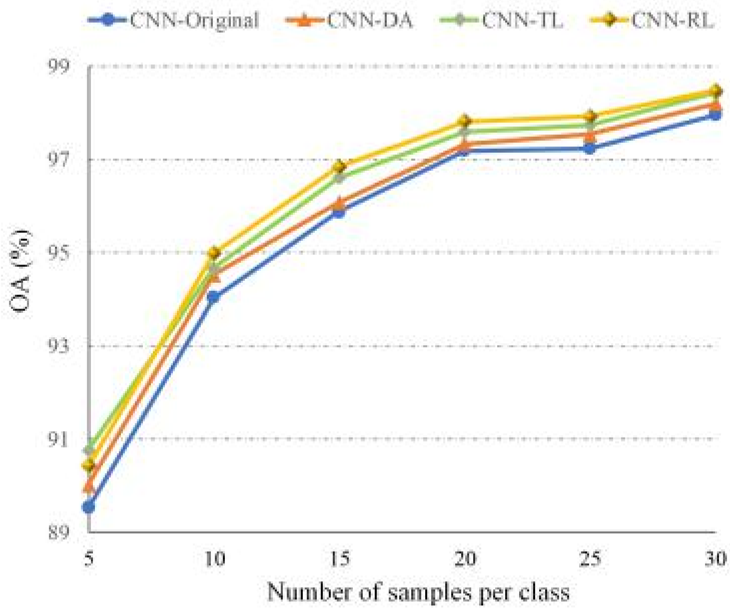


Figure 4: Values obtained by different methods versus number of samples.

ISSN: 2366-1313

6. CONCLUSION

Classification and recognition of hyper spectral images are important content of hyper spectral image processing. This paper discusses several methods of hyper spectral image classification, including supervised and unsupervised classification and semi supervised classification. Although the supervised and unsupervised classification methods described in this article have their respective advantages to varying degrees, there are limitations in the application of various methods. For example, supervised classification requires a certain number of prior conditions, and human factors will affect the classification results have an impact. Therefore, based on different application requirements, combined with the acquisition of hyper spectral images with massive information, multiple methods need to be combined with each other in order to achieve the desired classification effect. With the development of hyper spectral image technology, hyper spectral image classification has been widely used. Existing theories and methods still have certain limitations for more complicated hyper spectral image classification. Therefore, researching more targeted hyper spectral image classification methods will be an important research direction in the future.

ISSN: 2366-1313

7. REFERENCES

[1] M. Y. Teng, R. Mehrubeoglu, S. A. King, K. Cammarata, and J. Simons, "Investigation of epifauna coverage on seagrass blades using spatial and spectral analysis of hyperspectral images," in 2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), pp. 1–4, Gainesville, FL, USA, 2013.

[2] C. Ke, "Military object detection using multiple information extracted from hyperspectral imagery," in 2017 International Conference on Progress in Informatics and Computing (PIC), pp. 124–128, Nanjing, China, 2017.

[3] G. Notesco, E. Ben Dor, and A. Brook, "Mineral mapping of makhtesh ramon in israel using hyperspectral remote sensing day and night LWIR images," in 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), pp. 1–4, Lausanne, Switzerland, 2014.

[4] R. Pike, G. Lu, D. Wang, Z. G. Chen, and B. Fei, "A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging," IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 653–663, 2016.

[5] X. Zhang, A. Zhang, and X. Meng, "Automatic fusion of hyperspectral images and laser scans using feature points," Journal of Sensors, vol. 2015, Article ID 415361, 9 pages, 2015.

[6] H. Kwon, X. Hu, J. Theiler, A. Zare, and P. Gurram, "Algorithms for multispectral and hyperspectral image analysis," Journal of Electrical and Computer Engineering, vol. 2013, Article ID 908906, 2 pages, 2013.

[7] Y. Ma, R. Li, G. Yang, L. Sun, and J. Wang, "A research on the combination strategies of multiple features for hyperspectral remote sensing image classification," Journal of Sensors, vol. 2018, Article ID 7341973, 14 pages, 2018.

ISSN: 2366-1313

[8] H. Binol, "Ensemble learning based multiple kernel principal component analysis for dimensionality reduction and classification of hyperspectral imagery," Mathematical Problems in Engineering, vol. 2018, Article ID 9632569, 14 pages, 2018.

[9] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: an overview," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 9, pp. 6690–6709, 2019.

[10] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 6, pp. 1351–1362, 2005.

[11] G. Camps-Valls, T. V. Bandos Marsheva, and D. Zhou, "Semisupervised graph-based hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 10, pp. 3044–3054, 2007.

[12] G. Hughes, "On the mean accuracy of statistical pattern recognizers," IEEE Transactions on Information Theory, vol. 14, no. 1, pp. 55–63, 1968.

[13] X. Wang and Y. Feng, "New method based on support vector machine in classification for hyperspectral data," in 2008 International Symposium on Computational Intelligence and Design, pp. 76–80, Wuhan, China, 2008.

[14] A. Fu, X. Ma, and H. Wang, "Classification of hyperspectral image based on hybrid neural networks," in IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 2643–2646, Valencia, Spain, 2018.

[15] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 11, pp. 4085–4098, 2010.

[16] P. Deepa and K. Thilagavathi, "Feature extraction of hyperspectral image using principal component analysis and folded-principal component analysis," in 2015 2nd International Conference on Electronics and Communication Systems (ICECS), pp. 656–660, Coimbatore, India, 2015.

[17] M. Imani and H. Ghassemian, “Principal component discriminant analysis for feature extraction and classification of hyperspectral images,” in 2014 Iranian Conference on Intelligent Systems (ICIS), pp. 1–5, Bam, Iran, 2014.

[18] C. Jayaprakash, B. B. Damodaran, V. Sowmya, and K. P. Soman, “Dimensionality reduction of hyperspectral images for classification using randomized independent component analysis,” in 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 492–496, Noida, India, 2018.

[19] J. Wang and C.-I. Chang, “Independent component analysisbased dimensionality reduction with applications in hyperspectral image analysis,” IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 6, pp. 1586–1600, 2006.

[20] Q. Du, “Modified Fisher's linear discriminant analysis for hyperspectral imagery,” IEEE Geoscience and Remote Sensing Letters, vol. 4, no. 4, pp. 503–507, 2007.

[21] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” Science, vol. 290, no. 5500, pp. 2323–2326, 2000.

[22] J. B. Tenenbaum, V. D. Silva, and J. C. Langford,“A global geometric framework for nonlinear dimensionality reduction,” Science, vol. 290, no. 5500, pp. 2319–2323, 2000.

[23] J. D. Carroll, P. Arabie, and L. J. Hubert, “Multidimensional scaling (MDS),” Encyclopedia of Social Measurement, vol. 45, no. 2, pp. 779–784, 2005.

[24] J. Jost, Riemannian Geometry and Geometric Analysis, Universitext, Springer, 5th edition, 2008.

[25] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” Neural Computation, vol. 15, no. 6, pp. 1373–1396, 2003.

[26] Y. Yan, Y. Zhao, H.-f. Xue, X.-d. Kou, and Y. Liu, “Integration of spatial-spectral information for hyperspectral image classification,” in 2010 Second

ISSN: 2366-1313

IITA International Conference on Geoscience and Remote Sensing, pp. 242–245, Qingdao, 2010.

[27] X. Xu, J. Li, Y. Zhang, and S. Li, “A subpixel spatial-spectral feature mining for hyperspectral image classification,” in IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 8476–8479, Valencia, Spain, 2018.