

# Machine Learning strategies for Yield Prediction using various soil parameters

Lakshmi CB<sup>1</sup>, Dr. Malarvizhi S<sup>2</sup>

Asst.Professor, St. Francis de Sales College, Bangalore<sup>1</sup>  
Asst.Professor, Tiruvallur Govt Arts College, Rasipuram<sup>2</sup>

[lakshmicb@sfscollege.in](mailto:lakshmicb@sfscollege.in)<sup>1</sup>  
[smalarvizzhi@gmail.com](mailto:smalarvizzhi@gmail.com)<sup>2</sup>

## Abstract

In agriculture, crop yield estimation is critical. Remote sensing is being used as in farming systems to increase yield efficiency and lower operating costs. Remote sensing-based strategies, on the other hand, necessitate extensive processing, necessitating the use of machine learning models for crop yield prediction. Descriptive analytics is a form of analytics that is used to accurately estimate crop yields. This paper discusses the most recent research on machine learning-based strategies for efficient crop yield prediction. In general, the training model's accuracy should be higher, and the error rate should be low. As a result, significant effort is being put forward to propose a machine learning technique that will provide high precision in crop yield prediction.

## Keywords:

Crop yield, Machine learning, Support Vector Machine, Regression Model, Naive Bayes, Agriculture.

## INTRODUCTION

Agriculture is a significant industrial sector, and the country's economy is dependent on its sustainability in rural areas, as well as factors such as climate change, rainfall, water levels, and pesticides. Despite the fact that recent research has revealed statistical knowledge about agriculture, few studies have looked into crop prediction based on historical evidence. Forecasting crop yields is critical food production. Crop yield estimation, on the other hand, is incredibly difficult due to a variety of complex factors. High-dimensional marker data, for example, is used to identify genotype details, which can include tens of thousands or even millions of makers per plant. In the field of crop prediction, artificial neural networks such as grouping, clustering, correlation, approximation, forecasting, and optimization are extensively used. Several models, such as main component regression, partial least squares, multiple linear regression, and neares,

---

are developed as primarily dependent methods for crop yield estimation during the background study of several research papers in artificial neural networks [1][2].

Agro-climatic input parameters affect crop production in general. However, agricultural input parameters differ from region to region, and gathering such data over a wider area is a difficult task. Climate data is also collected in various parts for each sq.m area. The resulting data sets are huge, and they can be used to forecast massive crops. Agricultural researchers are experimenting with different forecasting methodologies. Agricultural scientists have shown that the pro-pesticide state's policies have resulted in an alarmingly high use of pesticides. Also, there is a negative association between pesticide use and crop yield, according to the report. Crop yields would be greatly reduced if there is a prolonged dry spell or heavy rain during important crop growth stages [3].

Agriculture has spatial attributes that help determine patterns in agricultural production based on input availability. The K-means method is primarily used to predict emissions in the atmosphere; however, the k nearest neighbor method can also be used to simulate weather variables and shifts in weather scenarios. Soil characteristics are studied using data mining methods, and the k-means method is used to identify soils. Spatial data mining was developed specifically for decision-making and is also used in agriculture land grading. Droughts and floods can be predicted using climate inputs and a data mining technique called association laws. Data mining methods based on Spatio-Temporal data can be used to forecast irrigation water demand. In order to make the optimized data sets available for demand forecasting, machine learning algorithms in crop yield prediction must be investigated [4][5].

A comprehensive evaluation of the linear association between yield and these interactive variables, as well as precise datasets and computational methods to expose these relationships, are needed for accurate yield prediction. Environmental conditions have a significant impact on the production of agricultural products. Crop growth is affected by the weather, and yield varies greatly from season to season. Furthermore, soil materials connect with the atmosphere, resulting in spatial variability in development. Crop agronomic management (planting, fertilizer application, harvesting, tillage, and other practices) may be used to compensate for yield losses caused by weather [6][7][8][9]. As a result, yield forecasting is a valuable method for increasing crop yields and evaluating crop insurance contracts. The aim of this research is to come up with new machine learning methods for predicting crop yields.

The following is how the paper is structured: The background analysis is discussed in Section II. The proposed machine learning methods for crop prediction are discussed in Section III, and the study is concluded in Section IV.

## 2. RELATED STUDY

For crop yield, the statistical model used a weather-based yield forecasting model with multi linear regression. The monthly crop yield returns are used as a dependent variable in the stepwise MLR process, while agro-meteorological variables in the cumulative time-lapse of harvesting are used as independent variables. The primary goal of statistical modelling is to develop monthly yield statistical models for more than six months in advance by illustrating the perfect correlation between time-lag weather phenomena for a young mature crop's first six harvesting years. Furthermore, in the absence of weather measuring instruments, planters may use modeling with limited meteorological data as a fast strategy [10][11].

To forecast crop yields, the model uses Fuzzy logic (FL), Adaptive Neuro Fuzzy Inference System (ANFIS), and Multiple Linear Regression (MLR). Radiation, biomass, irrigation, and harvestable soil water are all used in this procedure (esw). Missing values, redundant data, outliers, and inaccurate data were removed from the dataset during pre-processing. Crop yield is predicted using the FL, ANFIS, and MLR. By adjusting the parameters of the ANFIS model, crop yields can be predicted. Back propagation and least mean square techniques are combined in ANFIS' hybrid training algorithm. Its main goal is to reduce the approximation error as much as possible. The number of epochs used to minimize the error will be used when training a fuzzy inference file (FIS). The greater the number of lower epochs, the greater the error. Epochs are used to change the weights in ANFIS to eliminate errors. According to the evaluation, the ANFIS model is used to extract complex interrelationships between information variables that are different from FL and MLR [12][13][14].

An artificial neural network is focused on the biological neural pathways of an animal brain. To predict the output, a neural network requires training; once trained, it can predict crop yields with patterns even if the prior input contains errors. Even if the input is complex, multivariate, and nonlinear, and the output is extracted later, neural networks are known to produce precise results. Speech recognition, computer vision, character recognition, signature recognition, and human face recognition are only a few of the applications for Artificial Neural Networks (ANN)[15][16].

Several linear regression techniques are used to forecast plant growth. A linear regression variation is multiple regression. It's a method for estimating the value of a variable dependent on the importance of two or more other variables. The answer variable is the one we're trying to figure out (or sometimes the variable result, target or criterion). Independent factors are the variables we use to estimate the value of the dependent variable (or sometimes the variables of predictors, explanations or regressors). Using multiple regression to see whether exam performance will be measured using variables such as revision time, test anxiety, lecture attendance, and gender. Multiple regression can be used to determine how well a model matches the data (variance), as well as how much each predictor refers to the achievement variance [17][18].

Nodes, branches, terminal values, policy, payoff distribution, some alternatives, and rollback techniques are all used in decision tree models. There are three different types of nodes and two different types of branches to choose from. A decision node, seen as a square, is a point at which a decision must be taken. The starting point for decision branches is a choice node. A terminal value, also known as a payoff value, outcome value, or endpoint value, is assigned to each terminal node. Each terminal value denotes the outcome of a situation or a set of choices. Building a decision tree algorithm involves two steps: first, large decision tree development, followed by size reduction and overfitting the data, and finally, tree pruning. The pruned decision tree is the name given to the classification tree that has been identified. Multiple variables have an effect on crop forecasting. To build yield mapping and prediction yield, agronomic variables, nitrogen application, and weed control were used. They come to the conclusion that using ANNs results in higher forecast accuracies [19][20][21].

A Bayesian network or probabilistic guided acyclic graphical models used for crop prediction are examples of statistical models. The variance of recent weather parameters such as temperature, radiation, rainfall, and crop growth information, as well as the uncertainty of impacts of climate change, are all factors considered in this network of beliefs. It's thought that the network would support agricultural policymakers. The model is validated using synthetic weather conditions, and its performance is higher than that of traditional crop prediction models [22][23].

From the standpoint of food security, high accuracy crop yield estimation is important. Data is typically prepared uniformly, and crop model implementation is typically difficult in all regions. Deep learning can extract important features from input data to estimate the object. As a result, the reliance on input data can be minimized. The best algorithm was a network model with two Inner Product Layers that achieved a root mean square error of 6.298, which is a normal value. The research also emphasizes the benefits of deep learning for estimating crop yields [24][25][26].

Rice, wheat, and various pulses are among the foods produced. Crop-growing areas' long-term viability is contingent on favorable climatic conditions. Climate change has the potential to reduce demand. So, in order to make better crop-choice decisions, better techniques for predicting crop productivity in various climatic conditions are needed. Crop yield prediction under climatic instability is improved using machine learning techniques. An analysis of machine learning techniques is collected, as well as experimental findings. Precipitation, temperature, and reference crop evapotranspiration are the parameters used in the dataset. Mean absolute error, root mean squared error, relative absolute error, and root relative squared error are some of the parameters that are measured. Machine learning techniques outperformed conventional methods in the simulations [27].

Agricultural surveillance will help to avoid starvation and aid humanitarian efforts. Predicting yield estimation before harvest is the most difficult task. The use of remote sensing to predict crop yields is implemented as a scalable, reliable, and low-cost tool. The remote sensing group has

Site description	<ul style="list-style-type: none"> <li>• Latitude and longitude, elevation, average annual temperature</li> <li>• Slope and aspects of the site</li> </ul>
Weather	<ul style="list-style-type: none"> <li>• Daily global soil radiation, daily maximum and minimum temperature, daily rainfall</li> </ul>
Soil	<ul style="list-style-type: none"> <li>• Soil type, soil depth (divided by n layers), soil texture, soil organic carbon, bulk density, soil nitrogen, pH</li> </ul>
Initial condition of the system	<ul style="list-style-type: none"> <li>• Previous crop, residues left on the soil (if any), initial soil water and soil nitrogen</li> </ul>
Crop and field management	<ul style="list-style-type: none"> <li>• Cultivar name and type, planting date and type, row space, plants per square meter</li> <li>• Irrigation/nitrogen amount, method, dates of irrigation/fertilization, fertilizer type</li> </ul>

## Output Variable

- Crop Yield

Machine Learning (ML) is a technique for resolving problems where the relationship between input and output variables is uncertain or difficult to obtain. Learning refers to the automated acquisition of structural data. Figure 1 depicts a block diagram of crop yield prediction using machine learning techniques.

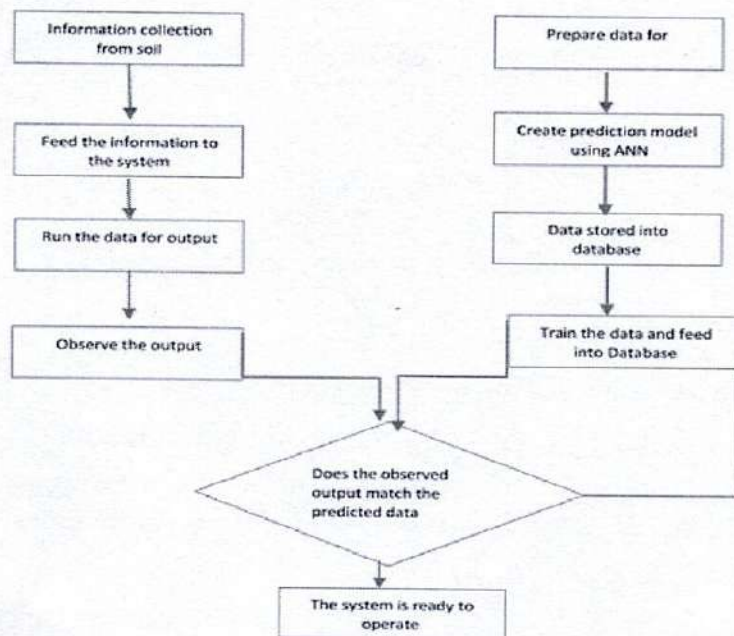


Figure 1. Proposed Crop Yield Prediction

historically used hand-crafted features. However, a creative dimension reduction technique is used to train and learn important features from labelled training data. Eventually, a Gaussian Process variable is being used to increase accuracy. The method is tested in county-level yield prediction and shown to outperform competing methods [28].

Various parameter of soil has a major contribution in prediction of yield i.e., based on CWD (crop water demand) parameter is used to avoid reductions in yield [29].

Climate has a major influence on crop yields. To forecast climate change impacts on the industry, a machine learning approach to yield modelling is presented. The model employs a semi-parametric version of a deep neural network. Complex nonlinear interactions, known parametric structure, and unobserved cross-sectional heterogeneity are all represented in high-dimensional datasets. The method outperforms all classical statistical approaches and entirely nonparametric neural networks of years withheld during model training in terms of yield prediction, according to yield results. Using scenarios from variable climate models, the most extreme negative effects of climate change on crop yield can be observed, but they are less drastic than impacts predicted using traditional statistical approaches. In particular, in the hottest regions, the new strategy is less negative.

### 3. SYSTEM MODEL

Crop yield is predicted using a machine learning method. Predictive analytics can be used to predict an event's likely future outcome. In prediction, proposed prediction models take into account historical data about the object. The climate change input parameters that are taken into account when making predictions.

#### Input Variables

##### Weather

- Precipitation (Daily, Weekly, Monthly, Crop Growth Phases)
- Minimum and Maximum Temperatures
- Solar Radiation
- Evapotranspiration

##### Non weather

- Seed variety, soil type, fertilizer, pesticides, farm equipment, and labour are all factors to consider.

**Table 1. Input for Crop Simulation Model**

Factor	Description
--------	-------------

- Sensitivity

Analyses of sensitivity can help to examine the main sources of uncertainty of model prediction. This helps identify important control points, prioritize additional data collection or analysis that is required, and select parameters for calibration.

- Specificity

Specificity represents crop-specific biological attributes for respiration and photosynthesis, essential stages and cycles of growth which characterize periods of vegetative and grain loading and canopy structure.

- Positive Predictive Value (Precision)

Precision is calculated as the number of true positive predictions divided by the total number of positive predictions.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Where

TP- True Positive

FP - False Positive

- F1 Score

The accuracy of the measurement test is known as the weighted harmonic mean of the tests ' accuracy and recall.

$$\text{F1 Score} = 2 * \left( \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \right)$$

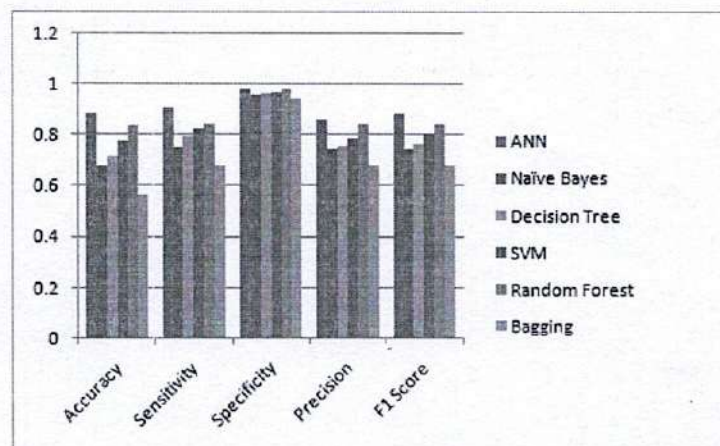


Figure 2. Comparison Analysis- ANN Vs Conventional methods

## Algorithm

Step 1: Import the set of data

Step 2: The variables in the data set are converted into a specific range of values. This ensures that the data collection is reliable, preventing anomalies. To standardize data, remove any missing values and use normalization. Once the data collection is standardized, redundancy is reduced, and data processing is simplified.

Step 3: Then the continuous set of data is discrete.

Step 4: The classification methodology is applicable to the preprocessed data collection. ANN is the classification technique used.

The artificial neural network is a non-linear model that performs better when the input parameters have a non-linear relationship, while machine learning is a linear model that performs better when the input parameters have a linear relationship.

## Performance Metrics

The true efficacy passes through the treatment with it – but only after it requires precision, and is recognised as such. The final results indicate that using the bagging procedure as an optimization technique within the category process could improve classification efficiency. The following performance metrics are used to assess crop prediction performance.

- Accuracy

Historically, the accuracy rate has become the most widely used statistical metric. However, in the case of unbalanced data sets, reliability is no longer an appropriate metric because it does not distinguish between the number in correctly classified instances of different groups. As a result, it can lead to incorrect conclusions; for example, a classifier that achieves 90% accuracy in a data set with an IR value of 9 is not accurate if it classifies all examples as negative.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$$

Where

TP- True Positive

TN- True Negative

FP - False Positive

FN - False Negative



Figure 2 shows how the ensemble approach compares to current KNN, Naive bayes, Decision tree, SVM, and Random forest methods in terms of output metrics. With the ANN method, all output parameters show a fair improvement and improved prediction accuracy.

## 5. CONCLUSION

Crop yield estimation is an active field of study and experimentation in agriculture. In agriculture, crop yield estimation is critical. Remote sensing is being used as decision support in farming systems to increase yield efficiency and lower operating costs. Since remote sensing-based strategies necessitate extensive processing, machine learning methods for crop yield prediction have become popular. However, accurate yield estimation for agricultural planning is a critical problem. The machine learning method is a strategy for finding realistic solutions to this problem. There are numerous machine learning methods for yield prediction that have been implemented in recent years. In this study, ANN-based crop prediction is proposed that tends to boost crop yield estimation and to investigate in the agricultural environment that machine learning algorithms need to focus on and resolve.

## REFERENCES

1. K. Aditya Shastry, H.A. Sanjay and Abhijeeth Deshmukh 2016, "A Parameter Based Customized Artificial Neural Network Model for Crop Yield Prediction", *Journal of Artificial Intelligence*, Volume 9 (1-3): 23-32.
2. Agus Bueno 2012, "An implementation of fuzzy inference system for onset prediction based on Southern Oscillation Index for increasing the resilience of rice production against climate variability", *Advanced Computer Science and Information Systems (ICACSIS)*, 281-286.
3. Andrew Crane-Droesch 2018, "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture", *Environ. Res. Lett.* 13.
4. L. Breiman 2001, "Random forests," *Mach. Learn.* 45 (1) 5-32.
5. Coates and A.Y. Ng 2012, "Learning feature representations with K-Means, in: G. Montavon, G.B. Orr, K.R. Müller (Eds.), *Neural Networks: Tricks of the Trade*," *Lecture Notes in Computer Science*, vol. 7700, Springer.
6. David B. Lobell and Marshall B. Burke 2010, "On the use of statistical models to predict crop yield responses to climate change", *Agricultural and Forest Meteorology*, Volume 150, Issue 11.
7. Daniel F. Polan, et al. 2016, "Tissue segmentation of Computed Tomography images using a Random Forest algorithm: a feasibility study," *Phys Med Biol.* 61(17): 6553-6569.
8. Eswari and Vinitha 2018, "Crop Yield Prediction in Tamil Nadu using Bayesian Network", *International Journal of Intellectual Advancements and Research in Engineering Computations*, Volume-6 Issue-2.
9. Gandhi, Niketa & Petkar, Owaiz & Armstrong, Leisa & Tripathy, Amiya. (2016). Rice crop yield prediction in India using support vector machines. 1-5. 10.1109/JCSSE.2016.7748856.
10. GONZALEZ-SANCHEZ, Alberto, FRAUSTO-SOLIS, Juan, OJEDA-BUSTAMATE, Waldo 2014, "Predictive ability of machine learning methods for massive crop yield prediction", *Spanish Journal of Agricultural Research*, [S.l.], v. 12, n. 2, p. 313-328, apr, ISSN 2171-9292.
11. H.V. Jagadish, et al. 2005, "I distance: an adaptive b+-tree based indexing method for nearest neighbor search," *ACM Trans. Database Syst. (TODS)* 30 (2) 364-397.
12. Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, Stefano Ermon 2017, "Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data", *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*.

13. Kanungo, T., et al. 2002, "An efficient k-means clustering algorithm: analysis and implementation." *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 881–892.
14. Khaki, S and Wang, L 2019, "Crop Yield Prediction Using Deep Neural Networks", arXiv:1902.02860.
15. Kim, N and Lee, Y.W 2016, "Machine learning approaches to corn yield estimation using satellite images and climate data: A case of Iowa State", *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* 34, 383–390.
16. K. Kuwata and R. Shibasaki 2015, "Estimating crop yields with deep learning and remotely sensed data," 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, pp. 858-861. doi: 10.1109/IGARSS.2015.7325900
17. Nari Kim, Kyung-Ja Ha, No-Wook Park, Jaecil Cho, Sungwook Hong and Yang-Won Lee 2019, "A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction: Case Study of the Midwestern United States, 2006–2015", *ISPRS Int. J. Geo-Inf.* 2019, 8, 240; doi:10.3390/ijgi8050240.
18. Samad Emamgholizadeh, M. Parsaeian, Mehdi Baradaran 2015, "Seed yield prediction of sesame using artificial neural network", *European Journal of Agronomy*, 68, 89-96
19. Sakshi Indolia, et al. 2018, "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," Pooja Asopa, *International Conference on Computational Intelligence and Data Science (ICCIDIS 2018)*, *Procedia Computer Science* 132, 679–688.
20. Su, Y. et al. 2017, "Support vector machine-based open crop model (SBOCM): Case of rice production in China. Saudi J," *Biol. Sci.* 24, 537–547.
21. S.Suthaharan 2016, "Decision tree learning, in *Machine Learning Models and Algorithms for Big Data Classification.*" Springer, pp. 237–269.
22. Veenadhari, Bharat Mishra and CD Singh 2011, "Soybean Productivity Modelling using Decision Tree Algorithms ", *International Journal of Computer Applications (0975 – 8887) Volume 27– No.7.*
23. SML Venkata Narasimhamurthy and AVS Pavan Kumar 2017, "Rice Crop Yield Forecasting Using Random Forest Algorithm", *International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98.*
24. Wang, A.X., Tran, C., Desai, N., Lobell, D and Ermon, S 2018, "Deep transfer learning for crop yield prediction with remote sensing data", *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, Menlo Park/San Jose, CA, USA, 20–22 June.*
25. Yang, Q, Shi, L, Han, J, Zha, Y and Zhu, P 2019, "Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images", *Field Crops Res.* 235, 142–153.
26. Zafer CÖMERT and Adnan Fatih KOCAMAZ 2017, "A study of artificial neural network training algorithms for classification of cardiocography signals," *Journal of Science and Technology* 7(2) 93–103.
27. Zingade, Omkar Buchade, Nilesh Mehta, Shubham Ghodekar and Chandan Mehta 2017, "Crop Prediction System using Machine Learning", *International Journal of Advance Engineering and Research Development Special Issue on Recent Trends in Data Engineering Volume 4, Special Issue 5.*
28. Zoppis, Italo, et al. 2018, "Methods, Kernel, Kernel Methods: Support Vector Machines, Reference Module in Life Sciences," Elsevier.
29. Lakshmi CB, Dr.Malarvizhi.S, "SOIL MOISTURE CONTENT PREDICTION USING MACHINE LEARNING: A REVIEW", *IJRAR - International Journal of Research and Analytical Reviews (IJRAR)*, E-ISSN 2348-1269, P- ISSN 2349-5138, Volume.6, Issue 2, Page No pp.117-125, June 2019, <http://www.ijrar.org/IJRAR19K4693.pdf>

## CERTIFICATE OF PUBLICATION

ISSN: 0031-4773

This is to certify that the paper entitled

*"Design of Ann Based Machine Learning Method for Crop Yield Prediction"*

Authored By

**Lakshmi CB**

From

**Asst.Professor, St. Francis de Sales College, Bangalore**

Has been published in

**PENSEE JOURNAL, VOLUME 51 ISSUE 05 - 2021**

PAPER ID: PNS-0521-68

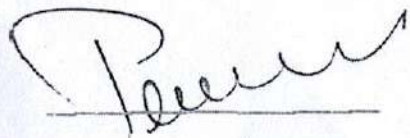
**MIAR**



Scopus®



ज्ञान-विज्ञान विमुक्तये



Editor-In-Chief

<https://penseereseach.com>

