MACHINE LEARNING FOR SPATIAL DATA

Mr.V.Balu[1], Mr.E.Pranay[2], Mr. J.Parthava Naidu[3]

[1]*Assisstant professor, Department of Computer Science and Engineering, SCSVMV, Kanchipuram*

[2] *B.E graduate(IV year), Department of Computer Science and Engineering, SCSVMV, Kanchipuram*

[3]*B.E graduate(IV year), Department of Computer Science and Engineering, SCSVMV, Kanchipuram*

Abstract - Nowadays, machine learning, artificial neural networks, and support vector machines are important tools for spatial and environmental data analysis and visualization. Machine learning for spatial data is used to predict and classify unknown locations according to known locations. Spatial data exist in different formats like vector data, raster data, geographic co-ordinate system. Usage of spatial data has been increased in recent times. Geographic information system (GIS) is the most common way to analyse the spatial data. Spatial data is used to track infectious diseases, climate change simulations, etc., There are many issues in spatial data handling like huge amounts of available data, missing data, duplicate data, classification of spatial images. This paper presents the review about the classification of images in a dataset using machine learning algorithm and finds the accuracy of prediction between support vector machine and Convolutional neural networks. The intersection of machine learning and spatial data involves four steps, including analysis and applicability assessment, extending the algorithm by embedding spatial properties, tuning parameters for good results and extending the algorithm in multiple ways.

From this study, we learned about the spatial properties and algorithms of machine learning with spatial data. Machine learning methods are more effective for spatial image classification and have wide applications these days.

Key Words: *Convolutional neural networks, Classification and Regression, Geographic information system, Machine learning, Prediction Accuracy, Support Vector Machine.*

INTRODUCTION

To gain a fundamental understanding of image classification, one should begin with conventional machine learning methods before switching to deep learning methods. To investigate the underlying concepts and methods of picture categorization, and to highlight the advantages that deep learning approaches have over more conventional machine learning methods. As previously noted, we started with the more conventional categorization techniques. After going through most of them, we decided to use SVM.

We believed that SVM would be a better starting point and we weren't mistaken, even if it made more sense to develop Neural Networks (NN), the more contemporary technique. We were suprised when we started to understand this strategy, which may not come as a surprise. Here, it should be noted that SVM is a very effective classification method on its own terms. But as was already stated, it is hopeless in a world overflowing with data. After more investigation, we discovered that breaking the topic down into three subtopics improved our comprehension. The sub-topics include Implementation and Analysis, Tuning Parameters, and Visualization. The approach produces an optimal hyperplane, which refers to the line after transformation/kernelling, given labelled training data (supervised learning).

Even though we created a classifier, it is not the most effective. The most effective classifier is determined by SVM using a technique known as the greatest gap trick optimization, with support vectors playing a vital role and all Other points are useless in determining the most effective classifier. After completing all the aforementioned steps, we were disappointed with the accuracy. After further research, we came up with the idea of deep learning, which improves signal transmission through multiple layers to neurons linked to the appropriate broad concepts by varying the strength of connections.

Each artificial neuron that fires when data is input into a network sends signals to specific neurons in the following layer, which are likely to fire if many signals are received. The development of software for spatial data has a rapid growth in the late 1980's by Haining. At first attention moved on conceptual issues. For example, integrating spatial measurable strategies and GIS climate. But today the situation is different because spatial data analysis software ware is readily available for different programs, customized scripts and GIS packages. R, java, python is used in spatial data software.

PROBLEM STATEMENT

There are many problems in spatial data handling some of them are huge amounts of available data, data quality, inconsistent data, missed data, spatial image classification. Due to irrelevant features in data, feature selection is also main problem. Machine learning algorithms used today to perform feature selection does not give the guaranteed reduction of huge spatial data. One of the main problems is spatial image classification that may be affected by many factors. In this paper we examined the current practices and prospects of image classification. Using machine learning algorithms, we have classified the images. We have used SVM and neural networks and calculated their accuracy.

EXISTING SYSTEM

Spatial entities and phenomena have object information like geometric, spectral, textural, and statistical that can be used as new features and added to the observation matrix. object information is added to spatial reference information contextual and relational. Geometric data for a specific geographic location. The length, area, and ratio of entities can all be stated in many ways. The remote sensing technology is highly dependent on spectral and textural data for land-cover and land-use classification. However, due to its heterogeneity, spectral features are insufficient, particularly in urban areas. Machine learning algorithms, especially Artificial Neural Networks of various architectures and Statistical Learning Theory, such as kernel-based methods, Support Vector Machines, and Support Vector Regression Some of the data-driven approaches are very dependent on the quality and quantity of data. As a result, multiple statistical/geostatistical methods can be used to control the quality of data analysis and modelling when utilizing using machine learning. Previous studies examined the foundational techniques for categorising photos. We have enough understanding after going through them to advance to more advanced methods. In this review study, two types of image categorization approaches are covered: supervised learning and unsupervised learning. Artificial Neural Networks, Support Vector Machines, and Directed Acyclic Graphs (SVM) are examples of supervised learning techniques. Fuzzy Decision Trees are examples of unsupervised learning approaches. In this article, we discuss three alternative classification methods, including There are three types of learning: unsupervised learning, in which labelled examples are not available for use as examples, and semi-supervised learning, in which the training data combines labelled and unlabelled data. Supervised learning requires the learner to study numerous input-output examples of the function in order to understand its behaviour.

PROPOSED SYSTEM

Rather than producing new spatial features and processing them with traditional non-spatial machine learning methodologies, we may directly use spatial properties in the learning algorithm. Support vector machines, decision trees, random forests, data mining. In recent times rough set theory is an efficient technique which has become an important tool for processing the spatial data. The earlier studies covered the fundamental methods for classifying images. After going through them, we had enough knowledge to move on to more sophisticated techniques. The image classification techniques discussed in this survey paper can be divided into supervised and unsupervised learning. Supervised techniques include Artificial Neural Networks, Support Vector Machines, and Directed Acyclic Graphs (SVM), while unsupervised learning techniques include Fuzzy Decision Trees. Here, we talk about three different classification techniques, including Three types of learning exist: supervised learning, in which the learner must study numerous input-output examples of the function in order to understand its behaviour; unsupervised learning, in which labelled examples are not available for use as examples; and semi-supervised learning, in which the training data combines labelled and unlabelled data. For classification and regression problem Support vectors machines are used. The SVM is used to map the original input space to a higher dimensionality feature space where the observations are separated by hyperplanes. To avoid overfitting Regularization terms are usually added to functions to control the complexity of the model. SVM performs well in high dimensional spaces. It is less sensitive to social class and more effective at simplification. A support vector random field extension of SVM that clearly captures spatial dependencies in classification.

ARCHITECTURE

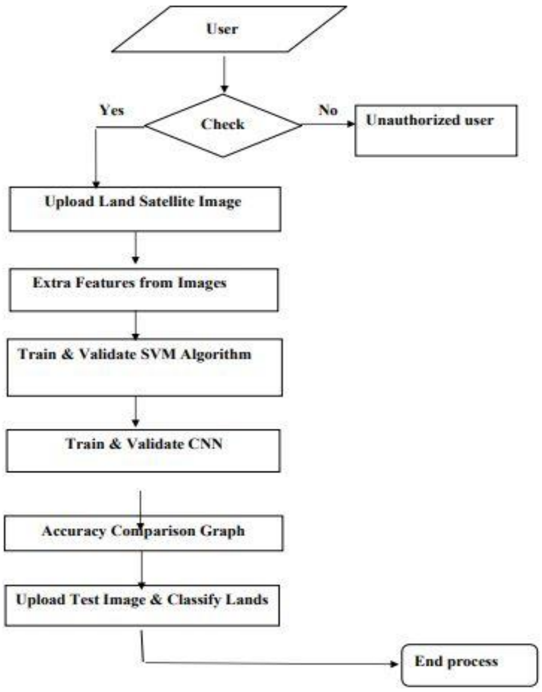

Figure1. Architecture

EXPERIMENTAL SETUP

This section describes the experiment that was carried out to classify the spatial image by support vector machine and Convolutional neural networks. To implement this project, we have used LAND satellite images which contains images of FOREST, AGRICULTURE LAND, URBAN AREA, and Range LAND. However, only a few studies have compared the performances of these classifiers with different training sample sizes for the same remote sensing images, particularly the Sentinel-2 Multispectral Imager (MSI). In this study, we examined and compared the performances of the CNN, and SVM classifiers for land use/cover classification using Sentinel-2 image data. An area of 30 × 30 km2 within the Red River Delta of Vietnam with six land use/cover types were classified using 14 different training sample sizes, including balanced and imbalanced, from 50 to over 1250 pixels/class.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

HARDWARE REQUIREMENTS:

- System : intel core i5
- Hard Disk : 10GB.
- Monitor : 15'' LED
- Input Devices : Keyboard, Mouse
- Ram :1 GB

SOFTWARE REQUIREMENTS:

- Operating system : Windows 10
- Coding Language : python
- Compiler : Python Idle(3.10)
- Libraries used : pandas, matplotlib.pyplot, keras, Tkinter Sklearn

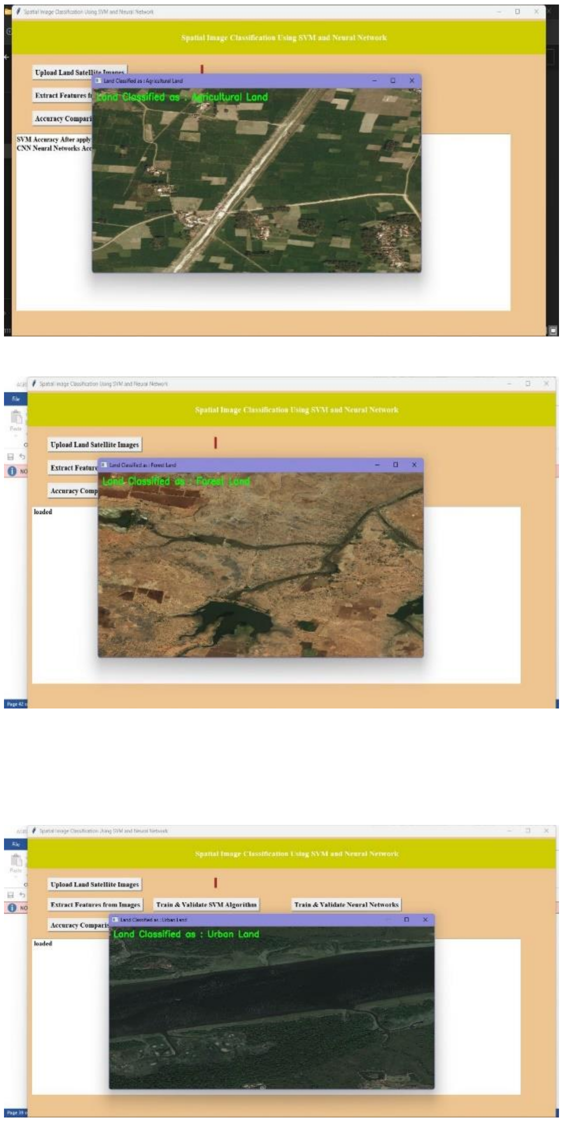

Figure 3. images classified as agricultural land, forest land and urban land.

RESULTS

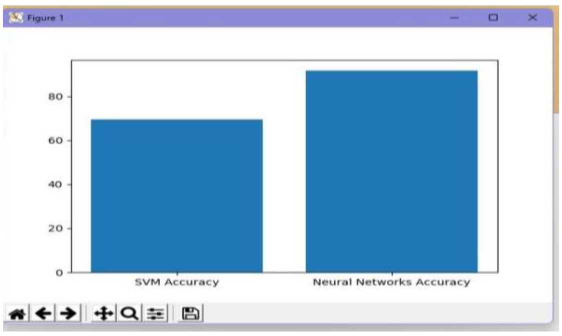

Figure 2.bar graph showing accuracy.

CONCLUSION

In conclusion, even though SVM is a highly powerful technique, attaining such a high accuracy is still unusual. We implemented SVM using a relatively short dataset to reach an accuracy of 93%. We eventually came to the conclusion that the reason our results had such a high accuracy was because the dataset wasn't big enough. Due to data augmentation, we were able to more than treble the amount of our dataset, and when we ran the SVM algorithm again, the accuracy was only 82%, which was significantly lower than before. We chose to switch to other deep learning methods because we weren't happy with the outcomes. Our investigation led us to CNN and neural networks. On the same dataset, we successfully implemented CNN and attained an astounding accuracy of 93.57%.We have learned about SVM algorithm and CNN. How images can be classified using the machine learning algorithms. Over the last few years, the research in image classification has gained momentum, showing signs of rapid growth. The latest to appear on the scene is bringing convenience in how classification can be done by employing various computational technologies. Using machine learning algorithms like SVM and neural networks we have classifies the images by extracting the features from images and calculated the accuracy. Classification results showed the high accuracy in neural networks ranging from 85% to 90% then followed by the SVM. Finally neural networks showed the higher accuracy when compared to SVM in image classification. Accuracy depends upon various factors like system speed.

REFERENCES:

[1] Luc Anselin, Ibnu Syabri, Youngihn Kho;" GeoDa: An Introduction to Spatial Data Analysis", Geographical Analysis ISSN 0016-7363, 2006

[2] Behnam Nikparvar and Jean-Claude Thill; "Machine learning of spatial data", preprints, August 6, 2021.

[3] Ibrahim Sabek, Mohamed F. Mokbel;" Machine Learning Meets Big Spatial Data".

[4] M. Kanevski, A. Pozdnoukhov, V. Timonin; "Machine Learning Algorithms for Geo Spatial data", Research gate 2016.

[5] Peijun Du, Xuyu Bai, Kun Tan, Zhaohui Xue, Alim Samat, Junshi Xia, Erzhu Li, Hongjun Su, Wei Liu; "Advances of Four Machine Learning Methods for Spatial Data Handling", Springer,2020.

[7] Francky Fouedjio· Jens Klump; "Exploring prediction uncertainty of spatial data in geostatistical and machine learning approaches", Springer, 2019

[8] Ibrahim Sabek, Mohamed F Mokbel; "Machine learning Meets Big spatial data", IEEE, 2020.

[6] Nicolas Gilardi and Samy Bengio; "Local Machine Learning Models for Spatial Data Analysis", JGIDA, 2000.

[9] Fateha Khanam Bappee, Amilcar Soares Junior and Stan MatWin; "Predicting Crime Using Spatial Features"12 Mar 2018.

[10] Hanna Meyera, Christoph Reudenbach, Stephan Wollauer, Thomas Nauss," Importance of spatial predictor variable selection in machine learning applications - Moving from data reproduction to spatial prediction", 21 Aug 2019.

[11] Pengfei Li, Hua Lu, Qian Zheng, Long Yang, Gang Pan;" LISA: A Learned Index Structure for Spatial Data", SIGMOD, June 14, 2020

[12] Trevor C. Bailey, Anthony C. Gatterl;" Interactive Spatial Data Analysis".

[13] Seongin Ahn, Dong-Woo Ryu and Sangho Lee;" A Machine Learning-Based Approach for Spatial Estimation Using the Spatial Features of Coordinate Information", 6 October 2020.