International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 1s
DOI: https://doi.org/10.17762/ijritcc.v10i1s.5823
***Article Received**: 29 August 2022 **Revised**: 01 November 2022 **Accepted**: 26 November 2022*

A Novel Hybrid AI Federated ML/DL Models for Classification of Soil Components

Dr. Mahendra Eknath Pawar[1], Dr. Rais Allauddin Mulla[2], Dr. Sanjivani H. Kulkarni[3], Sajeeda Shikalgar[4], Dr. Harikrishna B. Jethva[5], Prof. Gunvant A. Patel[6]
[1]Associate Professor, Vasantdada Patil Pratishthan College of Engineering and Visual Arts, Mumbai, Maharashtra, India
[2]Associate Professor, Vasantdada Patil Pratishthan College of Engineering and Visual Arts, Mumbai, Maharashtra,India
[3]Assistant Professor, Computer Science and Engineering, MIT World Peace University, Pune, Maharashtra, India
[4]Assistant Professor, School of computer Science and Technology, MIT World Peace University, Pune, Maharashtra, India
[5]Associate Professor, Department of Computer Engineering, Government Engineering College, Patan, Gujarat, India
[6]Associate Professor, Department of Electrical Engineering, Government Engineering College, Patan, Gujarat, India
mahendraepawar@gmail.com[1],mtechraismulla@gmail.com[2], sanjivani.kulkarni@mitwpu.edu.in[3], sajeeda.shikalgar@mitwpu.edu.in[4], hbjethva@gmail.com[5], gapatel09@gmail.com[6]

Abstract— The soil is the most fundamental component for the survival of any living thing that can be found on this planet. A little less than 41 percent of Indians are employed in agriculture, which accounts for approximately 19 percent of the country's gross domestic product. As is the case in every other industry, researchers and scientists in this one are exerting a lot of effort to enhance agricultural practices by utilising cutting-edge methods such as machine learning, artificial intelligence, big data, and so on. The findings of the study described in this paper are predicated on the assumption that the method of machine learning results in an improvement in the accuracy of the prediction of soil chemical characteristics. The correlations that were discovered as a result of this research are essential for comprehending the comprehensive approach to predicting the soil attributes using ML/DL models. A number of findings from previous study have been reported and analysed. A state of the art machine learning algorithm, including Logistic Regression, KNN, Support Vector Machine and Random Forest are implemented and compared. Additionally, the innovative Deep Learning Hybrid CNN-RF and VGG-RNN Model for Categorization of Soil Properties is also implemented along with CNN. An investigation into the significance of the selected category for nutritional categorization revealed that a multi-component technique provided the most accurate predictions. Both the CNN-RF and VGG-RNN models that were proposed were successful in classifying the soil with average accuracies of 95.8% and 97.9%, respectively, in the test procedures. A study was carried out in which the CNN-RF model, the VGG-RNN model, and five other machine learning and deep learning models were compared. The suggested VGG-RNN model achieved superior accuracy of classification and real-time durability, respectively.

Keywords—Soil Classification, Soil Nutrients, Machine Learning, feature extraction, segmentation

I. Introduction

Farming is, and always has been, one of the key pillars on which the economy of the Indian subcontinent is built, and this remains true today. This is because over two-thirds of Country's population derives their primary source of income, or subsistence, directly from agriculture. Another point that should not be overlooked is the fact that it is responsible for 20 percent of India's gross domestic product (GDP). The farmer, also known as the Annadatta (Food Producer) of our country, is at the epicenter of the agricultural industry and is currently up against a number of challenges, including the following:

1) Due to the wide variety of soil types found across the United States, farmers typically have a difficult time determining which crop will be most ideal and profitable for their particular soil, conditions, and region, and as a result, they end up suffering a great deal of financial loss.

2) Because of the unpredictability of the weather, it is currently very challenging for farmers to estimate both the yield that will result from a specific growing season and the revenue that they will make from their labour.

3) Due to the "farm to market" system, which involves hundreds of middlemen who consume the majority of the revenues by carrying and selling goods, farmers get exceptionally low prices for their produce. This is a direct result of the "farm to market" approach. This mechanism results in farmers earning dismally low profits for their produce.

The health of the earth's soil is the single most critical factor in maintaining life. The continued existence of an unlimited number of lives is dependent on the soil. To create soil, a large quantity of minerals, air, water, organic and inorganic components, and the decomposing remains of living or nonliving creatures are required. In addition, the soil must be able to hold water. In the process of weathering, several kinds of organic materials are combined with the regolith that is derived from the surrounding soil. The process of weathering can be either physically, chemically, or biologically driven.

Sand, clay, silt, peat, chalk, and loam soil are the six different types of soil that may be distinguished from one another based on the characteristics of the soil. When it comes to cultivating any kind of plant, you need specific soil nutrients in the right amount, as well as the right temperature and level of moisture, among other things.

It is essential for the soil to have the right quantity of nutrients in order to produce a crop of higher quality. It's possible that the lack of nutrition will result in a significant loss. Whether the level of nutrients is too low or too high, either scenario is detrimental to the cultivation of the crop [13]. The growth of the crop will be halted if the nutrient level is too low, and the crop will not continue to develop if the nutrient level is too high; for instance, more foliage but fewer or no fruits indicates that the nitrogen level is very high; yellowing of the leaves indicates that there is too much manganese, etc. A farmer needs to approach an approved laboratory in order to test the contents of their soil. They also need to wait for the results of the test, and some of the chemical characteristics are neglected during the testing process; only a select number are examined. Because of the delay in the results, the crop may suffer damage, which could result in a significant financial loss for the farmer [1].

The field of modern agriculture is a fertile testing ground for the use of both artificial intelligence and machine learning techniques. Technologies including such precision farming and crop suggestion engines can be used to enhance the overall harvest quality, yield forecast, detection of pests in plants, and poor nutrition on farms. These improvements can be made possible via the use of technology. These improvements are possible because of advances in technology. The troubled agricultural industry may benefit from the use of AI technologies, which may provide it a much-needed boost.

The field of artificial intelligence is the one that is expanding at the fastest rate and is becoming interwoven into practically every facet of human life. It has been demonstrated to be a useful tool that offers a second viewpoint, draws attention to information that is difficult to see, and forecasts behaviour based on previous experiences and learning algorithms. In most cases, the findings are dependent on a number of different aspects like the amount of the study dataset, the parameters of the algorithm, the kind of soil, and the categories that are to be evaluated. It is very challenging to replicate any published results with the same level of precision when using various research datasets since there is a significant amount of variation in the variables and the combinations of those elements.

Machine Learning and Deep Learning has played a vital role in various sectors of the business and corporates. Few of our previous works has included the use of ML/DL models for Competitor mining [16] [17], for predicting the trend of stock market [18][19] and also for categorization of various languages [20]. Thus, in this paper we describes the model for soil analysis and crop prediction mechanism that uses a Hybrid of VGG-RNN and CNN-RF Algorithm to solve some of the long-standing problems of the Indian agricultural sector and increase profitability for the average farmer. We also present the performance analysis of this algorithms in comparison to various state of art ML/DL algorithms like, LR, KNN, SVM, RF and CNN.

II. Literature Survey

In the field of agriculture, data mining (DM) is finding more and more applications, particularly in the areas of soil classification, management of wastelands, crop management, and pest management. In [1] researchers examined a number of association tactics in DM and applied them to a soil science database to anticipate important associations and give association rules for diverse soil types in agriculture. Similarly, numerous data mining tools have already been applied to analyse agriculture forecast, disease detection, and pesticide optimization [2].

In [3], the author conducted research on the J48 classification system to determine how accurately it can estimate the fertility rate of soil. [4] investigated the use of a variety of DM methods for the purpose of information discovery in the agricultural industry. Additionally, it provided a variety of displays for knowledge discovery in the form of association rules, clustering, classification, and correlation.

Estimating the soil fertility classes using classification methods was accomplished in [5] with the help of the Nave Bayes, J48, and K-Nearest Neighbor algorithms. In [6,] approaches from the field of data mining were used to make estimates of agricultural yields. The method known as Multiple Linear Regression (MLR) was utilised in order to ascertain the existence of a linear connection between the dependent and independent variables. Rainfall was chosen as the primary criterion, and the K-Means clustering approach was utilised in order to divide the data into four distinct categories.

The article [7] investigates the vegetative factors that determine the location and severity of landslides in the Shimen reservoir watershed in northern Taiwan. We made use of non-linear approaches, in addition to decision trees and Bayesian

Network data mining tools. In comparison to the non-linear method, it was determined that the optimization-based Bayesian Network strategy was more effective.

The author of [8] investigated the relative significance of soil fertility and crop management characteristics in the process of predicting maize yields, as well as in the process of affecting yield variability and farmer differences. In order to make an accurate prediction of the outcome, we used classification and regression tree analysis.

Both a production-related yield gap and a soil fertility-related nutrient balance were analysed in [9], where two comprehensive methods for calculating them were presented and discussed. Using this technology, which allows information to be transported from micro scale to larger sizes, an assessment of the land's quality is carried out.

The author of [10] employed classification algorithms to evaluate soil data and make predictions about the qualities of the soil. Classification algorithms such as Nave Bayes, J48, and JRip were used to categories soil parameters such as pH, Electrical Conductivity (EC), Potassium, Iron, Copper, and others. Among all of the algorithms, J48 stood out as a straightforward choice because it produced the best outcomes.

The author of [11] utilised three different machine learning algorithms, namely "Logistic Regression," "Naive Bayes algorithm," and "Random forest algorithm," in order to select appropriate crops and forecast production value. The author then compared the outcomes of these methods. In order to train their models, they gathered historical data regarding the weather, temperature, and a number of other aspects. The Random Forest Algorithm came out on top as having the most accurate results out of the three algorithms.

The authors of this study [12] developed a method that would determine the crop that would be most successful by taking into account the PH of the soil, its contents, the weather, and the amount of rainfall. In order to make accurate forecasts of rainfall, they made use of a Support Vector Machine (SVM) method. The result was input into the Decision Tree Algorithm, which was used to make crop predictions.

The authors of [13] employed an algorithm called the Random Forest Algorithm to make agricultural yield predictions for the Indian state of Tamil Nadu. Their model included variables such as the amount of precipitation, the highest temperature, the time of year, and the production.

The Random Forest Algorithm was utilised by the author in the previous referenced work [14] in order to make crop yield projections for the state of Maharashtra. In order to train a machine learning model, they collected data from various government websites as well as data relating to the environmental parameters (precipitation, temperature, cloud cover, and vapor pressure) on a monthly basis.

The author of presented a method in [15] that would identify the types of crops that would be best suited for a farmer depending on the nutritional characteristics of the soil. These characteristics would be determined in a laboratory based on a sample of soil that was gathered by the farmer. For the purpose of prediction, the Naive Bayes algorithm was utilised, and the accuracy of the model was 75%.

III. Implementation Methodology

This system makes use of four key classification algorithms for comparative purposes, in addition to three deep learning algorithms, one of which is an innovative hybrid model that is more accurate than models that are already in use. In this section, we evaluate the efficacy of a variety of classifiers that were utilised in the method, and we proceed to integrate a total of four algorithms in order to develop the suggested model.

A. Support Vector Machine

The Support Vector Machine is an example of a common supervised learning technique that may be used for service classification and regression. To make it easier to place newly acquired data points in the appropriate category, the SVM algorithm is intended to locate the line or decision limit in the class of dimensional space that provides the best estimate of the true value of the data. The SVM is used to choose the extreme points and vectors that ultimately result in the construction of the hyperplane. The algorithm is known as the supporting vector machine, and in the most extreme circumstances, it uses help vectors. Consider, for instance, the decision limit or hyperplane shown in figure 1, which is utilised to define two groups in a distinct manner:

Hyperplane: When dividing classes across an n-dimensional space, there are multiple lines and decision boundaries that can be used. However, in order to assist in the definition of data points, we need to select the optimal judgement boundary. In SVM, the term "hyperplane" is used to refer to the limit that provides the best results.

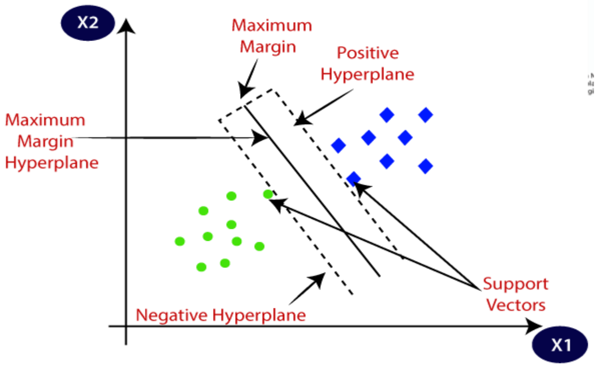


Figure 1 SVM classification

DOI: https://doi.org/10.17762/ijritcc.v10i1s.5823
***Article Received**: 29 August 2022 **Revised**: 01 November 2022 **Accepted**: 26 November 2022*

The dimensions of the hyperplane are determined by the properties of the dataset. If you locate two features, the hyperplane will be in the form of a straight line (as seen in the image). If all three of these conditions are met, the hyperplane in question takes on the form of a two-dimensional plane. Always representing the distance between data items, a high-level hyperplane shows how far apart they are. We are looking for a classifier that can distinguish between green and blue coordinates (x1, x2). Take a look at the example given in picture 2 below:

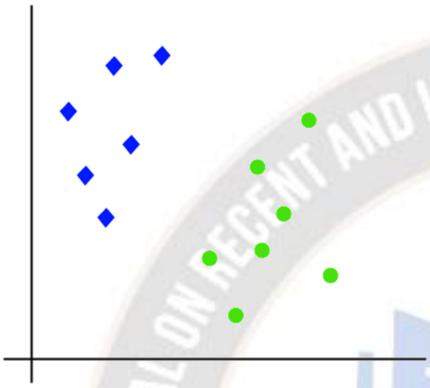


Figure 2 2D space of data plotting

Given that it is only two dimensions, Figure 3 in the previous illustration illustrates how the data were represented in the space using blue and green hues. The following picture illustrates the process of creating hyperplanes.

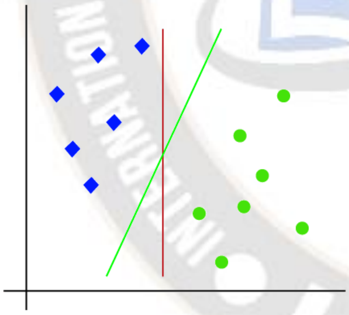


Figure 3 Drawing lines or creating hyperplanes

As a consequence of this, the SVM Algorithm is helpful in determining the optimal decision line or boundary; the optimal decision line or boundary is called a hyperplane, and the optimal decision area is also called a hyperplane. The SVM algorithm represents the middle ground between the two groups. These locations have been given the names of vectors of assistance by the staff. The objective of the SVM is to achieve the greatest possible distance. The space that separates the hyperplane and the vectors is referred to as a margin. The ideal hyperplane is the one with the greatest possible profit margin.

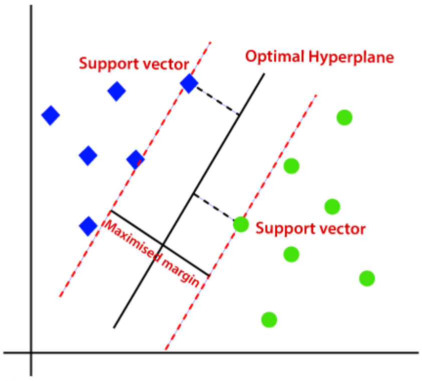


Figure 4 Hyperplane and support vector drawing

B. Logistic Regression

The procedure known as logistic regression is used in machine learning to visualise the correlations that exist between binary dependent variables and independent variables. In the case of logistic regression, the output must invariably take the form of a discrete or absolute value. The result should be consistent with the values provided by 0 and 1, and it can either be Yes or No. Logistic Regression, which is very similar to Linear Regression, is used to address problems that include classification. On the other hand, rather than having a linear line, it employs an S-shaped logistic function, which shows the likelihood of an event occurring in a given set of values. In other words, this function takes the place of a linear line. Because it is able to generate probabilities for fresh data sets, this approach is particularly valuable for machine learning. It is possible to categorise numerous sets of data with the assistance of Logistic Regression, which can also be used to find the classification methods that are the most effective.

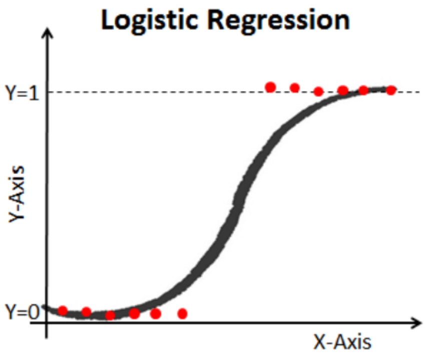


Figure 5 Logistic Regression

The logit function is used in the practise of logistic regression, which is a method for forecasting the likelihood of occurrence of something like a binary event. Logistic regression is a methodology. The target variable in this particular instance of linear regression is of a categorical nature, rather than an ordinal nature. It is an example of linear regression. The log of the chances is the dependent variable that we are looking at.

DOI: https://doi.org/10.17762/ijritcc.v10i1s.5823
***Article Received**: 29 August 2022 **Revised**: 01 November 2022 **Accepted**: 26 November 2022*

Linear Regression Equation:

$$\Phi = \delta 0 + \delta 1x1 + \delta 2x2 \dots\dots + \delta mxm$$

Where y is the dependent variable and x1, x2 … xn are descriptive variables.

Sigmoid Function:

$$g(x) = \frac{1}{1 + e^{-(x)}}$$

The sigmoid function is a straightforward mathematical operation that, in addition to being known by its other name, the logistic function, the sigmoid function transfers the values of definite values to the corresponding probabilities. In logistic regression, the threshold value is what determines the probability that a certain value would be found. The values 0 and 1 are required for the logistic regression function, and these values cannot go above the limit set by the S-form curve.

$$p = 1/\{1 + e^{-(\delta 0 + \delta 1x1 + \delta 2x2 \dots\dots + \delta nxn)}\}$$

D. KNN - The K Nearest Neighbor method is a type of supervised learning that is typically utilised for the purposes of classification and regression. It takes into account the connections between the data points and then makes a prediction about the category or continuous value of the new data point. In this particular illustration, we have two categories, which we will refer to as Category A and Category B respectively. In order to determine which of the two categories the new data point falls under, which is a necessary step for solving this problem, we must first determine which of the two groups the new data point belongs to.

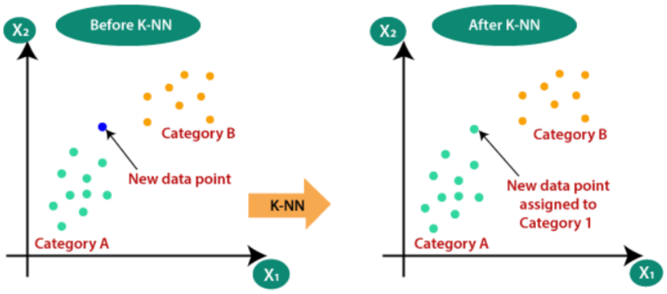


Figure 6 KNN classifier

D. Hybrid CNN-RF Model

In the hybrid CNN-RF model, the RF algorithm is used to categories high-level characteristics that have been derived from the CNN. This approach takes into account the benefits that are offered by both CNN and RF. As input for the classification process, CNN-RF makes use of spatial information that have been retrieved from the optimal structure and parameters of the CNN. Therefore, additional stages of feature extraction or selection are not needed prior to the RF-based classification because they are not necessary. Bagging is one example of the more complex classification algorithms that may be applied in RF as opposed to CNN's fully connected layer, which is used as a classifier. [38] Furthermore, the potential benefits of RF, such as its resistance to outliers and its capacity to reduce overfitting, can improve the performance of the classifier even when proper or informative spatial features can indeed be extracted from CNN as a result of insufficient training data and input data. This is possible owing to the advantages of RF's ability to reduce fitting problem and its reliability to outliers.

In contrast to other machine learning models, such as SVM, RF calls for the setting of a relatively small number of parameters, such as the number of trees that are to be grown in the forest (ntree) and the number of variables that are to be used in the node partitioning (mtry) [16]. In order to discover the best possible values for ntree and mtry, a grid search technique was carried out using different permutations of the parameters. In order to acquire an acceptable level of classification performance from a CNN, numerous factors, such as the number of layers, the size of the image patches, and the number as well as the size of the convolution filters, need to be properly calculated.

The CNN-RF model utilises the same network architecture as the CNN model; however, before to carrying out the RF-based classification, certain parameters will need to be modified. The characteristics that are recovered from the fully connected layer are utilised as inputs for the RF classifier. In order to identify mtry and ntree, a grid search approach that is analogous to the one used in the RF model is performed. The architecture of the CNN-RF model that was created for this investigation is displayed in Figure 7.

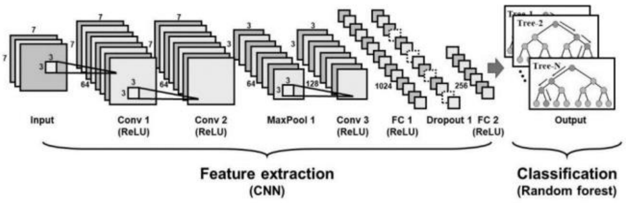


Figure 7. Architecture for Hybrid CNN-RF Classification Model

For the most effective model training, we used a five-fold cross-validation procedure. Only four of the five different breakdowns of the training data were utilised in the actual training of the model. The remaining partition was utilised as validation samples in order to locate the classification models with the best possible hyper-parameter settings.

E. Hybrid VGG-RNN Model

This paper studies and applies an improved deep hybrid network model in order to construct a model that is more accurate for the classification of the soil nutrients. This model is composed of a one-dimensional serial convolutional neural network (also known as a visual geometry group network, or VGG network), as well as a Recurrent Neural network (RNN network). Figure 8 provides an illustration of the structure of the model. The parameters of the soil's composition are sent into the VGG–RNN hybrid network model as an input, and the model's output is a classification of the soil according to one of the five standard classifications provided in the dataset. The model can be constructed and enhanced to the point where it can be implemented in real time if sufficient training is received.

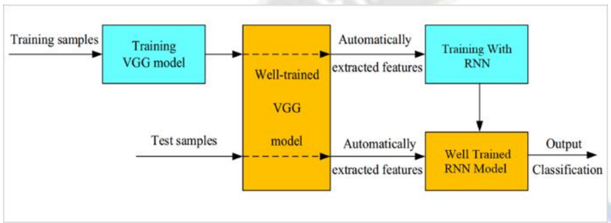


Figure 8. Architecture for Hybrid VGG-RNN Model

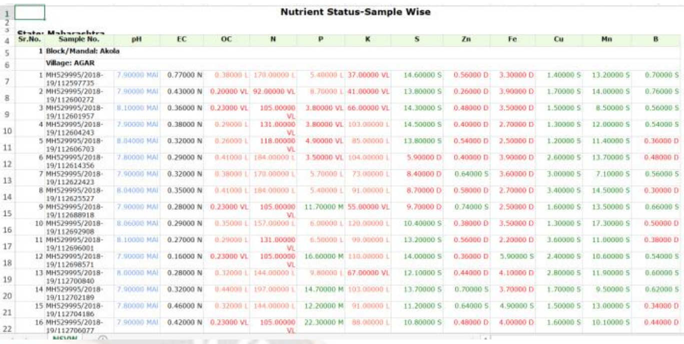
In order to improve the depth of the model and give a more sophisticated nonlinear transformation for the extraction of higher dimensional features, the convolutional neural model is connected serially rather than being set with a huge convolutional core. This is done for two reasons. The maximum pooling layer is utilised to control the risk of over-fitting, reduce the dimensionality of features, and ensure translation invariance. RNN network then mines the time-order properties of inertial data via RNN network, thereby realizing the effect of retaining long-term memory on the basis of selective memory. This is done on the basis that RNN network gets the feature fragments of convolutional neural network.

IV. Result and Analysis

A. Dataset Discription

The soil dataset was obtained by downloading it from the website http://www.soilhealth.dac.gov.in, which is run by the Department of Agriculture, Cooperation, and Farmers Welfare within the Ministry of Agriculture and Farmers Welfare under the Government of India. The Department of Agriculture in India had its own soil testing labs located in several District areas across the country. These labs accept soil samples from local farmers and recommend appropriate fertilisers to improve agricultural yields in the area.

On the internet page, the dataset can be obtained by first selecting the option labelled SOIL HEALTH DASHBOARD and then selecting the option labelled REPORT. Right now, the database only includes fifteen districts, but it contains more than fifty thousand samples; this is one reason why the state of Maharashtra is so highly regarded; this is because of the wide variety of cultural crops that are farmed throughout the entire state.



Nutrient Status-Sample Wise

State: Maharashtra

Sr.No.	Sample No.	pH	EC	OC	N	P	K	S	Zn	Fe	Cu	Mn	B
1	Block/Mandal: Akola												
	Village: AGAR												
1	MH529995/2018-19/112597735	7.90000 MAl	0.77000 N	0.38000 L	170.00000 L	5.40000 L	37.00000 VL	14.60000 S	0.56000 D	3.30000 D	1.40000 S	13.20000 S	0.70000 S
2	MH529995/2018-19/112600272	7.90000 MAl	0.43000 N	0.20000 VL	92.00000 VL	8.70000 L	41.00000 VL	13.80000 S	0.26000 D	3.90000 D	1.70000 S	14.00000 S	0.76000 S
3	MH529995/2018-19/112601957	8.10000 MAl	0.36000 N	0.23000 VL	105.00000 VL	3.80000 VL	66.00000 VL	14.30000 S	0.48000 D	3.50000 D	1.50000 S	8.50000 S	0.56000 S
4	MH529995/2018-19/112604243	7.90000 MAl	0.38000 N	0.29000 L	131.00000 VL	3.80000 VL	103.00000 L	14.50000 S	0.40000 D	2.70000 D	1.30000 S	12.00000 S	0.54000 S
5	MH529995/2018-19/112606703	8.04000 MAl	0.32000 N	0.26000 L	118.00000 VL	4.90000 VL	85.00000 L	13.80000 S	0.54000 D	2.50000 D	1.20000 S	11.40000 S	0.36000 D
6	MH529995/2018-19/112614356	7.80000 MAl	0.29000 N	0.41000 L	184.00000 L	3.50000 VL	104.00000 L	5.90000 D	0.40000 D	3.90000 D	2.60000 S	13.70000 S	0.48000 D
7	MH529995/2018-19/112622423	7.90000 MAl	0.32000 N	0.38000 L	170.00000 L	5.70000 L	73.00000 L	8.40000 D	0.64000 S	3.60000 D	3.00000 S	7.10000 S	0.56000 S
8	MH529995/2018-19/112625527	8.04000 MAl	0.35000 N	0.41000 L	184.00000 L	5.40000 L	91.00000 L	8.70000 D	0.58000 D	2.70000 D	3.40000 S	14.50000 S	0.30000 D
9	MH529995/2018-19/112688918	7.90000 MAl	0.28000 N	0.23000 VL	105.00000 VL	11.70000 M	55.00000 VL	9.70000 D	0.74000 S	2.50000 D	1.60000 S	13.50000 S	0.66000 S
10	MH529995/2018-19/112692908	8.06000 MAl	0.29000 N	0.35000 L	157.00000 L	6.00000 L	120.00000 L	10.40000 S	0.38000 D	3.50000 D	1.30000 S	17.30000 S	0.50000 D
11	MH529995/2018-19/112696001	8.10000 MAl	0.27000 N	0.29000 L	131.00000 VL	6.50000 L	99.00000 L	13.20000 S	0.56000 D	2.20000 D	3.60000 S	11.00000 S	0.38000 D
12	MH529995/2018-19/112698571	7.90000 MAl	0.16000 N	0.23000 VL	105.00000 VL	16.60000 M	110.00000 L	14.00000 S	0.36000 D	5.90000 S	2.40000 S	10.60000 S	0.54000 S
13	MH529995/2018-19/112700840	8.00000 MAl	0.28000 N	0.32000 L	144.00000 L	9.80000 L	67.00000 VL	12.10000 S	0.44000 D	4.10000 D	2.80000 S	11.90000 S	0.60000 S
14	MH529995/2018-19/112702189	7.90000 MAl	0.32000 N	0.44000 L	197.00000 L	14.70000 M	103.00000 L	13.70000 S	0.70000 S	3.70000 D	1.70000 S	9.50000 S	0.62000 S
15	MH529995/2018-19/112704186	7.80000 MAl	0.46000 N	0.32000 L	144.00000 L	12.20000 M	91.00000 L	11.20000 S	0.64000 S	4.90000 S	1.50000 S	13.00000 S	0.34000 D
16	MH529995/2018-19/112706077	7.90000 MAl	0.42000 N	0.23000 VL	105.00000 VL	22.30000 M	88.00000 L	10.80000 S	0.48000 D	4.00000 D	1.60000 S	10.10000 S	0.44000 D

Figure 9. Structure of Dataset

B. Performance Parameter

We tested the models on the held-over test set after they had been trained. The confusion matrix was then used to calculate the performance measures. Elements of confusion matrix are used to signify expected and actual classifications. Classification process yields two classes: right and wrong.

In order to compute the confusion matrix, we evaluated four basic scenarios:

- The proportion of genuine positives that are accurately detected is measured by true positive (TP).
- False negative (FN) refers to incorrect predictions. It identifies instances that are malicious yet are wrongly predicted as normal by the model.
- False positive (FP) refers to an inaccurate positive prediction when the detected assault is actually normal.
- The fraction of true negatives that are correctly identified attacks is measured by true negative (TN).

Accuracy, recall, f1 score, and precision are used to evaluate proposed work performance. Measurements are derived from

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (1)$$

$$Precision = \frac{T_p}{T_p + F_p} \quad (2)$$

$$Recall = \frac{T_p}{T_p + T_n} \quad (3)$$

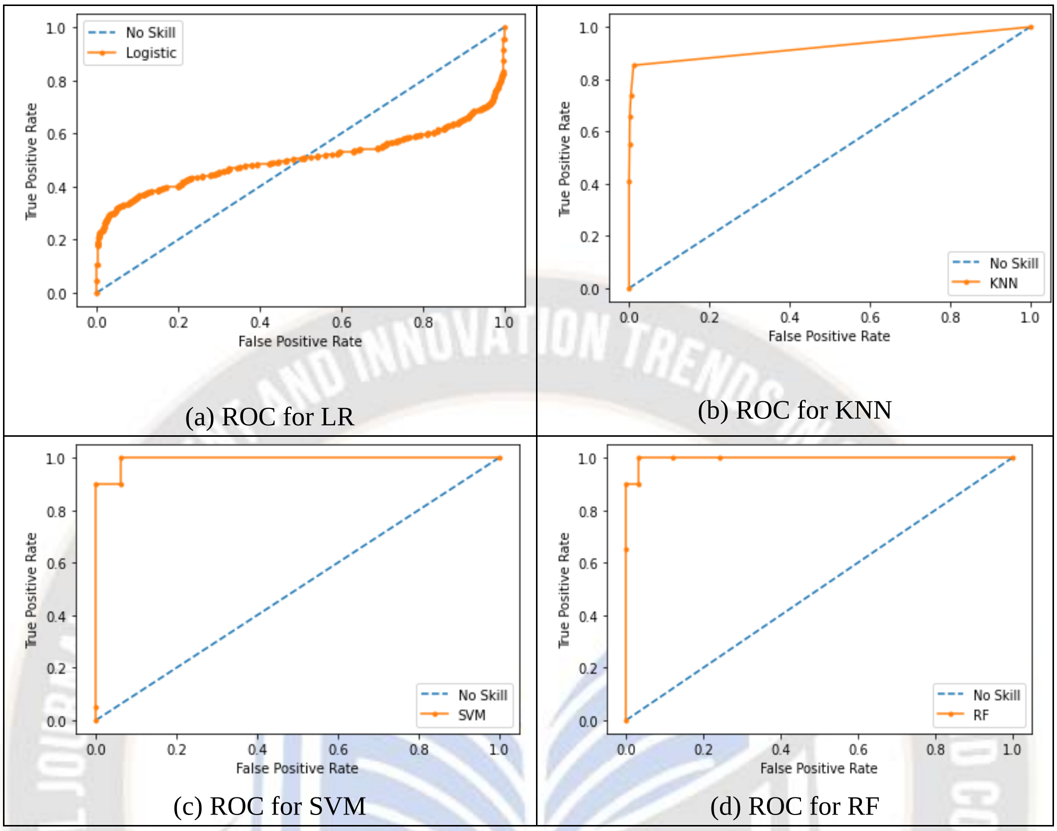
$$F1 = 2X \frac{Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

C. Experimental Results

In this section, we will give our findings from the comparative analysis of both systems. The comparison of the proposed systems with typical state-of-the-art mechanisms such as logistic regression, KNN, support vector machine, and

***Article Received**: 29 August 2022 **Revised**: 01 November 2022 **Accepted**: 26 November 2022*

random forest is also shown here for the purpose of providing a more accurate evaluation. Figure 10 illustrates the ROC Curve, which compares all of the different algorithms.



(a) ROC for LR

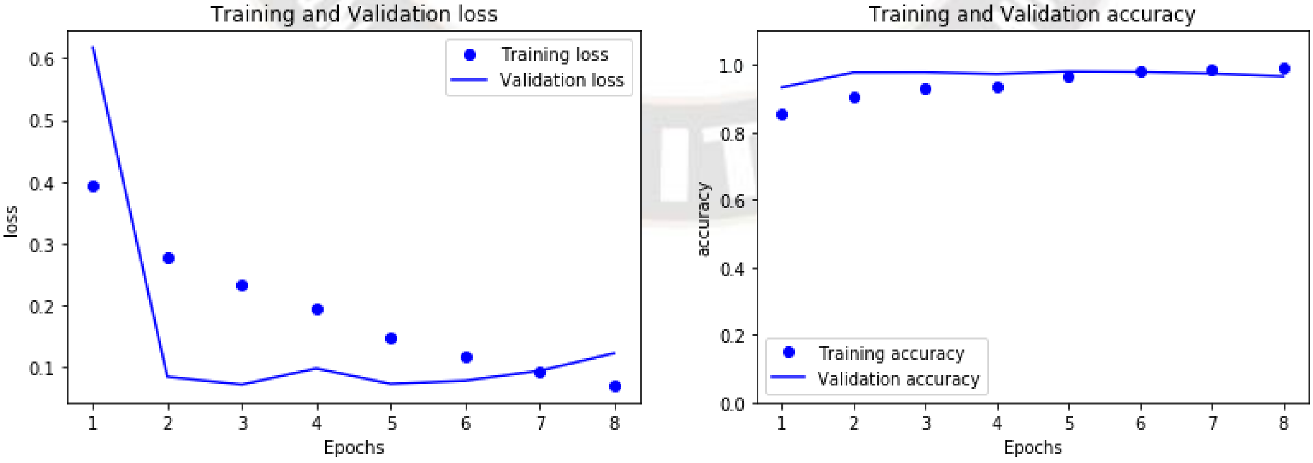
(b) ROC for KNN

(c) ROC for SVM

(d) ROC for RF

Figure 10. ROC Curve for (a) LR (b) KNN (c) SVM (d) RF

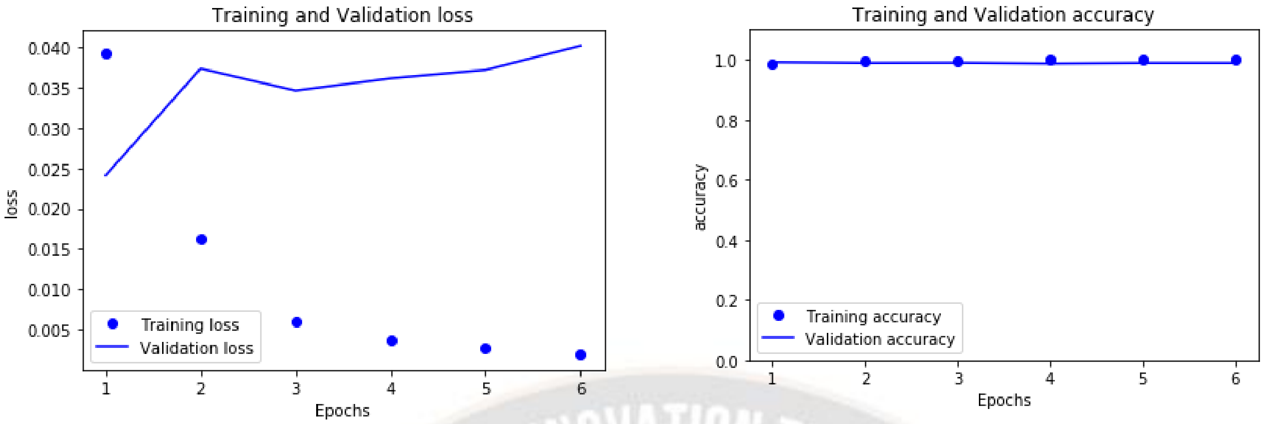
Apart from the above mentioned ML algorithms we experimented on 3 deep learning models. The deep learning models includes the CNN, a hybrid CNN-RF and Hybrid VGG-RNN models. In following figures, we presented the results of all the three algorithm in the form of Training and validation Loss and Accuracy. Figure 11 represents the training and validation loss and training and validation accuracy for CNN, similarly Figure 12 and 13 represents the training and validation loss and validation Accuracy for Hybrid CNN-RF and Hybrid VGG-RNN respectively.



(a) (b)

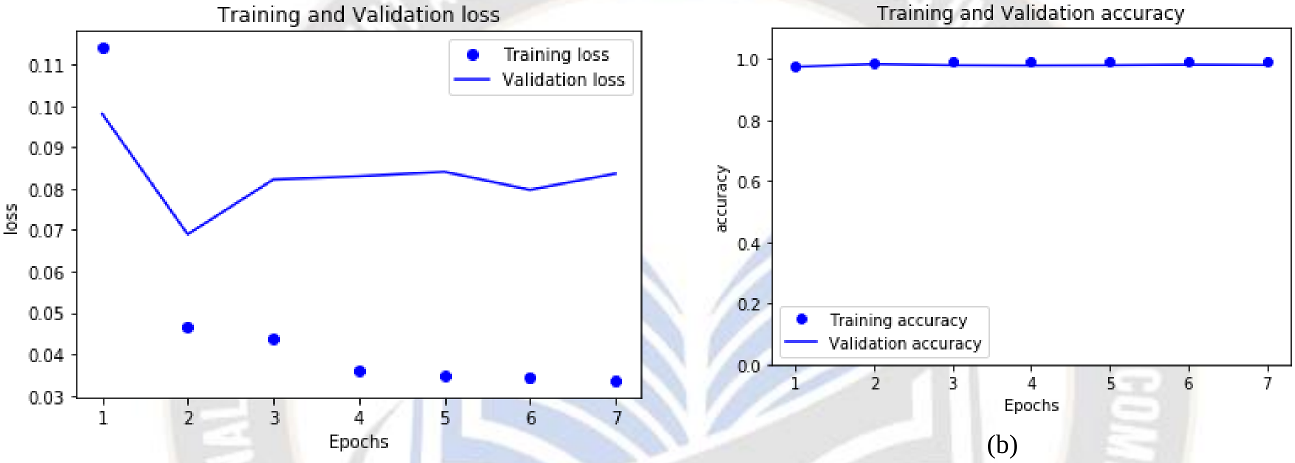
Figure 11. Training and Validation Loss and Accuracy for CNN Algorithm

DOI: https://doi.org/10.17762/ijritcc.v10i1s.5823
***Article Received**: 29 August 2022 **Revised**: 01 November 2022 **Accepted**: 26 November 2022*



(a) (b)

Figure 12. Training and Validation Loss and Accuracy for CNN-RF Algorithm



(b)

(a)

Figure 13. Training and Validation Loss and Accuracy for VGG-RNN Algorithm

From the above results it can be seen that the Hybrid VGG-RNN model has the least loss and highest accuracy among the three followed by the CNN-RF model. The detailed classification report of all three deep learning algorithms along with the four ML algorithms is presented in Table 1 as follows.

Table 1. Detailed Classification Report for all the Algorithms

Algorithms	Precision	Recall	F1 Score	Accuracy
LR	78	72	77	74.39
KNN	86	73	80	82.51
SVM	87	74	80	84.21
RF	88	75	81	85
CNN	82	92	84	95
CNN + RF	84	83	83	96
VGG + RNN	94	80	84	98

From the above table 1, it can be seen that among the Machine learning algorithms, RF has the highest accuracy with 85%, whereas the other ML algorithms i.e, LR has the accuracy of 74.39%, KNN has the accuracy of 82.51% and SVM is having the accuracy of 84.21%. Among the Deep learning algorithms, CNN yields the least accuracy of 95% whereas the hybrid CNN-RF has the accuracy of 96% and the hybrid VGG-RNN algorithm has the highest among all with 98% accuracy. From the above results we can say that the experiment model of hybrid VGG-RNN is the best performing algorithm among all the algorithms. Following figure 14 shows the comparative analysis of accuracy of all the algorithms.

DOI: https://doi.org/10.17762/ijritcc.v10i1s.5823
***Article Received**: 29 August 2022 **Revised**: 01 November 2022 **Accepted**: 26 November 2022*

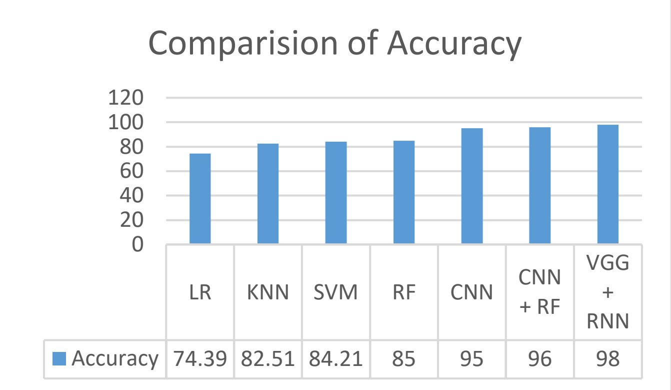


Figure 14. Comparative Analysis of Accuracy of all the algorithms

V. Conclusion

The success of India's agriculture industry is essential to the country's long-term economic expansion. Our objective was to provide autonomy to family farmers who only had a few acres of land by boosting their profitability and optimising crop productivity. The primary goal is to design a novel hybrid classifier that has an optimal feature selection approach, with the end goal of improving the accuracy of the soil categorization. The results demonstrated some progress made in comparison to the established procedures. The current standard algorithm, which consists of CNN, RF, VGG, and RNN, has been used as the basis for the hybridization of the new approach that has been proposed. In order to evaluate how useful the suggested method is, we compared it to other algorithms that represent the current state of the art. In preparation for the experiment, we visited the website of the agriculture department and downloaded the standard dataset. The experiment had a sample size of more than 80,000 instances. We conducted some research using a split of 70:30 between training and testing. It is possible to draw the obvious conclusion from the findings that the Hybrid VGG-RNN model provides greater performance than traditional methods, and that performance improves even further if better features are selected. Similarly, it can be seen that among the Machine Learning algorithms, RF has the highest accuracy with 85%, whereas the other ML algorithms, such as LR, have an accuracy of 74.39%, KNN has an accuracy of 82.51%, and SVM is having an accuracy of 84.21%. Similarly, it can be seen that among the Machine Learning algorithms, RF has the highest accuracy with 85%. The CNN algorithm has the lowest accuracy of all the deep learning algorithms, coming in at 95%, while the hybrid CNN-RF method has the accuracy of 96%, and the hybrid VGG-RNN algorithm has the best accuracy of all the deep learning algorithms, coming in at 98%. Based on the findings shown above, we are able to conclude that the experiment model of hybrid VGG-RNN is the method that performs the best among all of the algorithms. The subsequent figure 14 presents a comparative examination of the levels of accuracy offered by each method.

References

[1] Geetha MCS. Implementation of association rule mining for different soil types in agriculture. International Journal of Advanced Research in Computer and Communication Engineering. 2015 Apr; 4(4):520–2.

[2] Solanki J, Mulge Y. Different techniques used in data mining in agriculture. International Journal of Advanced Research in Computer Science and Software Engineering. 2015 May; 5(5):1223–7.

[3] Bhuyar V. Comparative analysis of classification techniques on soil data to predict fertility rate for Aurangabad District. International Journal of Emerging Trends and Technology in Computer Science. 2014 Mar-Apr; 3(2):200–3.

[4] Fathima NG, Geetha R. Agriculture crop pattern using data mining techniques. International Journal of Advanced Research in Computer Science and Software Engineering. 2014 May; 4(5):781–6.

[5] Suman, Naib BB. Soil classification and fertilizer recommendation using WEKA. International Journal of Computer Science and Management Studies. 2013 Jul; 13(5):142–6.

[6] Ramesh D, Vardhan VB. Data mining techniques and applications to agricultural yield data. International Journal of Advanced Research in Computer and Communication Engineering. 2013 Sep; 2(9):3477–80.

[7] Tsai F, Lai JS, Chen WW, Lin TH. Analysis of topographic and vegetative factors with data mining for landslide verification. Ecological Engineering. 2013 Dec; 61:669–77.

[8] Tittonell P, Shephered KD, Vanlauwe B, Giller KE. Unraveling the effects of soil and crop management on maize productivity in small holder agricultural systems of Western Kenya - An application of classification and regression tree analysis. Agriculture, Ecosystems and Environment. 2008 Jan; 123(1-3):137–50.

[9] Bindraban PS, Stroorvofel JJ, Jansen DM, Vlaming J, Groot JJR. Land quality indicators for suitable land management: Proposed methods for yield gap and soil nutrient balance. Agriculture, Ecosystems and Environment. 2000; 81:103–12.

[10] Gholap J, Lngole A, Gohil J, Shailesh, Attar V. Soil data analysis using classification techniques and soil attribute prediction. 2012 Jun; 9(3):1–4.

[11] Venugopal, S. Aparna, J. Mani, R. Matthew. and V. Williams, “Crop Yield Prediction using Machine Learning Algorithms,” Int. J. of Eng. Res. and Technol., issue 13, vol. 9, pp. 87-91, Aug 2021

[12] Mahendra N. , Dhanush V., Nischitha K., Ashwini and Manjuraju M. R, “Crop Prediction using Machine

Learning Approaches," Int. J. of Eng. Res and Technol., issue 8, vol. 9, pp. 23-26, Aug 2020

[13] Priya P., Muthaiah U. and Balamurugan M., "Predicting Yield of the Crop Using Machine Learning Algorithm," Int. J. of Eng. Sci and Res. Technol., issue 11,vol. 29, pp. 1248-1255, 2020

[14] M. Champaneri, D. Chachpara, C. Chandvidkar and M. Rathod, "Crop Yield Prediction using Machine Learning," Int. J. of Sci. and Res., issue 1, vol. 10, pp. 01-03, Apr 2018

[15] Bharath K.R., Balakrishna K., Bency C.A.., Siddesha M. and Sushmitha R., "Crop Recommendation System for Precision Agriculture," Int. J. of Comput. Sci. and Eng., issue 5, vol. 7, pp. 1277-1282, May 2019

[16] Pawar, M.E., Saini, S. (2022). Mining Top-K Competitors by Eliminating the K-Least Items from Unstructured Dataset. In: Saini, H.S., Singh, R.K., Tariq Beg, M., Mulaveesala, R., Mahmood, M.R. (eds) Innovations in Electronics and Communication Engineering. Lecture Notes in Networks and Systems, vol 355. Springer, Singapore. https://doi.org/10.1007/978-981-16-8512-5_54

[17] Mahendra Eknath Pawar, Satish Saini. (2021). MINING HIGH QUALITY ITEMSET FROM ONLINE REVIEWS USING ASPECT-BASED OPINION MINING AND MULTI-CLASS HYBRID CLASSIFICATION. Harbin Gongye Daxue Xuebao/Journal of Harbin Institute of Technology, 53(12), 271–279.

[18] Mulla, R.A., Saini, S. (2022). An Improved Stock Market Index Prediction System Based on LSTM. In: Pundir, A.K.S., Yadav, N., Sharma, H., Das, S. (eds) Recent Trends in Communication and Intelligent Systems. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-19-1324-2_15

[19] Rais Allauddin Mulla, Satish Saini. (2021). Machine Learning Based Framework For Making Adaptive Stock Market Index Prediction System. Harbin Gongye Daxue Xuebao/Journal of Harbin Institute of Technology, 53(12), 243–263.

[20] V. Khetani, Y. Gandhi and R. R. Patil, "A Study on Different Sign Language Recognition Techniques," 2021 International Conference on Computing, Communication and Green Engineering (CCGE), 2021, pp. 1-4, doi: 10.1109/CCGE50943.2021.9776399.

[21] S. Chaudhary, R. Harsh, "Big Data Hysteria, Cognizance and Scope", 4th International Conference for Convergence in Technology (I2CT) 2018, IEEE, ISBN: 978-1-5386-5432-9/18, 2018.

[22] S. Chaudhary, R. Harsh, "Scope of Big Data Analytics in Bikaner Urban Water Management", Proceeding of International Conference on Computing Intelligence & Internet of Things (ICCIIoT)2018, International Journal of Computational Intelligence & IoT, Vol. 2,No. 3, Available at: HTTPS://www.ssrn.com/link/ijciiot-pip.html. ELSEVIER-SSRN (ISSN: 1556-5068), 2018.

[23] S. Chaudhary, R. Harsh, "Paradigm Shift of Water demand Forecasting Techniques", 3rd International Conference on Soft Computing: Theory and Applications, ScienceDirect, Procedia Computer Science 00(2018)000-000, Published by Elsevier Ltd. Selection 2018, Available at: www.elsevier.com/locate/procedia.

[24] S. Chaudhary, R. Harsh, "Epistemological View: Data Ethics, Privacy & Trust on Digital Platform", 2018 IEEE International Conference on Systems, Computation, Automation, Networking (ICSCAN 2018)" Manakula Vinayagar Institute of Technology, Pondicherry, 6-7 July 2018, pp: 1-6, DOI: 10.1109/ICSCAN.2018.8541166. Added on IEEE Explore Digital Library"22 Nov 2018, Available at: https://ieeexplore.ieee.org/abstract/document/8541166

[25] S. Chaudhary, R. Harsh, " Role of Ethics in Big data & Issues Faced by Indians", IEEE International Conference on Advances in Computing, Communication Control and Networking (ICACCCN2018), IEEE ISBN No: 978-1-5386-4119-4, 12-13 Oct 2018.

[26] S. Chaudhary, J. Manocha, "Finest Execution Time Approach for Optimal Execution Time in Mobile and Cloud Computing", International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), ISSN: 2321-8169, PP: 166 – 171, Vol: 6, Issue: 6 , June 18

[27] S. Chaudhary, P. Choudhary, "Motif and Conglomeration of Software Process Improvement Model", International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169,Vol:6, Issue:6, PP:163-165, June 2018.

[28] S. Chaudhary, A. Kiradoo, "CBIR by Using Features of Shape and Color", International Journal on future revolution in computer science & communication engineering, ISSN: 2454-4248, Vol: 4, Issue:9, PP:73-76, Sep 2018.

[29] S. Chaudhary and A. Jain, "Storage Security and Predictable Folder Structures in Cloud Computing", International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), ISSN: 2321-8169, vol.- 6,Issue no.- 5,pp. 109-116, May 2018.

[30] S. Chaudhary, M. Dave and A. Sanghi, "Enhance the Data Security in Cloud Computing by Text steganography", Springer/LNNS proceeding of the World Conference on Smart Trends in Systems, Security and Sustainability, ISSN: 2367-3370, Series:15180, pp. 1-8, Feb 2017.

[31] S. Chaudhary, G.Khatri, M. Dave and A. Sanghi, "Advancing the Potential of Routing Protocol in Mobile Ad Hoc Network" International Journal on Future Revolution in Computer Science & Communication Engineering ,ISSN: 2454-4248,Volume: 3,Issue: 11,PP: 125–128, November 2017.