JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Soil Quality Classification using Machine Learning for Precision Agriculture

Dr.Ayesha Banu[1] , Kaitha Rajitha[2], Pastam Srinivas [3], Komakula swathi[4], Bathini Shashikiran[5]

[1]Associate Professor, Department CSE(Data Science), Vaagdevi College of Engineering, Bollikunta, Warangal. ayeshabanuvce@gmail.com

[2]Assistant Professor, Department CSE(Data Science), Vaagdevi College of Engineering, Bollikunta, Warangal. rajithakaitha1@gmail.com

[3]Assistant Professor, Department CSE(Data Science), Vaagdevi College of Engineering, Bollikunta, Warangal. pastamsrinivas@gmail.com

[4] Assistant Professor, Department CSE(Data Science), Vaagdevi College of Engineering, Bollikunta, Warangal. swathijanu89@gmail.com

[5]Assistant Professor, Department CSE(Data Science), Vaagdevi College of Engineering, Bollikunta, Warangal. b.shashikiran6@gmail.com

ABSTRACT

Precision agriculture is focused on the detailed segmentation of the soil types in order to ensure better productivity and increased efficiency. Current manual practices in the field are not only tedious but also require proper expertise in the matter, making them both time and cost-consuming. Furthermore, the possibility of error during the process is not ruled out, thus complicating the decision-making process. This paper proposed an automated solution on the basis of advanced algorithms and artificial intelligence to address the issues introduced by conventional methods and provide a reliable, accurate, and quick alternative that is built upon the principles of convolutional networks and traditional machine learning approaches. The suggested methodology utilizes the process of image processing to convert soil images into numerical data necessary for analysis and subsequent segmentation. The collected data is utilized to train machine learning and deep learning models, including the Extra Trees Classifier and Convolutional Neural Network (CNN). To facilitate easy and intuitive user interaction, the system is implemented through a graphical user interface (GUI) developed using Tkinter, allowing individuals without technical expertise to operate the application efficiently. The system also incorporates visualization features, such as confusion matrices and training history graphs, which enable effective assessment and monitoring of model performance and accuracy. By automating the process of soil classification, the proposed system seeks to enhance conventional soil assessment practices by minimizing dependence on limited expert resources while providing rapid and precise results. It enables reliable identification of soil types within a shorter time, supporting better resource utilization and facilitating informed decision-making in agricultural activities. Overall, the proposed AI-based solution has considerable potential to improve the efficiency and effectiveness of precision agriculture, thereby promoting sustainable farming practices and contributing to enhanced crop productivity.

Keywords: Precision agriculture ,Soil type classification, Machine Learning, Predictive analytics, Deep CNN.

I. INTRODUCTION

Soil classification is the process of grouping soils that exhibit similar engineering characteristics based on a defined set of physical and mechanical properties. Unlike conventional engineering materials such as steel and concrete, soil exhibits complex behavior because of its natural origin, geological formation, and particulate structure. The evaluation of soil properties and the understanding of its behavior constitute fundamental aspects of geotechnical engineering. Soil classification is generally based on key properties such as particle-size distribution, liquid limit, and plasticity index. Classifying soils according to their common characteristics helps engineers predict their behavior under different loading

conditions and supports appropriate engineering applications. In particular, accurate soil classification is an essential preliminary step in foundation design, as the engineering performance of a foundation is strongly influenced by the properties of the underlying soil. Therefore, conducting a detailed soil investigation is necessary to determine the relevant soil characteristics during the research, analysis, and design stages of geotechnical engineering projects.

From an engineering perspective, reliable soil classification requires systematic analysis of soil properties. Sieve analysis is one of the primary methods used to determine the distribution of soil particle sizes. In this procedure, the soil sample is passed through a series of sieves with different opening sizes, and the quantity of material retained on each sieve is measured. The resulting particle-size distribution provides important information for classifying the soil. Another important parameter in soil classification is soil plasticity. The consistency and moisture-dependent behavior of fine-grained soils are commonly described using the liquid limit (LL), plastic limit (PL), and shrinkage limit (SL). As the moisture content decreases, a soil sample gradually changes from a liquid state to a plastic state and subsequently to semisolid and solid states, as illustrated in Figure 1. Among these parameters, the liquid limit and plasticity index are particularly significant for identifying and classifying fine-grained soils.

The liquid limit represents the moisture content at which soil transitions from a plastic state to a liquid state, whereas the plastic limit corresponds to the moisture content at which the soil changes from a semisolid state to a plastic state. The liquid limit can be experimentally determined using methods such as the Casagrande cup and fall cone techniques. The plastic limit is generally determined by repeatedly rolling a portion of soil by hand on a smooth glass plate until the soil reaches the specified consistency and begins to crumble. The difference between the liquid limit and plastic limit is known as the plasticity index (PI), which provides an indication of the range of moisture content over which the soil behaves plastically. Consequently, the liquid limit and plasticity index values can be plotted on the Casagrande plasticity chart to assist in the identification and classification of different soil types.

II. AI & ML Algorithms for Soil Classification: Survey

Artificial intelligence (AI) technologies have been used to predict the behaviour of nonlinear systems and have helped in adjusting various factors in order to optimize the conditions for system operation. A recent study highlights the increasing role of artificial intelligence as a means of improving agricultural productivity. Sharma et al. [1] suggested that solar-powered IoT sensor nodes should be used to monitor and control the agricultural sector, the tasks involved in crop management, harvesting, the regulation of water supply, animal control, the distribution of pesticides, as well as the measurement of moisture and temperature being among those that will also be monitored and controlled in agriculture. Suchithra [2] suggested that sensors can be used to monitor various field conditions, including temperature, humidity, and soil fertility. The data obtained from the sensors is authenticated and then transmitted to Wi-Fi, after which the verified information is sent via the cloud to the farmer's mobile phone or laptop. When action is needed on the field, the farmers are informed by means of an SMS message. An algorithm is created to set thresholds for temperature, humidity, and fertility in order to control water usage in an MCU node. Farmers are able to control the engine from anywhere in the world.

Joshi [3] looked at the development of wireless environmental sensor nodes for use in agriculture, with the aim of monitoring climatic conditions and identifying the ideal external conditions for achieving optimal crop yields in certain agricultural areas. The study highlights the current body of literature on wireless environmental sensor nodes that are designed to monitor climatic factors and establish the best external conditions for maximizing crop yields in specific farming regions. The agriculture and food production industry has recently turned its attention to wireless sensor networks (WSNs) in order to improve production and agricultural yield standards by making use of these cheap modern technologies. There has been a great deal of interest in wireless sensor networks (WSNs) recently. According to Mekonnen [5], the current study provides an extensive analysis of several machine learning algorithms' application in the field of sensor data analytics in agro ecosystem. It contains a case study on the IoT-based smart farm prototype for integrated food, energy, and water (FEW). According to Sangeeta et al. [4], the current study presents a machine learning strategy that is designed to predict the best crop yield in a specific area by assessing such a number of environmental factors as precipitation, temperature, and wetness, soil type and pH, and past plant crop records.

Based on data extraction analyzes, Ghadge [6] recommended that farmers monitor the soil fertility. Therefore, the approach focuses on the monitoring of the soil quality to determine what crop is suitable for the production depending on

the type of the soil and to maximize the crop production using the prescribed fertilizer. According to Sujawat [7], artificial intelligence has huge applications across numerous fields. Due to its capacity to comprehend issues, identify the causes of them, and identify the best solutions, artificial intelligence can be of great assistance in treating agricultural disorders. The paper provides a brief overview of the use of artificial intelligence in agriculture, the various agricultural techniques that are accessible, and the various methods that may be used to identify plant diseases. In addition to explaining the use of hybrid artificial intelligence for feature extraction, classification and prediction, and modeling using various algorithms and optimization techniques, Kshirsagar and Akojwar [8-11] expounds on the application of artificial intelligence for various classification and prediction problems. Artificial intelligence, case-based reasoning, multiagent optimization, scheduling, data mining, web crawlers, understanding and interpreting natural languages, and virtual vision reality have all shown notable results [12-14].

III. CNN Method for the Classification of Soil Images.

The proposed methodology employs a Convolutional Neural Network (CNN) to classify soil images based on their visual characteristics. CNNs are particularly well suited for image-based analysis because their convolutional layers can automatically learn and extract discriminative spatial features, including edges, textures, shapes, and structural patterns. These learned features enable the network to effectively differentiate among various soil classes.

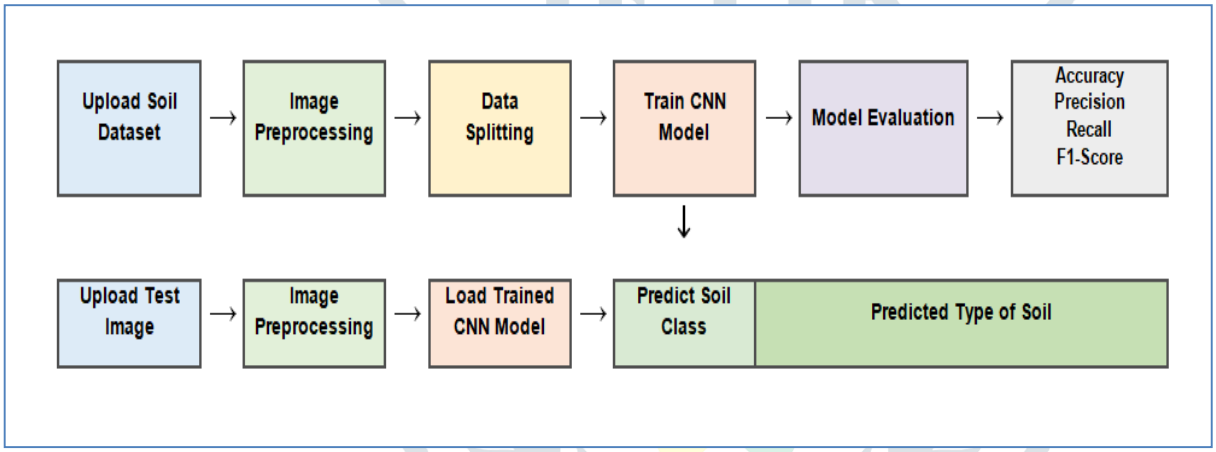

Fig. 1. CNN-Based Soil Image Classification – Block Diagram

Phase-1: The first step is to gather and arrange a variety of soil image datasets, with a particular emphasis on four different types of soil: red soil, clay soil, black soil, and alluvial soil. Alluvial soil, which is found in river basins, is usually pale in color and has a fine texture. Cotton growing is frequently linked to black soil, which is productive and dark. Red soil is more porous and reddish because of its iron oxide concentration, while clay soil has a compact texture and holds water effectively. In order to effectively depict differences in soil parameters like moisture content or texture, the collection should contain high-quality photos shot under consistent lighting circumstances. The classification system uses these pictures as its primary input.

Phase-2: In order to enhance the quality of the images of soil samples and guarantee consistency across the entire database, it is essential to prepare the dataset. To unify the input dimensions for machine learning models, the images are downsized to a specific resolution, such as 224x224 pixels. To accelerate the converging process during the model training stage, normalization of the pixel values to a range between 0 and 1 is applied. Image augmentation techniques, including flipping, rotation, and noise addition, are implemented to enlarge the size of the dataset and increase the model's resilience to variations in the appearance of the samples. Additionally, in order to amplify the appearance of particular types of soil, noise reduction methods, such as the application of the Gaussian filter, are utilized to remove undesirable artifacts from the images.

Phase-3: In the final part, quantitative measures should be utilized to compare the performance of the CNN model and the Extra Trees Classifier. While precision and recall assess the model's dependability and sensitivity in identifying each

type of soil, accuracy measures demonstrate how accurate forecasts are generally. For a fair assessment, the F1-score, which is the harmonic mean of recall and precision, should also be taken into account. To evaluate the viability of each approach, computational efficiency (such as training and prediction timeframes) should also be examined. The comparison validates the superiority of CNN as the suggested solution by demonstrating its ability to achieve higher classification accuracy through capturing complex information from the images of the soil.

MODEL BUILDING

Several crucial processes are involved in creating a machine learning (ML) model. First, choose the right procedure according to the dataset's properties and the type of issue (classification, regression, clustering). Preprocess the dataset by dividing it into training and testing sets, scaling features, and cleaning and transforming the data. Use a package like scikit-learn or TensorFlow to initialize the model, then adjust the hyper parameters as necessary. Fit the training data to find patterns and relationships in order to train the model. Depending on the type of problem, use measures such as accuracy, precision, recall, or RMSE to assess the trained model on the testing data. Optimize the model, if needed, by experimenting with other algorithms or optimizing hyper parameters. Then, use Joblib or Pickle to store the model for use in production if the model attains an acceptable performance level. Finally, apply the model as part of an application or system to make predictions on new data sets.

An artificial neural network known as a convolutional neural network or CNN is a deep learning model used in processing and analyzing visual data such as images and videos. CNNs are highly efficient at learning visual patterns in images through the automatic discovery of image features and eliminate the need for manual feature extraction. This model employs multiple layers to progressively learn image features from lower to higher levels while eliminating the requirement for human expertise in feature engineering. Convolutional neural networks are primarily used for image classification, object detection, and pattern recognition tasks by learning hierarchical patterns of visual data through convolutional, pooling, and fully connected layers that are inspired by the human visual system.

The design involves the input layer with the input images with a shape of 64 × 64 × 3 (for RGB images). The convolutional part of the architecture processes images with several convolutional layers that use filters to detect edges, patterns, and objects. These layers generate feature maps that are fed to the max-pooling layers. The feature maps are reduced to lower resolutions through max-pooling to minimize the model's complexity. The resolution is reduced while maintaining the most informative parts of the feature maps. A Rectified Linear Unit (ReLU) activation function introduces non-linearity and enables the model to learn complex relationships. The fully connected layer takes the output from the previous layers, flattens it into a one-dimensional array, and learns to associate the features with the classes of images. Lastly, for the purpose of categorization through the use of a softmax output layer, probabilities are assigned to various categories. In order to extract relevant information layer by layer, and in turn develop a deeper understanding of the image, CNNs process images by sliding filters through the input. CNNs are ideal for a variety of computer vision applications due to their robust and efficient structure.

IV. CNN MODEL IMPLEMENTATION AND RESULTS

All 1215 photos in the given collection are categorized based on the visual appearance of the soil, which can be useful in the work of precision agriculture. The four categories include Alluvial Soil (523 photos), Black Soil (228 photos), Clay Soil (197 photos), and Red Soil (267 photos). Alluvial Soil is a type of soil found in alluvial deposits of river valleys and plains and is very fertile. Black Soil is a black-colored clayey soil, which is highly moisture-resistant and common in ancient volcanic areas, therefore it is very suitable for cotton production. Clay Soil is a fine-textured soil, which has a decent level of fertility, but it is hard to work with due to its structure. Finally, Red Soil gets its name from the characteristic red color caused by iron oxide. In order to enable the creation of AI models capable of recognizing and classifying soils automatically, this dataset is intended for image-based categorization of soil types. The data will contribute to the improvement of techniques such as irrigation, fertilization, and crop choice, which in turn will lead to increased farming production and efficiency.

RESULTS

Based on precision and recall, the performance of each type of soil classification is shown in Fig. 2. Both Black Soil and Alluvial Soil have high recall and precision, suggesting that the model can accurately classify and identify them with few

false positive and negative cases. By contrast, Clay Soil has high recall but low precision, indicating that although the model can correctly classify most of the real Clay Soil cases, it may frequently misidentify other types of soil as Clay Soil. Finally, with a precision and recall of zero, Red Soil poses a significant challenge to the model. This indicates that the model is either completely unable to predict Red Soil or mistakenly classifies all examples of Red Soil , suggesting that the model's performance is severely deficient and may require further investigation, such as additional data, features, or modifications to the model itself.

The comparison of the CNN and Extra Trees Classifier (ETC) models' performance for the four soil classes is demonstrated in Fig. 3. The CNN model shows better precision and recall than the ETC for the majority of the classes, including Alluvial and Black Soils. For the CNN, the precision for the Clay Soil is significantly increased without sacrificing its recall. This implies that the CNN correctly recognizes more positive samples and reduces the number of negative samples identified as positive. On the contrary, both models totally fail to recognize Red Soil, with precision and recall equal to zero. This suggests that none of the samples were predicted as positive, and thus the class might be completely misclassified. The CNN model has been found to be more effective in capturing the true classifications as witnessed in the comparison confusion matrix. In most instances, the predicted and actual values are found to be in unison.

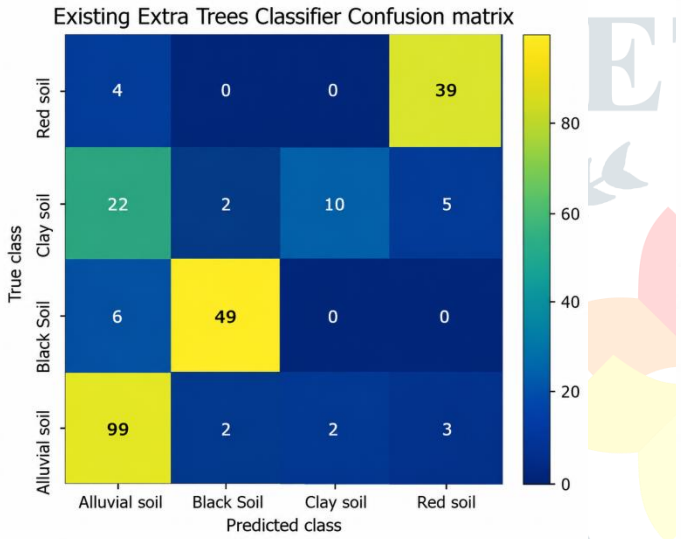

Fig 2. Confusion matrix of Extra Tree Classifier.

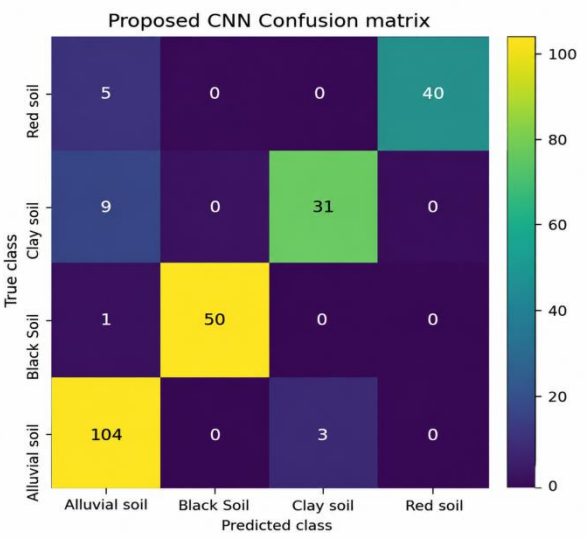

Fig. 3. Confusion matrix CNN

Table 1: Comparative Analysis: Existing Extra Trees Classifier vs Proposed CNN Algorithm

Metric	ETC model	Proposed Deep CNN model
Accuracy	81.07%	92.59%
Precision	83.58%	94.64%
Recall	74.71%	90.41%
F1-Score	75.04%	92.24%
Sensitivity	98.02%	100%
Specificity	89.09%	98.04%

Comparisons were made between the Existing Extra Trees Classifier (ETC) and the Proposed CNN Algorithm where the latter was found to be better than the former in all key statistical measures. The accuracy of the ETC model at 81.07% is lower than that of the CNN at 92.59%. Furthermore, CNN's recall of 90.41% is significantly higher than ETC's 74.71%,

while its precision of 94.64% is higher than ETC's 83.58%. Besides, in terms of F1-score (92.24% vs. 75.04%), sensitivity (100% vs. 98.02%), and specificity (98.04% vs. 89.09%), the CNN also performs better than the ETC. For some classes, like Black Soil and Red Soil, the CNN even achieves extraordinary class-wise performance, such as flawless precision/recall, which indicates its ability to correctly classify soil types with fewer errors. Therefore, it can be concluded that the proposed CNN model can provide a more reliable solution to the task of soil categorization.

V. CONCLUSIONS AND FUTURE WORK

The proposed CNN model shows significant improvement over the current model in terms of accuracy, precision, recall, and overall classification performance compared to the Extra Trees Classifier (ETC). The ETC model had an accuracy of 81.07%, with excellent precision (83.58%) and specificity (89.09%), but it had poor recall (74.71%) and F-score (75.04%). This indicates that even though it is proficient in detecting true positives, it cannot correctly identify all relevant cases. In addition, the confusion matrix shows that the ETC model performs poorly on Red Soil, with 0 precision and recall, but well on Alluvial and Black Soil. With the accuracy of 92.59%, precision of 94.64%, the recall is 90.41%, and the F-score accounts for 92.24%, the CNN model is on the other hand able to classify samples better, which guarantees a balanced performance. A high specificity (98.04%) means a lower false positive rate while the sensitivity reaches 100%, which means that the model can correctly recognize all the positive cases. On the other hand, the CNN model can better classify the Clay Soil, which solves the low precision problem of the ETC model. By contrast, both models failed to classify the Red Soil well, implying that further improvements or more features need to be considered to enhance the performance of Red Soil classification. The overall improvements in the confusion matrix verify that CNN provides a more reliable and efficient way for classification. It is clear that the suggested model is a better option for soil classification tasks, providing improved reliability and generalization, given the CNN's superior accuracy and balanced precision-recall trade-off. Thus, using deep learning models like CNN for soil categorization can result in significant gains in predictive performance, improving land management and agricultural decision-making.

REFERENCES

[1] H. Sharma, A. Haque, and Z. A. Jaffery, "Smart agriculture monitoring using energy harvesting Internet of Things (EH-IoT)," *An International Scientific Journal*, vol. 121, pp. 22–26, 2019.

[2] M. Suchithra, "Sensor data validation," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 12, pp. 14327–14335, 2018.

[3] P. Joshi, "Wireless sensor network and monitoring of crop field," *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, vol. 12, no. 1, pp. 23–28, 2017.

[4] S. G. Sangeeta, "Design and implementation of crop yield prediction model in agriculture," *International Journal of Scientific & Technology Research*, vol. 8, no. 1, 2020.

[5] Y. Mekonnen, "Review—machine learning techniques in wireless sensor network based precision agriculture," *Journal of The Electrochemical Society*, vol. 167, no. 3, Art. no. 037522, 2020.

[6] R. Ghadge, "Prediction of crop yield using machine learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 2, pp. 2237–2239, 2018.

[7] G. S. Sujawat, "Application of artificial intelligence in detection of diseases in plants: A survey," *Turkish Journal of Computer and Mathematics Education*, vol. 12, 2021.

[8] P. Kshirsagar and S. Akojwar, "Optimization of BPNN parameters using PSO for EEG signals," in *Proc. Int. Conf. Communication and Signal Processing*, India, 2016, pp. 385–394.

[9] S. Oza, A. Ambre, S. Kanole, et al., "IoT: The future for quality of services," in Proc. ICCCE 2020, Springer, Singapore, 2020, pp. 291–301.

[10] P. K. Kollu, K. Kumar, P. R. Kshirsagar, et al., "Development of advanced artificial intelligence and IoT automation in the crisis of COVID-19 detection," Journal of Healthcare Engineering, vol. 2022, Art. no. 1987917, 2022.

[11] Ayesha Banu , Ranjith kanna Kanna, Jayalaxmi marripadagala, Veldandi Kiran. "Using Machine Learning Algorithms For Detecting Cyber Attacks". Journal of Interdisciplinary Cycle Research. Volume 18, Issue 9,Pages 1-6.

[12] P. R. Kshirsagar, P. P. Chippalkatti, and S. M. Karve, "Performance optimization of neural network using GA incorporated PSO," Journal of Advanced Research in Dynamical and Control Systems, vol. 10, no. 4, pp. 156–169, 2018.

[13] N. Ahmed, D. De, and I. Hussain, "Internet of Things (IoT) for smart precision agriculture and farming in rural areas," IEEE Internet of Things Journal, vol. 5, no. 6, pp. 4890–4899, 2018.

[14] S. Sundaramurthy, C. Saravanabhavan, and P. Kshirsagar, "Prediction and classification of rheumatoid arthritis using ensemble machine learning approaches," in Proc. 2020 Int. Conf. Decision Aid Sciences and Applications (DASA), India, 2020, pp. 17–21.

[15] Ayesha Banu, Dr. J Sravanthi, Dr. B. Swathi, Dr. P. Latha, B Maheshwari, Mohammad Sohail. "Data Analytics in Machine Learning". Journal of Information Systems Engineering and Management. Volume 10, Issue 31s, Pages 114-122.

[16] M. Padmaja, S. Shitharth, K. Prasuna, A. Chaturvedi, P. R. Kshirsagar, and A. Vani, "Grow of artificial intelligence to challenge security in IoT application," Wireless Personal Communications, 2021.