A novel approach to segment and classify remote sensing pictures more accurately than random forest by employing FFNN

K.Venkata Vinay Kumar Reddy[1], John Justin Thangaraj.S[2]

[1]Research Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University Tamil Nadu, India, Pin:602 105.

[2]*Project Guide, Corresponding Author, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University Tamil Nadu, India, Pin:602 105.

ABSTRACT

Aim: Image classification includes analyzing and categorizing specific objects over the pictures. Our main purpose is to assess the performance of two types of classifiers used in image segmentation. The image of every location is randomly selected and changed into the bitmap formation to create the test type dataset consisting of 5 locations from the 400 different satellite images. Total 10 number of iterations computed with400 images which were collected from 5 various locations. The images were collected at different views. **Materials and Methods:** The classifiers are making use to segment the remote sensor images using the FFNN framework and Random Forest. In this research work, the ORL(Olivetti Research Laboratory) kind image database is used for analysis, and the implementation is processed with the assistance of Python programming. **Results and Discussion:** The outcome of novel FFNN compared with Random Forest, which has an accuracy of 97.15%, is compared with the previous classifier accuracy of 93.75%. The two algorithms are statistically satisfied with the independent sample t-Test two tailed is 0.008 (α = 0.001) value ($p<0.05$) with a confidence level of 95%. **Conclusion:** Based on the outcomes, it is finalized that the suggested novel FFNN framework makes better outcomes in the image segmentation system than the Random Forest classifier. The essential computational time takes place at a faster stage of 0.5 seconds.KNN classifier is used in the existing research work, with a mean value of accuracy rate of 93.75%. ResNet is recommended, which has a mean accuracy value of 97.15%.

Keywords: Image Segmentation, Remote Sensing, novel FFNN, Random ForestAlgorithm, Computational Time.

INTRODUCTION

Image classification is different from object prediction in that it works at the pixel state for determining the contours of the images over an object while doing satellite imagery. Those objects are like buildings, cars, trees, roads etc. (Gong et al. 2014). The applications of this method of Aerial imaging are widely spread from analyzing traffic to examining environmental differences occurring because of the global heat (Aber, Marzolff, and Ries 2010). Here in remote sensing, the calculation occurs in the region of the microwave, open to light and infrared rays of the electromagnetic spectrum. The images from remote sensing are utilized in the detection of coastal, geology, or formation of the classified maps, monitoring of ocean, agriculture cover lands etc. (Guo, Goodchild, and Annoni 2019). The spectral and spatial features of the resolution of the picture offer the deepest information over the final target. Those pictures are mostly utilized in the sectors of habitat management, a cover of

land mapping etc. The spatial information in such an imaging method is limited by the number of pixels that are in the particular image. The order is utilized to bundle pixels present in a picture into one of the land cover class (Thenkabail 2018).
In Last 5 years 2017-2021 the Google Scholar has published more than 196 articles and the IEEE published more than 200 articles about Remote sensing. Remote Sensing (RS) utilizes the energy of the electromagnet in the form of heat, the waves of radio, and exposure of light for calculating and predicting the features of the final goal. The aerial and the satellite fields are employed in remote sensing (Rustamov, Hasanova, and Zeynalova 2018). The arrangement procedure is used to investigate the advanced picture and concentrate data from that picture contingent on the application. The most common way of creating topical guides with subjects like vegetation types, land use, and geography is called picture order. In remote detecting different quality pictures are delivered relying on a sort of sensor. Order exactness relies upon picture quality (Green 2020).(Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020)
The disadvantages of Remote detecting satellite pictures are considered as perhaps the main datum hotspots for land use planning because of their broad geological inclusion at a proficient expense while giving indispensable data on the world's surface (Janssen and Bakker 2000). Land use maps are generally delivered in light of remote detecting picture arrangement draws near. Nonetheless, the precision and handling season of land use maps utilizing remote detecting pictures is as yet a test to the remote detecting local area. To comprehend a characterization of satellite pictures, the initial step is to perceive the items and afterward perceive the classification of the scene (Thomas 2018). The research gap is mainly due to the fact that remote sensing images are a fairly expensive method of analysis especially when measuring or analyzing smaller areas. Remote sensing requires a special kind of training to analyze the images. The aim of the study involves older cartographic techniques,mapping a satellite image is the simplest method, but the most significant effort is still analyzing specific targets in an image, such as trees or landmarks. Knowing about the newest news about the barriers, as well as information about them, is critical. And it's for this reason that deep learning models, which is the sub topic of machine learning, have become so popular.

MATERIALS AND METHODS

This research work was performed at Deep Learning Laboratory, Saveetha School of Engineering, SIMATS (Saveetha Institute of Medical and Technical Sciences). The proposed work contains two groups such as group 1 is taken as FFNN and group 2 as Random Forest. The FFNN algorithm and Random Forest algorithm were assessed several times with a sample size of 400. After collecting the dataset, unwanted contents were removed by pre-processing and data cleaning stages. Afterward, it opens the data sets, and the accuracy of both the FFNN algorithm and Random Forest structure is evaluated. The results were calculated using G* power software of version 3 and the minimum power of the analysis is fixed as 0.8 and

the maximum accepted error is fixed as 0.5 with a threshold value of 0.05% and the Confidence Interval is 95%. Computational time for this method is very less as it takes more pixels for calculation.

The ORL type database of satellite images is employed in this research task for analysis reasons. There are 10 iterations as shown in Table 2 and 400 images were collected from 5 various locations at different views. In the program layout, that program is a standard bitmap based format consisting of a 4 lines header, and data stored in the unsigned char type, providing a maximum of 256 gray scale levels or 8-bit data per pixel, the size of the image is set at 92 x 112 pixels (Yin and Ying 2020). This research process uses Python programming for segmenting satellite images. It is one of the most commonly used computer-based programming languages for image segmentation. Its outstanding library functions and tools aid to complete the image processing in a smooth way.

FFNN:

Feed-forward neural networks are used to learn the relationship between independent variables that act as network inputs and dependent variables that function as network outputs. Input variables are commonly observable variables like spectral intensities at specific wavelengths, and output variables are frequently category assignments like the presence or absence of organic functional groups iStep 5 : Now use model selection for importing, do molecules using infrared (IR) spectra. When a collection of 'training set' samples with known spectra and class labels are introduced to the network, the network weights are changed to minimize the disparities between the network outputs and the known 'actual' outputs, learning occurs. The network can be used to predict the class membership of unknown samples from their spectra after the weights have been changed using the samples in the training set. This chapter looks at feed-forward neural networks from the perspective of an applied scientist, with a focus on modeling data structures encountered in chemical research. A feed-forward neural network is a classification algorithm that is biologically inspired. It is made up of a number of basic neuron-like processing units that are arranged in layers, with each layer's units coupled to those in the previous layer. Each link may have a different strength or weight, therefore they are not equal. The weights on these connections represent a network's knowledge (Brownlee 2018).

The pseudocode for the novel FFNN approach :

Step 1 : Import the data from the required library.

Step 2 : Give directory of the dataset in the csv extension file.

Step 3 : Data as indexes with price prediction validation.

i . Price of the data.

ii . Data analysis as per the dataset.

iii . Plot the graph by using Matplot.

Step 4 : Import MLP models, Library Classification.

train and test split, use FFNN for the importing and Sequential models.

Step 6 : Give the sample size , the test size and train size then fit the train and test.

Step 7 : Then print the accuracy score and

end the program.

Random Forest algorithm:

Random Forest is a well-known machine learning algorithm that uses the supervised learning method. In machine learning, it can be utilized for both classification and regression issues. It is based on ensemble learning, which is a method of integrating several classifiers to solve a complex problem and increase the model's performance. "Random Forest is a classifier that contains a number of decision trees on various subsets of a given dataset and takes the average to enhance the predicted accuracy of that dataset," according to the name. Instead than relying on a single decision tree, the random forest collects the forecasts from each tree and predicts the final output based on the majority votes of predictions (VanderPlas 2016). The bigger the number of trees in the forest, the more accurate it is and the problem of overfitting is avoided. Random Forest application is that it is most commonly utilized in the following four sectors, This algorithm is mostly used in the banking industry to identify loan risk. This method can be used to identify disease trends as well as disease risks. Using this technique, one can find places with comparable land use. This algorithm can be used to identify marketing trends. The benefits of Random Forests are that it can handle both classification and regression problems. It can handle huge datasets with a lot of dimensionality. It improves the model's accuracy and eliminates the problem of overfitting.Random Forest's Disadvantages are despite the fact that random forest can be used for both classification and regression tasks, it is not better suited to regression.

The pseudocode for the Random forest classifier :

Step 1 : Import the data from the required library.

Step 2 : Give directory of the dataset in the csv extension file.

Step 3 : Data as indexes with price prediction validation.

i . Price of the data.

ii . Data analysis as per the dataset.

iii . Plot the graph by using Matplot.

Step 4 : Import MLP models, Library Classification.

Step 5 : Now use model selection for importing use train and test split, use Random Forest for the importing and Sequential models.

Step 6 : Give the sample size , the test size and train size then fit the train and test.

Step 7 : Then print the accuracy score and end the program.

Hardware setup references the details and system resource settings selected for specific devices, the following are minimum hardware specifications to develop this model. Processor: Intel i5, RAM 8GB, 1TB of HDD storage. The required software specifications for this model are Windows OS, version 10, Python programming language version 3 or above, IDE PyCharm, Jupyter.

Statistical Analysis

Statistical software IBM SPSS with the standard version 26.0 to find the SD (Standard Deviation), mean deviation, significance level and also plot the graphs, etc. The SPSS software was used in this research work for statistical analysis.

Group statistics and independent sample tests were executed on the experimental results and the graph was constructed for two graphs with two parameters under the concerned study. The independent variables are the pixel values and the dependent variables are image quality and embedding capacity in the study, for example, image steganography.

Training type dataset and test type datasets are assigned to the database (400 different images, photos of 5 locations). One image of every location is randomly selected and changed into the bitmap formation to create the test type dataset consisting of 5 locations from the 400 different satellite images. The training type dataset is created by removing the test type dataset images from the ORL type database, providing 400 image data as a whole. Entire images are converted to bitmap format once again.

RESULTS

Both the pseudocodes show how to perform the improved segmentation and classification of remote sensing images. For the given image datasets, the suggested algorithm FFNN framework will offer more precise results than Random Forest. The accuracy rate of novel FFNN architecture and the Random Forest classifier are described in Table 1. FFNN and Random Forest accuracy are 97.15 percent and 93.75 percent, respectively according to the data shown in Table 1. These are the datasets which we are going to use for the searching of the project title. Accuracy of novel FFNN and Random Forest are shown in Table 1. The FFNN algorithm is 1.55% more accurate than the Random Forest algorithm shown in Table 2. Independent sample T-test is performed for the two groups for significance and standard error determination. Two Tailed Significance value is 0.0001 ($p<0.01$) and it is statistically significant shown in Table 3. TheBar graph analysis of the novel FFNN algorithm and Random Forest algorithm is shown in Figure 1. Graphical representation shows the mean accuracy of 96.90% and 93.35% for the proposed algorithm FFNN and Random forest respectively. X-axis : FFNN vs Random Forest, Y-axis : Mean precision ± 1 SD.

DISCUSSION

Random Forest classifier is used in the existing research work, with a mean value of accuracy rate of 93.75 percent. FFNN is recommended, which has a mean accuracy value of 97.15 percent.CNN is one of the familiar models that produce a better ranking of attributes. Jonathan Long et al., 2014 say that CNN is trained based on pixel-level in semantic image segmentation. Their research work aims to design a fully CNN that accepts various inputs and generates equivalent size outcomes with effective learning and inference. Here the authors define and elaborate the importance of CNN, its application, and make the associations to the existing models (Gong et al. 2014). Fully connected CNN produces 20% associative improvement to 72.2% average IU value in 2012. The limitationof this article depends on the image information by the number of pixels that are present in the particular image (Oliva and Hinojosa 2020).

This research process provides the investigational study to assess the efficiency of the various ImageNet trained FFNN models with many network kinds later used as extracting features in the proposed model to resolve the image split-

up process. Computational time for this method is less as it takes more pixels for calculation. (Deng and Yu 2014). (Gong et al. 2014) uses SVM and Random Forest classifiers for classifying remote sensing images. This new system contains various processes like enhancing images, segmenting images, choosing training data, and classifying data (Dey, Ashour, and Borra 2017). MSVM (Multi-class SVM) and Random Forest models classify the images into water region 1, soil region 2, and vegetation region 3 (Sanz 2012). FFNN has the advantage of being able to train networks with a large number of layers (even thousands).

The disadvantage of the Random Forest is the accuracy value depends on the data quality and the predicting process may be very slow. Concurrently, the Random Forest also needs extra memory storage to keep the total training data and it is more costly (Damji et al. 2020). The future scope of this technique can be improved in the upcoming times by providing another classifier to increase the accuracy rate. The feature removal process can produce the use of a variety of newly created techniques. The accurate values of the extracted features are more, hence the effectiveness of the system improves (Jiang, Zhang, and Khosravi 2021).

CONCLUSION

When compared to Random Forest, the suggested FFNN system creates a better outcome with an accuracy value rate of 97.15%. The essential computational time takes place in a faster stage of 0.5 seconds.The segmentation and classification of remote sensing images makes each and every location to be captured.Based on the outcomes, it is finalized that the suggested novel FFNN framework makes better outcomes in the image segmentation system than the Random Forest classifier.

DECLARATION

Conflicts of Interests

No conflict of interest

Authors Contribution

Author KVVKR was involved in data collection, data analysis, manuscript writing. Author CMV was involved in the Action process, Data verification and validation, and Critical review of the manuscript.

Acknowledgement

The authors are grateful to Saveetha School of Engineering and Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for providing the required infrastructure for this research to be completed successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Sun Microsystems Technologies
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering.

REFERENCES

1. Aber, James S., Irene Marzolff, and Johannes Ries. 2010. *Small-Format Aerial Photography: Principles, Techniques and Geoscience Applications*. Elsevier.
2. Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. “A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural

Fibres." *Materials Today: Proceedings* 33 (January): 798–805.
3. Brownlee, Jason. 2018. *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. Machine Learning Mastery.
4. Damji, Jules S., Brooke Wenig, Tathagata Das, and Denny Lee. 2020. *Learning Spark*. O'Reilly Media.
5. Deng, Li, and Dong Yu. 2014. *Deep Learning: Methods and Applications*.
6. Dey, Nilanjan, Amira S. Ashour, and Surekha Borra. 2017. *Classification in BioApps: Automation of Decision Making*. Springer.
7. Gong, Shaogang, Marco Cristani, Shuicheng Yan, and Chen Change Loy. 2014. *Person Re-Identification*. Springer Science & Business Media.
8. Green, David R. 2020. *Unmanned Aerial Remote Sensing: UAS for Environmental Applications*. CRC Press.
9. Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.
10. Guo, Huadong, Michael F. Goodchild, and Alessandro Annoni. 2019. *Manual of Digital Earth*. Springer Nature.
11. Janssen, Lucas L. F., and Wim H. Bakker. 2000. *Principles of Remote Sensing: An Introductory Textbook*.
12. Jiang, Yizhang, Yu-Dong Zhang, and Mohammad Khosravi. 2021. *Advanced Deep-Transfer-Leveraged Studies on Brain-Computer Interfacing*. Frontiers Media SA.
13. Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.
14. Oliva, Diego, and Salvador Hinojosa. 2020. *Applications of Hybrid Metaheuristic Algorithms for Image Processing*. Springer Nature.
15. Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.
16. Rustamov, Rustam B., Sabina Hasanova, and Mahfuza H. Zeynalova. 2018. *Multi-Purposeful Application of Geospatial Data*. BoD – Books on Demand.
17. Sanz, Jorge L. C. 2012. *Advances in Machine Vision*. Springer Science & Business Media.
18. Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.
19. Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.
20. Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020.

"Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.

21. Thenkabail, Prasad. 2018. *Remote Sensing Handbook - Three Volume Set*. CRC Press.
22. Thomas, Ciza. 2018. *Ontology in Information Science*. BoD – Books on Demand.
23. VanderPlas, Jake. 2016. *Python Data Science Handbook: Essential Tools for Working with Data*. "O'Reilly Media, Inc."
24. Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.
25. Yin, Ruihan, and Shuyi Ying. 2020. "Partially Reversible Gray Image Data Hiding Based on Adjacent Pixel Difference." *Proceedings of the 2020 2nd International Conference on Information Technology and Computer Communications*. https://doi.org/10.1145/3417473.3417483.

TABLES AND FIGURES

Table 1: Accuracy of FFNN and Random Forest. The FFNN algorithm is 3.40% more accurate than the Random Forest algorithm. The essential computational time takes place in a faster stage of 0.5 seconds.

ITERATION NO.	FFNN (%)	Random Forest (%)
1	98.93	94.27
2	98.34	94.19
3	98.11	94.10
4	97.55	93.99
5	97.22	93.88
6	96.87	93.75
7	96.43	93.64
8	96.21	93.48
9	96.03	93.26
10	95.77	92.93

Table 2: Group Statistics of FFNN and Random Forest algorithm with the mean value of 97.15% and 93.75%.

GROUP	N	Mean(%)	Std.Deviation	Std.Error Mean
FFNN	10	97.1460	1.0703	0.33846
Random Forest	10	93.7490	0 .42927	0.13575

Table 3: Independent sample T-test is performed for the two groups for significance and standard error determination. Two Tailed Significance value is 0.0001 ($p<0.01$) and it is statistically significant.

		Levene's Test for Equality of Variance		T-test for Equality of Means						
	Equal Variances	F	Sig	t	df	Sig(2-tailed)	Mean Diffe--rence	Std. Error Differe nce	95% Confidence Interval of the Difference	
									Lower	Upper
Accuracy	Equal variances	8.919	0 .008							

	Assumed			9.315	18	0.0001	3.39700	.36467	2.63086	4.16314
	Equal variances Not Assumed			9.315	11.822	0.0001	3.39700	.36467	2.60113	4.19287

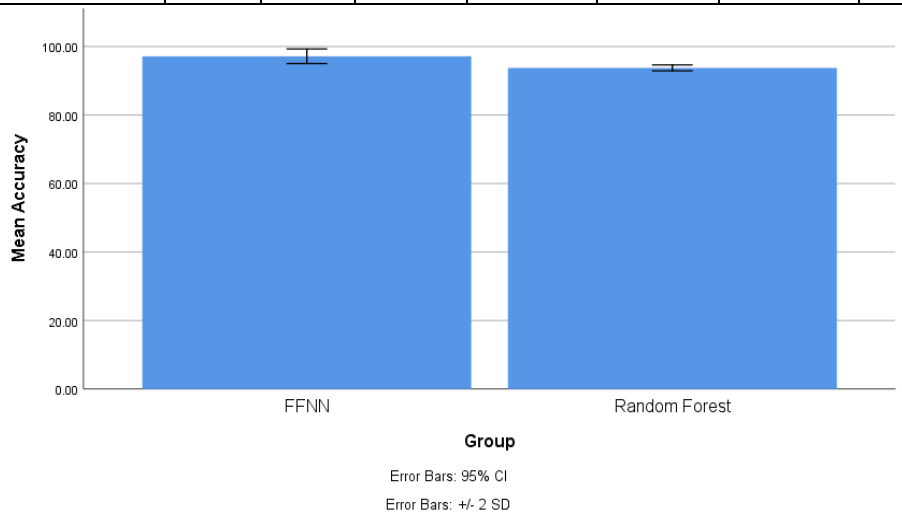

Fig 1. Bar graph analysis of FFNN algorithm and Random Forest algorithm. TheBar graph analysis of FFNN algorithm and Random Forest algorithm is shown in Fig 1. Graphical representation shows the mean accuracy of 97.15% and 93.75% for the proposed algorithm FFNN and Random Forest respectively. X-axis : FFNN vs Random Forest, Y-axis : Mean precision ± 2 SD.