STM JOURNALS

International Journal of
Algorithms Design and Analysis Review

Volume 1, Issue 1, 2023
January–June
DOI (Journal): 10.37591/IJADAR

https://journals.stmjournals.com/ijadar

Review IJADAR

Machine Learning Techniques for Predicting Industries Based on Region

Supriya Kapase[1], Chaitanya Bari[2,*], Chaitrali Bhambure[2], Pallavi Chopade[2], Padmini Kondhalkar[2]

Abstract

In recent years, there has been a growing interest and emphasis on agricultural land preparation and its implementation among researchers, primarily due to various factors. These factors include an increased focus within the research community, a rising demand for agricultural land, and the significance of assessing soil health for ensuring robust crop production. Picture request is one such philosophy for soil and land prosperity examination. It is a staggering measure having the effects of various dimensions. This paper has suggested an investigation into the stream, analyzing the problems it addresses, as well as its future possibilities. The emphasis is focused on the intelligent examination of various advanced and successful gathering frameworks and methodology. Here, various components have been taking into account and these techniques have been directed to work on the accuracy of the portrayal. Suitable use of the features of remotely recognized data and picking the best sensible classifier are by and large huge for working on the accuracy of the data collected. The data-based game plan or non-parametric classifier like brain network has procured pervasiveness for multisource data gathering lately. Nevertheless, there is at this point the degree of extra investigation, to decrease weaknesses in the improvement of accuracy of the image gathering instruments. Support vector machine calculation is utilized to suggest the harvests in light of the soil conditions. Within this project, we strongly advocate for the adoption of the K nearest neighbor model by industries as well. We have created a skilled dataset for industries. We have worked on five regions, namely Konkan, Marathwada, Vidarbha, Nashik, and Pune.

Keywords: Convolutional neural network, support vector machine, K nearest neighborhood, crop prediction system

***Author for Correspondence**
Chaitanya Bari
E-mail: chaitanyabari.nbnssoe.it@gmail.com

[1]Professor, Department of Computer Engineering, Nbn Sinhgad Technical Institutes Campus, Pune, India
[2]Student, Department of Computer Engineering, Nbn Sinhgad Technical Institutes Campus, Pune, India

Received Date: June 07, 2023
Accepted Date: June 23, 2023
Published Date: July 29, 2023

Citation: Supriya Kapase, Chaitanya Bari, Chaitrali Bhambure, Pallavi Chopade, Padmini Kondhalkar. Machine Learning Techniques for Predicting Industries Based on Region. International Journal of Algorithms Design and Analysis Review. 2023; 1(1): 1–8p.

INTRODUCTION

To tackle the issue using data analysis, information mining suggests the identification of latent patterns within extensive datasets and the establishment of connections or correlations between them. Data mining applications in agriculture have benefited science. In order to lay the groundwork, representation is essential in all branches of science. It can assist in locating the variety of objects and ideas. It similarly gives fundamental information through which investigation can be made in a deliberate manner. Soil plays a vital role in agricultural fields as it is crucial for crop production. The composition of soil is influenced by various factors, including environmental conditions and practical considerations. On the earth's land surfaces, soil collection creates a link between soil testing and various types of trademark substances. Soil grouping has developed as an extremely important issue in image processing and computer vision. Numerous new calculations are being developed utilizing convolutional models to make the calculation as precise as could really be expected. These convolutional approaches have enabled the separation of pixel nuances. This examination means to plan a double face classifier which can extricate the highlights like edges, variety, and surface, independent of its arrangement. This examination presents a strategy to produce precise soil order from any erratic size input picture.

Convolutional Neural Network

Techniques are employed to recognize soil images, which involve various chemical features. Based on the geographical characteristics, support vector machine (SVM) models are utilized to suggest potential crops suitable for that particular soil series. It is an internet program that is very beneficial to farmers. Because of this, the farmer can also sell products online, avoiding travel to the markets, as happened during the COVID-19 pandemic.

Motivation

The main goal of the proposed research was to create a solid model that can categorize various types of soil data, suggest appropriate crops, and provide comprehensive information on selling the crops, all to the advantage of farmers who may make online purchases. The knowledge engineering approach has a flaw in that categorization rules must be updated on a regular basis, which is very difficult. However, over the past 20 years, the use of machine learning technologies has become more popular as a result of a number of variables, including the quantity of large datasets and the rising need for effective data processing methods.

The goals included the need to assist the agriculture sector and design a system that classifies soil and suggests crops with greatest precision and in the shortest processing time.

LITERATURE SURVEY

Vlachopoulos et al. [1] communicated in their study the outcomes of a field attempt coordinated for looking over the yield prosperity status of a couple of grain and oat crop fields in Prince Edward Island, Canada. The crop fields were mapped with an unmanned aircraft system (UAS), and the crop health status was assessed through the green area index (GAI) and vegetation indices (VIs). GAI maps were produced from the UAS imagery and VIs used machine learning pipelines with several regression algorithms (multiple linear models, support vector machines, random forests, and artificial neural networks) feature framework strategy. The random forests estimation was exhibited to be the best computation for GAI assumption with a normal relative root mean square error of 10.86% and a mean absolute error of 0.67. The resulting GAI maps and the regression feature space were described with random forests to discriminate between vigorous and stressed crop areas. The authors achieved a mean overall accuracy of 94%.

Rao et al. [2] created an optimal crop recommendation model that assists farmers in choosing the most suitable crop based on local meteorological factors and soil enhancements. To achieve this, the study employed two distinct evaluation metrics, Gini and entropy, to analyze popular algorithms such as K-nearest neighbor (KNN), decision tree, and random forest classifier.

The findings indicate that among the three options, random forest yields the highest precision, as demonstrated by Nejad et al. [3]. This review presents two proposed frameworks. The primary model incorporates Long short-term memory-Contemplations, skip connections, and 2D-CNN, while the subsequent model combines ConvLSTM Attention, skip connections, and 3D-CNN. MODIS data, including Land-Cover, Surface-Temperature, and MODIS-Land-Surface, spanning from 2003 to 2018, were collected for over 1800 regions at the area level, predominantly focused on soybean cultivation in the United States. The proposed procedures have been compared with the most recent models. Then, the results showed that the second proposed system significantly outperformed the other techniques. In the case of Mean Absolute Error, the second proposed approach, DeepYield, ConvLSTM, 3DCNN, and CNN-LSTM received scores of 4.3, 6.003, 6.05, 6.3, and 7.002, respectively. Usha Devi and Sheela Selvakumari [4] communicated that their study is portrayed as a proposition system that uses a couple of simulated intelligence ways to suggest appropriate crops considering data soil factors. The target of this investigation was to use a machine learning methods to forecast agricultural production. The classifier models consolidated data from logistic regression, naive Bayes, and random forest, with random forest area showing the best precision. Machine learning algorithms can help farmers with picking whatever plant to grow depending on these factors, including precipitation, temperature, and area. This approach thus reduces the financial losses incurred by farmers when they choose to plant some incorrect crops, as well as aiding farmers in their search for new types of crops suitable for cultivation in their area.

Hina et al. [5] communicated that in their suggested structure, they used a vast dataset that consolidated India's states, while in the old system, just a singular state was considered. These suggestions can be extracted and used to educate the farmers. The farmer can gain a better understanding of the harvests by employing a pictorial representation. Machine learning techniques develop a well-defined model with the data and helps us with accomplishing assumptions. Agrarian issues like crop prediction, crop rotation, water requirement, fertilizer requirement and protection can be settled.

Considering the dynamic climatic factors of the environment, it is necessary to employ an effective approach for managing crop yield growth and providing support to farmers and agricultural managers in their production endeavors. This could help agriculturalists with having an unrivaled cultivation. By utilizing data mining techniques, farmers can receive a comprehensive set of recommendations to improve their crop cultivation. This strategy considers both the specific characteristics of the crops and the prevailing climatic conditions. Data assessment provides a strategy for creating supportive extraction off country informational index. Crop dataset has been analyzed and prediction of harvests is done considering productivity and season.

PROPOSED METHOD AND ALGORITHM

Proposed Methodology

We propose conducting an experiment on soil color detection within the suggested system using a limited amount of supervised data [6].

The initial phase involves gathering soil test images of specific types, known as sensor selection, which is a crucial step in image-based soil classification. Factors such as user requirements, scale of analysis, characteristics of the soils being studied, availability of soil data, as well as time and cost limitations need to be considered during this process. Multiple soil test images are captured using a color camera and integrated into the framework. The system architecture of the proposed model is illustrated in Figure 1.

The utilization of a convolutional neural network (CNN) for image classification requires a substantial amount of training samples. When conducting fieldwork, samples are collected while taking into account a number of variables, including the spatial resolution of the images that were taken, the accessibility of ground reference data, and the intricacy of the data being studied. Careful consideration was given to selecting training samples, considering these factors. Following the successful

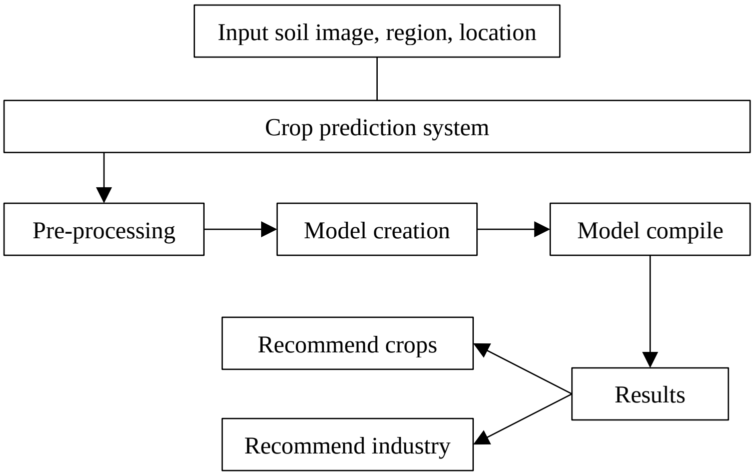

Figure 1. Proposed architecture.

classification of images, crop recommendations are made using the support vector machine (SVM) algorithm. Our contribution empowers farmers to sell their products online, eliminating the need to physically visit markets, especially during times of pandemic. The procedure comprises two stages: (1) training and (2) testing. Two datasets, specifically soil and crop datasets, are utilized. The soil dataset consists of images representing various soil classes such as alluvial soil, black soil, clay soil, and red soil. The crop dataset is employed for recommending suitable crops based on the corresponding soil class [7]. Additionally, in this paper, we also propose the adoption of the K-nearest neighbors (KNN) classifier for the benefit of the agricultural industry.

Dataset Gathering

We have collected the data from the Kaggle platform. There are a total of 440 images in the dataset. We split the dataset into two categories. We used 400 images for training and 40 images for testing.

Dataset Pre-processing

We scaled the image to 224 × 224 pixels size during the dataset pre-processing stage. We also removed some corrupt images from the folder.

Data Augmentation

By using data augmentation stage, we increased the size of training dataset as we zoom the image, changing the brightness and rotate the image [8].

Algorithms

CNN

The CNN model for advertisement images is illustrated in Figure 2, consisting of four blocks: (1) input, (2) capturing, (3) classification, and (4) output. The image comprises numerous components, with three key processing components: (1) settings and initialization, (2) CNN learning and visualization, and (3) sample test selection. The component “Settings and Initialise” has two inputs: (1) instance and (2) configurations. The instance input specifies the number of layers, layer names, number of filters in each convolution layer, classification technique, input image size, and kernel of the convolution layers. On the other hand, the configuration input defines model parameters such as learning rate, mini-batch size, weight decay, and momentum. The “CNN Learning and Visualization” component in the nLmF-CNN model represents image features across multiple layers, including input, convolutional (conv), rectified linear unit (relu), pooling (pool), fully connected (fc), and softmax layers as shown in Figure 2. It visualizes the current result of learning status. A convolution removes tiles of the information highlight guide, and applies channels to them to register new

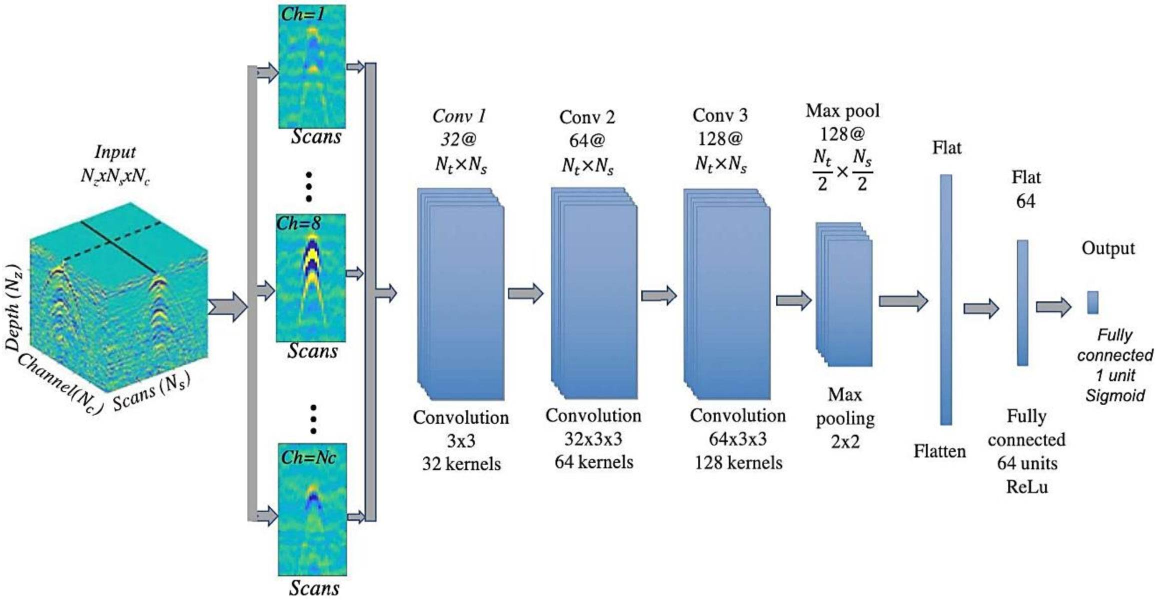

Figure 2. Convolutional neural network (CNN) architecture.

elements, delivering a result include map, or convolved highlight (which might have an alternate size and profundity than the info highlight map).

Two parameters are necessary for convolution operations: the size of the extracted tiles, which are commonly 3 × 3 or 5 × 5 pixels tiles and the stride value that determines the step size at which the tiles are moved over the input data.

The profundity of the result include map, which compares to the quantity of channels that are applied during a convolution, the channels (frameworks a similar size as the tile size) really slide over the information highlight guide's matrix evenly and in an upward direction, each pixel in turn, removing each relating tile [9]. Following every convolution activity, the CNN applies a corrected straight unit (ReLU) change to the convolved highlight, to bring nonlinearity into the model. The ReLU capability, F(x) = max(0,x), returns x for all upsides of x > 0, and returns 0 for all upsides of x ≤ 0. ReLU is utilized as an enactment capability in various brain organizations; After ReLU comes a pooling step, in which the CNN down examples the convolved highlight (to save money on handling time), diminishing the quantity of elements of the component map, while still safeguarding the most basic element data. A typical calculation utilized for this cycle is called max pooling. Max pooling works similar to convolution. We slide over the element guide and concentrate tiles of a predetermined size. For each tile, the greatest worth is result to another component guide, and any remaining qualities are disposed of. Max pooling tasks take two boundaries. Size of the maximum pooling channel (normally 2 × 2 pixels) Toward the finish of a convolutional brain network is at least one completely associated layer (when two layers are “completely associated,” each hub in the principal layer is associated with each hub in the subsequent layer). Their responsibility is to perform arrangement in light of the elements extricated by the convolutions. Commonly, the last completely associated layer contains a softmax enactment capability, which yields a likelihood esteem from 0 to 1 for every one of the characterization marks that the model is attempting to foresee.

Support Vector Machine

For the most part, the plan makes use of artificial intelligence (AI) calculations. The yields will be recommended by the help vector machine in this study. SVM is a regulated AI computation that operates on the concept of choice planes, which characterizes choice limitations. A choice limit distinguishes objects of one class from objects of another. The information focuses closest to the hyperplane are called support vectors [10]. Part capability is utilized to isolate non-direct information by changing contribution to a higher layered space. Gaussian spiral premise capability part is utilized in our proposed strategy as shown in Figure 3 [11–13].

KNN

An advancement of the K-nearest neighbors (K-NN) feature estimation is to evaluate the contribution of each K neighbor based on their distance to the query location, assigning a higher weight to the neighbors that are closer in proximity. The KNN classifier suggesting the industries to the farmer for crop selling subject to the nearest distance as shown in Figure 4.

$d=\sqrt{((x2 - x1)^2 + (y2 - y1)^2)}$, d = distance
x1, x2, y1, y2 = data points

RESULTS AND DISCUSSION

In our experimental setup, we evaluated a total of 400 trained photographs representing four distinct soil types, as shown in Table 1. Additionally, an additional set of 40 images was subjected to testing. These images were processed using our image processing module's feature extraction technique and passed through the CNN framework. Subsequently, our trained soil classification model accurately assigns specific color labels to the analyzed images. This paper suggests the adoption of the KNN model by industries.

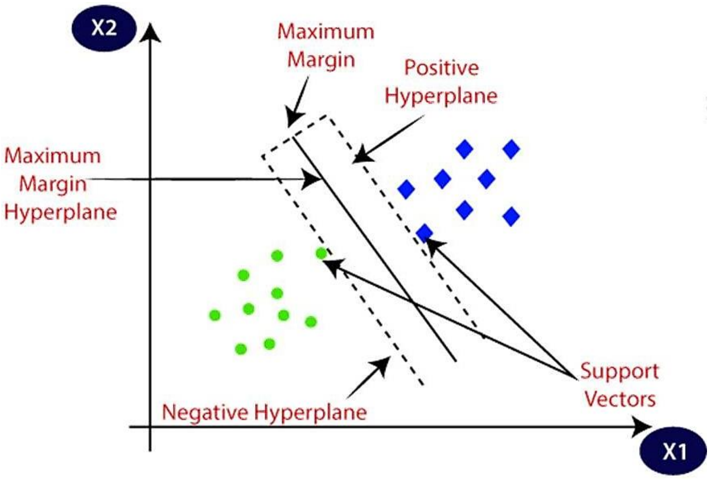

Figure 3. Support vector machine (SVM) architecture.

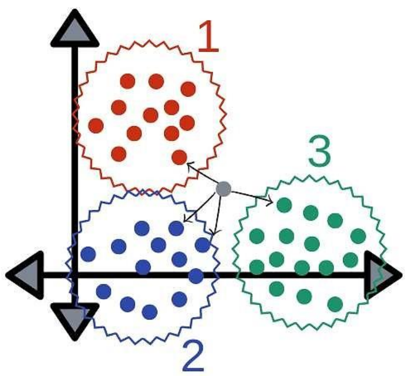

Figure 4. K nearest neighbors (KNN) architecture.

Table 1. Classification of data.

S.N.	Category	Number of images
1	Training images	400
2	Testing images	40

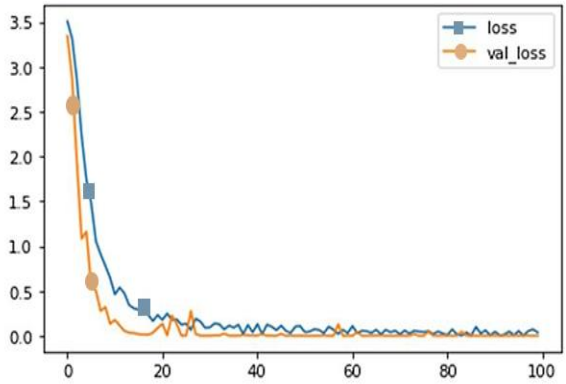

Figure 5. Loss of convolutional neural network (CNN).

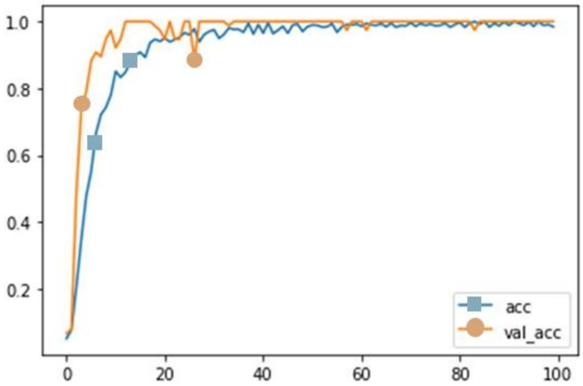

Figure 6. Accuracy of convolutional neural network (CNN).

In our experimental setup, shown in Table 1, the total numbers of images were 440. We obtained 95.45% accuracy for 100 epochs as shown in Figures 5 and 6.

CONCLUSION

This article presents a model that predicts soil conditions and provides recommendations for suitable crop yields based on the identified soil type. To accomplish this, we have developed an enhanced CNN model derived from a standard CNN model, constructed using ConvNet. The CNN model is employed for classifying soil images into four categories: (1) alluvial, (2) clay, (3) red, and (4) black soils. Using the SVM algorithm, we suggest specific crops for each soil class. Furthermore, we extend our recommendations to the agricultural industry to facilitate crop selling between industries and farmers. We are recommending the industries to farmer using KNN classifier. In future studies, we will try to improve the performance of model as well as increase the number of classes.

REFERENCES

1. Vlachopoulos O, Leblon B, Wang J, Haddadi A, LaRocque A, Patterson G. Evaluation of crop health status with UAS multispectral imagery. IEEE J Select Topics Appl Earth Observ Remote Sensing. 2022; 15: 297–308.
2. Rao MS, Singh A, Subba Reddy NV, Acharya DU. Crop prediction using machine learning. J Phys Conf Series. 2021; 2161: 012033. doi: 10.1088/1742-6596/2161/1/012033.
3. Nejad SMM, Abbasi-Moghadam D, Sharifi A, Farmonov N, Amankulova K, Lászlź M. Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches. IEEE J Select Topics Appl Earth Observ Remote Sensing. 2023; 16: 254–266. doi: 10.1109/JSTARS.2022.3223423.
4. Usha Devi R, Sheela Selvakumari NA. Crop prediction and mapping using soil features with different machine learning techniques. In: Proceedings of the International Conference on Innovative Computing and Communication (ICICC) 2022, February 19–120, New Delhi, India, 2022. doi: 10.2139/ssrn.4097213.
5. Hina F, Hasan MT. Agriculture crop yield prediction using machine learning. Int J Res Appl Sci Eng Technol. 2022; 10 (IV): 910–915.
6. Rajeswari SR, Khunteta P, Kumar S, Singh AR, Pandey V. Smart farming prediction using machine learning. Int J Innov Technol Exploring Eng. 2019; 8 (7): 190–194 .

7. Puno JCV, Bedruz RAR, Brillantes AKM, Vicerra RRP, Bandala AA, Dadios EP. Soil nutrient detection using genetic algorithm. In: 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Laoag, Philippines, November 29–December 1, 2019. pp. 1–6. doi: 10.1109/HNICEM48295.2019.9072689.
8. Reshma R, Sathiyavathi V, Sindhu T, Selvakumar K, SaiRamesh L. IoT based classification techniques for soil content analysis and crop yield prediction. In: Proceedings of the 2020 Fourth International Conference on I- SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), October 7–9, 2020. Palladam, India. pp. 156–160.
9. Jain S, Ramesh D. Machine learning convergence for weather based crop selection. In: 2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS), Bhopal, India, February 22–23, 2020. pp. 1–6. doi: 10.1109/SCEECS48394.2020.75.
10. Mariammal G, Suruliandi A, Raja SP, Poongothai E. Prediction of land suitability for crop cultivation based on soil and environmental characteristics using modified recursive feature elimination technique with various classifiers. IEEE Transac Comput Soc Syst. 2021; 8 (5): 1132–1142. doi: 10.1109/TCSS.2021.3074534.
11. Mahendra N, Vishwakarma D, Nischitha K, Ashwini, Manjuraju MR. Crop prediction using machine learning approaches. Int J Eng Res Technol. 2020; 9 (8): 23–26.
12. Vadalia D, Vaity M, Tawate K, Kapse D. Real time soil fertility analyzer and crop prediction. Int Res J Eng Technol. 2017; 4 ()3: 1497–1499.
13. Devi MPK, Anthiyur U, Shenbagavadivu MS. Enhanced crop yield prediction and soil data analysis using data mining. Int J Modern Computer Sci. 2016; 4(6).

